"It's a lot more to it than just research": integrating critical data literacy and reasoning with data into a STEM summer camp

Received 1 July 2025
Revised 8 October 2025
18 February 2026
Accepted 19 February 2026

Marc T. Sager
The Budd Center, Simmons School of Education and Human Development, Southern Methodist University, Dallas, Texas, USA

Saki L. Milton
Simmons School of Education and Human Development, Southern Methodist University, Dallas, Texas, USA, and

Candace Walkington
Department of Teaching and Learning, Simmons School of Education and Human Development, Southern Methodist University, Dallas, Texas, USA

Abstract

Purpose – This study aims to explore how middle-grade girls from predominantly underrepresented and underserved racially and ethnically minoritized (UUREM) backgrounds developed critical data literacy (CDL) through participation in a week-long residential science, technology, engineering and mathematics (STEM) camp. Given the increasing importance of data science education in a data-driven world, this research examines how informal learning environments can support CDL development among youth from historically marginalized groups.

Design/methodology/approach – The study draws on qualitative interview data from 11 participants and their group-produced artifacts to investigate how the camp experience supported engagement with data and the development of CDL. Interviews explored participants' experiences with data collection, organization, analysis and interpretation within the context of real-world, community-relevant topics.

Findings – The findings demonstrate that participants engaged with data by collecting, organizing and analyzing information relevant to authentic problems. Meaning-making was facilitated through social interactions with peers and mentors, enabling participants to collaboratively interpret and critique data. Girls demonstrated critical examination of data by questioning sources, identifying limitations and considering potential biases. Through presentations, participants applied and communicated their findings, connecting data analyses to community-relevant concerns. Reflection and iteration emerged as participants reconsidered their approaches and interpretations over time.

© Marc T. Sager, Saki L. Milton and CandaceWalkington. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/

Funding: This research was partially funded through support from the Simmons School of Education and Human Development at Southern Methodist University. It was also partially funded through support from the National Science Foundation under Grant DRL 2341948. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Declaration of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Information and Learning Sciences
Vol. 127 No. 5-6, 2026
pp. 315-344
Emerald Publishing Limited
2398-5348
DOI 10.1108/ILS-07-2025-0102

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

Originality/value – This study contributes to research on informal learning, equity in STEM and data literacy by illustrating how informal STEM programs can intentionally foster CDL among UUREM girls. The findings highlight the role of collaborative meaning-making, critical reflection and community relevance in supporting youth to engage with data as a tool for understanding and addressing societal challenges.

Keywords Critical data literacy, Data science education, Data feminism, Girls in STEM, Informal learning, Middle school, Reasoning with data, Underrepresented and minoritized girls

Paper type Research paper

Data Science Education (DSE) is becoming increasingly important in today's data-driven world [National Academies of Sciences, Engineering, and Medicine (NASEM), 2022]. As data plays a vital role in shaping various aspects of society, DSE is essential for preparing learners for future careers, particularly within science, technology, engineering and mathematics (STEM), so that learners are equipped with the skills and knowledge necessary to navigate an increasingly complex digital world (NASEM, 2022). However, traditional approaches to DSE have historically excluded and marginalized girls, particularly underrepresented and underserved racially and ethnically minoritized (UUREM) girls, perpetuating systemic inequities in STEM fields (Calabrese Barton and Tan, 2020; Philip and Azevedo, 2017). Further, feminist approaches to DSE recognize that democratizing access to data literacy requires intentionally centering the experiences and perspectives of marginalized groups, particularly young women of color who have been systematically excluded from data-driven decision-making that affects their communities (D'Ignazio and Klein, 2020).

Critical data literacy (CDL) extends beyond a simplistic understanding of data – CDL incorporates a more nuanced comprehension of data's uses, societal impacts and its role in shaping the world (Louie, 2022; Stornaiuolo, 2020). The *critical* dimension of CDL draws from critical pedagogy and feminist theory, emphasizing how data is never neutral but embedded within systems of power that can either perpetuate or challenge social inequities (D'Ignazio and Klein, 2020; Freire, 1970). This critical lens examines who gets to collect data, what questions are asked, whose voices are included or excluded and how data-driven narratives can either reinforce dominant perspectives or amplify marginalized experiences (D'Ignazio and Klein, 2020). CDL enables learners to critically assess, interpret and navigate data in various contexts, helping them make informed decisions and recognize data as a tool for both justice and social good (Louie, 2022; Pangrazio and Selwyn, 2023). When grounded in feminist frameworks, CDL empowers learners, particularly those from historically marginalized communities, to interrogate data systems, challenge biased interpretations and use data as a tool for community advocacy and social transformation (Philip and Azevedo, 2017; Tschida *et al.*, 2014). Understanding CDL is vital for fostering informed decision-making and promoting critical thinking skills, which are essential for addressing complex societal issues, such as food justice, environmental racism and reproductive health equity, all of which disproportionately affect UUREM communities (Calabrese Barton *et al.*, 2021).

Research has consistently documented the historical marginalization of girls and women, particularly UUREM women, in STEM fields through exclusionary practices, biased curricula and unwelcoming environments (Calabrese Barton and Tan, 2020; Ireland *et al.*, 2018; Ong *et al.*, 2011). In data science specifically, this marginalization is evident in the underrepresentation of women and people of color in data science careers, the lack of diverse perspectives in algorithm development and the perpetuation of biased data systems that harm marginalized communities (Broussard, 2018; Criado-Perez, 2019; Noble, 2018). Feminist approaches to DSE argue that centering the experiences and knowledge of UUREM girls is

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

not merely about inclusion, but about fundamentally transforming how data science is taught and practiced to serve social justice goals (D'Ignazio and Klein, 2020; Tschida *et al.*, 2014).

While interest in DSE is growing, significant gaps exist in current research, particularly regarding CDL in informal learning environments (Sager *et al.*, 2026). Middle school learners, particularly girls, are underrepresented in DSE research. This demographic represents a crucial intervention point, as middle school is when many girls, especially UUREM girls, begin to disengage from STEM fields due to stereotype threat, lack of representation and culturally irrelevant curricula (Calabrese Barton and Tan, 2020; Ireland *et al.*, 2018; Ong *et al.*, 2011). Feminist pedagogical approaches emphasize that engaging UUREM girls in DSE is essential for democratizing data science and ensuring that data-driven solutions address the needs of diverse communities rather than perpetuating existing inequities (Philip and Azevedo, 2017). To truly democratize DSE, it is necessary to explore new environments where this knowledge can thrive, such as informal settings (Bell *et al.*, 2009; Calabrese Barton *et al.*, 2021; Lee *et al.*, 2021). The development of open-source curricula tailored to informal environments is critical for providing educators with the resources to engage diverse learners and build on their prior knowledge (Sager *et al.*, 2025). The research question guiding this paper is:

> How do middle school girls' experiences with critical data literacy in an informal learning environment shape their reasoning with data, specifically, how they conceptualize, interpret, and apply data—to form perceptions of the CDL process, including the challenges they face and the tools they use?.

The paper will begin with a review of the literature, examining CDL and reasoning with data. We will then outline the theoretical framework, grounded in sociocultural perspectives of learning integrated with feminist and data feminism frameworks, followed by the methods, findings and discussion. The paper will conclude with the implications and limitations of the study.

Literature review

Critical data literacy

CDL encompasses several key components that are essential for learners to engage effectively with data. (Louie, 2022; see Figure 1). Grounded in critical pedagogy, CDL goes beyond technical skills to examine how data systems reflect and reproduce power structures, whose voices are privileged or silenced in data collection and interpretation and how learners can use data as a tool for social justice and community transformation (D'Ignazio and Klein, 2020; Freire, 1970). One of the foundational aspects is formulating relevant and structured questions, which guide the data collection process and ensure that the gathered data aligns with the goals of the analysis (Bargagliotti *et al.*, 2020). From a critical perspective, this questioning process is inherently political, as it determines whose problems get attention, what solutions are considered and which communities benefit from data-driven insights (D'Ignazio and Klein, 2020). Learners then focus on gathering or compiling appropriate data from reliable sources, honing their ability to assess data for relevance and quality. CDL emphasizes examining not just data quality, but understanding who collected the data, for what purposes, what assumptions guided collection methods and whose perspectives might be missing or marginalized (Philip and Azevedo, 2017; Philip *et al.*, 2016).

Employing analytical tools is another key step, where learners gain proficiency in numeracy and statistics, skills necessary for analyzing and interpreting data (Bowler *et al.*, 2019; Letourneau *et al.*, 2020). However, CDL recognizes that analytical tools and methods are not neutral but embed particular worldviews and assumptions that may privilege dominant perspectives while marginalizing others (D'Ignazio and Klein, 2020). For example, in one activity proposed by Letourneau *et al.* (2020), children recorded time spent

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

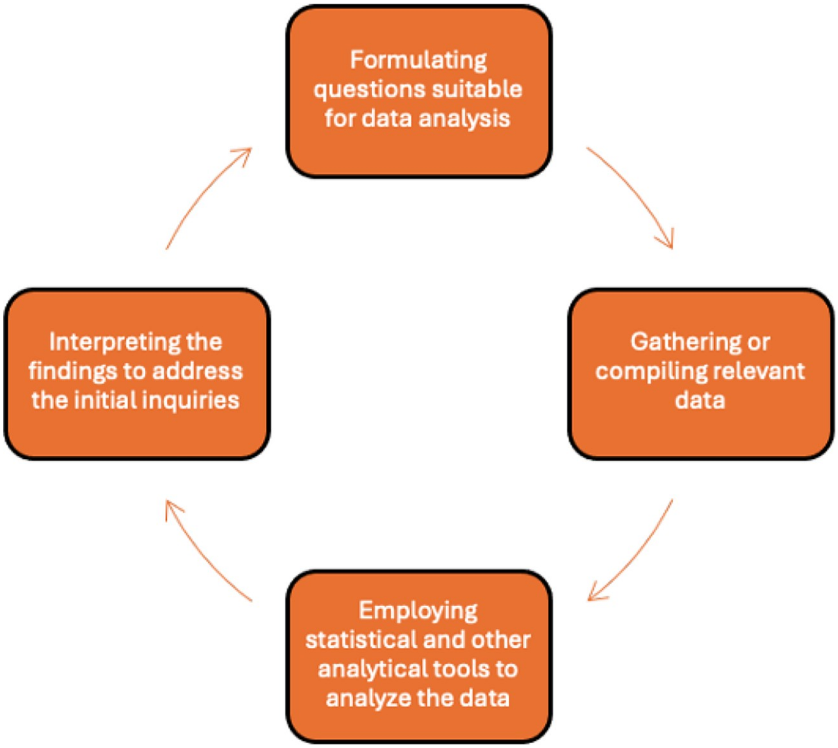

Figure 1. Four-step cyclical process of comprehending data literacy
Note(s): Derived from Bargagliotti *et al.* (2020); Reprinted from Sager (2026)

at various museum exhibits and collaboratively organized their data on a continuous line, engaging in an interpretive process that underscored how different visualizations can lead to varying conclusions.

In the next stage, learners focus on interpreting the findings, a crucial step in drawing meaningful conclusions from data. From a CDL perspective, interpretation involves not just technical analysis, but critical examination of whose stories are told through data, what narratives are privileged or silenced and how different communities might interpret the same data differently based on their lived experiences (D'Ignazio and Klein, 2020; Tschida *et al.*, 2014). Here, learners comprehend how different representations of data, such as graphs or spreadsheets, can influence decision-making and interpretation (Bargagliotti *et al.*, 2020; Louie, 2022). Furthermore, CDL involves developing an awareness of data privacy concerns and the ethical implications of data usage (Lee *et al.*, 2021; Lee *et al.*, 2022). This ethical dimension is particularly crucial from a feminist and social justice perspective, as marginalized communities have historically been subject to harmful data collection and surveillance practices, making questions of consent, ownership and benefit-sharing central to ethical data practice (Benjamin, 2019; Browne, 2015; Criado-Perez, 2019). Understanding these aspects ensures that learners are equipped not only to handle data responsibly but also to navigate the ethical challenges that arise in our increasingly data-driven world (Bowler *et al.*, 2019; Hautea *et al.*, 2017). Together, these skills form a comprehensive approach to

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

CDL, enabling learners to analyze, interpret and ethically engage with data in meaningful ways while working toward social justice goals and community empowerment.

Reasoning with data

Reasoning with data is fundamentally a socially and culturally mediated practice, where meaning-making with data emerges through interaction, language and participation in communities of practice (McClain and Cobb, 2001; Rosenberg, 2018). This social dimension of data reasoning is particularly important because it recognizes that different communities bring distinct knowledge systems, values and ways of understanding the world that can enrich and challenge dominant approaches to data interpretation (Bang and Medin, 2010; D'Ignazio and Klein, 2020; Tschida *et al.*, 2014). This perspective moves beyond viewing statistical reasoning as purely cognitive to understanding it as embedded in social contexts and cultural practices. Research (i.e. McClain and Cobb, 2001; Petrosino *et al.*, 2023) has established that statistical reasoning develops through social negotiation and discourse. Students construct understanding not in isolation, but through collaborative engagement with data, where different interpretations are shared, questioned and refined through group interaction (McClain and Cobb, 2001). This collaborative meaning-making process is particularly powerful when it centers the experiences and perspectives of marginalized groups, as it can challenge dominant narratives and create space for alternative interpretations that reflect diverse community needs and values (Philip and Azevedo, 2017; Philip *et al.*, 2016).

The cultural dimension of data reasoning becomes particularly evident when learners connect data to their lived experiences and local contexts. Wilkerson and Laina (2018) demonstrated this through their study of middle school students who repurposed publicly available data to explore questions about their own communities. This approach aligns with critical perspectives that emphasize the importance of connecting learning to lived experiences and recognizing the knowledge that marginalized communities possess about their own circumstances (hooks, 1994; Tschida *et al.*, 2014). Key reasoning opportunities emerged when learners worked to reconcile two critical relationships: first, their own knowledge and experiences of the context from which data were collected with the details of the data provided; and second, their different emerging stories about the data with one another (Wilkerson and Laina, 2018). This process of reconciliation required students to draw upon their cultural knowledge while engaging in collaborative meaning-making through data storytelling (Wilkerson and Laina, 2018).

The role of student choice in topic selection is also crucial for connecting personal experience to data interpretation (Fielding *et al.*, 2025; Kanari and Millar, 2004; Kazak *et al.*, 2023). From a critical perspective, this choice is not merely pedagogically effective but politically important, as it shifts power from educators to learners and enables marginalized students to investigate issues that directly affect their communities (Freire, 1970; Philip and Azevedo, 2017). When students can choose socio-scientifically relevant topics that connect to their communities and experiences, they are better positioned to engage in meaningful reasoning that bridges their cultural knowledge with statistical concepts (Fielding *et al.*, 2025; Lee *et al.*, 2021; Wilkerson and Laina, 2018). This approach is particularly important for UUREM girls, who may be more engaged when they can explore data related to issues, such as environmental justice, food access or health disparities, that directly impact their communities (Calabrese Barton and Tan, 2020). These findings directly relate to the current study, where campers had the choice of a socio-scientific topic and data to use for analysis and infographic creation. This design aligns with research showing that student agency in topic selection supports the social and cultural dimensions of data reasoning, allowing

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

learners to draw upon their experiences while engaging in collaborative meaning-making practices (Lee *et al.*, 2022).

Theoretical framework

Sociocultural perspectives on learning

Situated and sociocultural learning theories provide valuable frameworks for informal DSE, particularly in addressing social justice issues (e.g. food and environmental justice). Situated learning highlights the importance of engaging learners in authentic, real-world tasks (Bang, 2015; Greeno, 2006; Lave and Wenger, 1991; Vygotsky, 1978). By working with real data sets, such as analyzing local food deserts or pollution, learners can see the direct impact of their work on their communities, making learning more relevant (Lee *et al.*, 2021; Wilkerson and Polman, 2020).

Sociocultural perspectives emphasize that learning occurs through social interaction and is mediated by cultural and contextual factors (Rogoff, 2003; Rogoff *et al.*, 2016). *Reasoning with data* is not just an individual cognitive skill but a socially and culturally situated process that unfolds through participation in meaningful activities (McClain and Cobb, 2001). This process is mediated by:

- Tools, Discourse and Community Practices, in which learners engage with data visualization tools, discussions and collaborative inquiry to make sense of data within real-world contexts (Brown *et al.*, 1989; Lee *et al.*, 2022). From a critical perspective, the choice of tools and discourse practices can either perpetuate dominant ways of knowing or create space for alternative epistemologies and ways of understanding data that reflect marginalized communities' knowledge systems (D'Ignazio and Klein, 2020.
- Guided Participation and Scaffolding, where learners refine their reasoning through mentorship, structured supports and peer collaboration (Lee *et al.*, 2022). Critical pedagogical approaches emphasize that effective scaffolding for marginalized learners requires recognizing and building upon their existing strengths and knowledge rather than treating them as a deficit (Tschida *et al.*, 2014).
- Cultural and Contextual Factors, where learners' lived experiences shape how they interpret and critique data, influencing how they frame and address socio-scientific challenges (Lee *et al.*, 2021; Wilkerson and Polman, 2020).

For UUREM girls, these cultural and contextual factors include experiences of marginalization, community resilience and knowledge about social justice issues that can provide unique insights into data interpretation and application (Calabrese Barton and Tan, 2020; Philip and Azevedo, 2017).

CDL further informs this framework by emphasizing how reasoning with data involves questioning data sources, identifying biases and recognizing the role of power in data use (D'Ignazio and Bhargava, 2020; Lee *et al.*, 2021). This perspective aligns with informal DSE by fostering learners' ability to critically examine data within their own communities.

Data feminism

Data feminism provides a crucial framework for understanding how DSE can be transformed to serve social justice goals. D'Ignazio and Klein (2020) outline seven principles of data feminism that directly inform educational practice. The first principle, examining power, involves understanding how data systems reflect and reinforce existing power structures and working to challenge these structures. Building on this foundation, the second principle

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

encourages challenging power by using data science skills to question dominant narratives and center marginalized perspectives (D'Ignazio and Bhargava, 2020). The third principle, elevating emotion and embodiment, recognizes that data represents human experiences and that emotional and embodied ways of knowing are valuable alongside statistical analysis (D'Ignazio and Klein, 2020).

The fourth principle calls for rethinking binaries and hierarchies by challenging false dichotomies such as objective versus subjective and recognizing the value of multiple perspectives and ways of knowing. Similarly, the fifth principle emphasizes embracing pluralism, acknowledging that there are multiple valid ways to approach data analysis and interpretation, particularly from different cultural perspectives (D'Ignazio and Bhargava, 2020). The sixth principle stresses the importance of considering context, understanding that data is always collected and interpreted within particular social, cultural and historical contexts that shape its meaning (D'Ignazio and Klein, 2020). Finally, the seventh principle advocates for making labor visible by acknowledging the often-invisible work that goes into data collection, cleaning and analysis and ensuring that marginalized communities benefit from data work done about them (D'Ignazio and Klein, 2020).

These principles inform the current study by providing a framework for understanding how UUREM girls' engagement with data science can be both personally empowering and socially transformative. Rather than viewing DSE as politically neutral skill-building, data feminism recognizes it as inherently political work that can either perpetuate or challenge existing inequities (D'Ignazio and Bhargava, 2020; D'Ignazio and Klein, 2020).

Feminist approaches to data science

Feminist approaches to DSE recognize that traditional STEM education has systematically excluded and marginalized girls and women, particularly those from UUREM communities (Calabrese Barton and Tan, 2020; Ireland *et al.*, 2018; Ong *et al.*, 2011). These approaches argue that democratizing data science requires not just adding more girls to existing programs but fundamentally transforming how data science is taught and practiced, centering social justice goals and marginalized perspectives (D'Ignazio and Klein, 2020; Tschida *et al.*, 2014). Key principles of feminist approaches to DSE include:

- Centering Marginalized Voices: Rather than treating UUREM girls as a deficit population to be "fixed," feminist approaches recognize them as holders of valuable knowledge and perspectives that can enrich and challenge dominant approaches to data science (hooks, 1994; Philip and Azevedo, 2017).
- Connecting to Social Justice Issues: Feminist pedagogy emphasizes connecting learning to issues of social justice and community transformation, recognizing that meaningful engagement often occurs when learners can use their skills to address problems affecting their communities (Freire, 1970; Tschida *et al.*, 2014).
- Collaborative and Participatory Learning: Feminist approaches emphasize collaborative learning processes that value different ways of knowing and recognize that knowledge is constructed through community rather than individual achievement (hooks, 1994).
- Critical Examination of Power: Feminist DSE explicitly examines how data systems reflect and reproduce power structures, preparing learners to challenge biased practices and create more equitable data systems (D'Ignazio and Klein, 2020).

By framing *reasoning with data* as a socially and culturally mediated process, rather than a standalone cognitive skill and integrating feminist and data feminism frameworks, this

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

approach deepens learners' understanding of data science while empowering them to engage in community-driven change (Rosenberg, 2018; Wilkerson and Laina, 2018). When grounded in social justice and feminist principles, this approach is particularly powerful for UUREM girls, who can use data science skills to investigate and address issues affecting their communities while developing positive STEM identities grounded in social justice values (Calabrese Barton and Tan, 2020; Philip and Azevedo, 2017). Integrating these theoretical perspectives into informal DSE programs creates learning environments that are not only contextually relevant but also socially transformative (Esmonde and Booker, 2017; Sager *et al.*, 2025).

Research design

This study used a case study approach (Stake, 1995) to broadly examine how UUREM girls make meaning from data science experiences. A case study approach was selected as the most appropriate methodology for several reasons:

- It allows for in-depth exploration of complex phenomena within their real-world context, particularly suited for understanding how UUREM girls navigate and make meaning from CDL experiences (Stake, 1995).
- It enables examination of the interconnections between individual experiences and the social, cultural and educational contexts that shape them, aligning with our sociocultural and feminist theoretical frameworks (Bang, 2015; D'Ignazio and Bhargava, 2020).
- It provides flexibility to capture the nuanced ways that marginalized students engage with data science when their experiences and perspectives are centered (Patton, 2015).

Given the underrepresentation of UUREM girls in DSE research, this exploratory case study approach allows for rich description and analysis that can inform future interventions and research (Creswell and Guetterman, 2019).

Context

The project involved an established one-week residential STEM camp hosted by a local nonprofit. The camp took place at a small public university in the southwestern U.S. during the summer of 2024 (Milton *et al.*, 2025). The program's topics (see Table 1) were structured around daily themes to enhance the learning experience and ensure cultural relevance. Since 2010, the camp has been hosted at various universities. Ethical approval for the study was granted by the researchers' university Institutional Review Board, and parental consent was obtained. Girls were actively recruited from a variety of settings, including coed. and all-girls public, private and public charter middle schools across a major metropolitan area. Recruitment efforts also reached out to public libraries, predominantly black churches, cultural centers, youth groups and other venues where UUREM girls gather. This recruitment strategy was intentionally designed to reach UUREM girls who might not otherwise have access to DSE opportunities, aligning with feminist pedagogical principles of actively seeking out and centering marginalized voices rather than waiting for them to self-select into STEM programs (hooks, 1994; Tschida *et al.*, 2014). A partnership between the nonprofit running the camp and a statewide nonprofit school network played a key role in recruiting girls from local all-girls schools. Social media campaigns, including targeting relevant online groups like Facebook parent groups, were also used to boost recruitment.

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

Table 1. Overview of topics by grade level

Grade level	Math topics	Socio-scientific topics	Data science project driving questions
7th	Proportionality and data literacy	Food insecurity	How do food deserts, defined as areas with limited access to affordable and healthy food options, disproportionately affect marginalized communities?
8th	Linear equations	Environmental injustices	What are the environmental hazards or pollutants in the target area that may impact health, such as air pollution, water contamination, or exposure to harmful chemicals?
9th	Algebra 1 –Quadratic Equations	Mental health Vaccine access	What are the existing barriers to accessing reproductive health services, such as distance to clinics, cost, cultural stigmas, or lack of education?
		Reproductive health Healthy lifestyle	Are there disparities in vaccination rates among different demographic groups, and if so, what are the underlying causes?

During the camp, participants took courses on grade-appropriate mathematics, science and data science topics, as they related to the state's learning standards. The first two authors directed the curriculum development for the camp. The math curriculum was developed by an experienced middle school mathematics teacher and instructional coach, while the science curriculum was developed by an experienced science teacher and science curriculum lead. The data science curriculum was developed by the first author, a learning scientist who conducts research within DSE. The teachers at the camp were experienced mathematics and science teachers and held the appropriate teaching credentials relevant to the state in which the study occurred.

The camp emphasized data science and data literacy, particularly in the context of socio-scientific topics (Lee *et al.*, 2021; Milton *et al.*, 2023; Milton *et al.*, 2025). The integration of socio-scientific topics was intentional and theoretically grounded in feminist and critical pedagogical approaches that emphasize connecting learning to issues of social justice and community relevance (D'Ignazio and Klein, 2020; Freire, 1970). This approach recognizes that UUREM girls are more likely to engage with STEM learning when they can see connections to issues affecting their communities and when their existing knowledge about social justice issues is valued and built upon (Calabrese Barton and Tan, 2020; Philip and Azevedo, 2017). Throughout the week, campers worked collaboratively in groups to apply data science skills toward creating infographics that address socio-scientific topics. The camp's learning objectives include understanding the connections between secondary education, higher education and career pathways; developing effective written and verbal communication skills for various audience needs; fostering collaborative professionalism among peers and adults; and using technology proficiently to inform or persuade audiences (see Table 1 for an overview of the topics by grade level).

Data science curriculum. Throughout the camp, campers progressed through a structured data science process (Lee *et al.*, 2022) with the goal of designing infographics. The campers were divided into a total of 11 groups to collaborate on the data science capstone project. Each group comprised 3–5 campers, and the groups were separated by grade level. The balance between structure and student autonomy was carefully designed to align with feminist pedagogical principles. While the broad socio-scientific themes

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

were predetermined based on grade-appropriate content standards and social justice issues relevant to UUREM communities (see Table 1), campers exercised significant autonomy within these frameworks. Within each theme, groups had complete freedom to select their specific research questions, choose their data sources, determine their analytical approach and decide how to frame their findings for their target audience. For example, within the 7th-grade "Food Insecurity" theme, groups independently chose to focus on different geographic regions (Houston, Chicago, Dallas) and different aspects of food access (food deserts, urban farms, healthy food environments). This structure ensured curriculum alignment while honoring critical pedagogical principles that emphasize student voice and the importance of connecting learning to students' own interests and community concerns (Freire, 1970; Tschida *et al.*, 2014). This approach ensured that while learning objectives were met, students could bring their own interests, experiences and perspectives to bear on the data science process.

The campers began their projects by brainstorming topics relevant to their grade levels, identifying relevant data sources and assigning roles within their teams to foster effective collaboration. During this brainstorming phase, facilitators encouraged students to draw connections between the assigned themes and issues they had observed or experienced in their own communities, creating space for their lived experiences to inform their data science work (Philip and Azevedo, 2017). Using tools like the Common Online Data Analysis Platform (CODAP; The Concord Consortium, 2026), campers then analyzed data to identify trends and develop research questions to guide the creation of their infographics. Next, campers employed data visualization techniques, such as pie charts, bar charts, histograms and scatterplots, to visually depict trends or comparisons pertinent to their grade's themes. Finally, the groups participated in iterative feedback sessions, refined their designs based on input and completed their infographics, ensuring their visualizations effectively conveyed meaningful insights to a community audience (see Table 2 for an overview of the data science project and Table 3 for each group's topic; see Figure 2 for an example of learners' produced visualizations).

Table 2. Overview of data science project

Project phase	Activities	Tools/techniques
Topic brainstorming	Campers brainstorm campaign topics, identify relevant data sources and assign roles within their teams to encourage effective collaboration	Collaboration and research
Data analysis	Campers analyze data using tools like CODAP, pinpoint trends and develop research questions that will guide the creation of their infographics	CODAP, data analysis
Data visualization	Campers employ basic data visualization techniques to depict trends or comparisons relevant to their campaign themes	Pie charts, bar charts, histograms, scatterplots
Iterative feedback and refinement	Teams participate in feedback sessions, refine their infographic designs based on input and ensure their visualizations effectively convey insights to their target audience	Feedback sessions, design refinement
Infographic completion	Campers finalize their infographics, ensuring the visualizations effectively communicate the intended message to the audience	Final design and presentation

Note(s): The data science curriculum was grounded in project-based learning (Barron *et al.*, 1998; Petrosino *et al.*, 2023)

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

Table 3. Each groups' topic for their infographic

Group no.	Grade	Socio-scientific topic
1	8th	How a flood can affect soil pollution
2	8th	How our health is impacted by soil pollution
3	9th	Reproductive rights
4	9th	Healthy lifestyles and childhood obesity in the USA
5	8th	Flood risk in Florida
6	7th	Food access in Houston
7	7th	Food deserts and access to urban farms in Chicago
8	8th	Effects of air pollution
9	8th	Wildfires effect on air pollution
10	7th	Health food environments
11	7th	Food access in Dallas

Facilitator roles and scaffolding

The learning environment was supported by a team of experienced middle school mathematics and science teachers, along with high school and college-aged mentors assigned to each group. Facilitators employed a distributed scaffolding approach aligned with critical pedagogical principles of shared authority (Freire, 1970; hooks, 1994). Rather than directing student work, facilitators served as research consultants, providing guidance when requested and offering structured reflection prompts during daily check-ins.

Specific facilitation strategies included the following:

- Daily 15-minute reflection circles where groups shared challenges and received peer suggestions;
- "Critical friends" protocols where groups presented preliminary findings to other groups for feedback on source credibility and evidence quality;
- Open-ended questioning techniques that encouraged students to justify their analytical choices ("What makes you confident in this data source?" "How might someone from a different community interpret this differently?"); and
- Just-in-time technical support with CODAP and visualization tools without taking over student decision-making processes.

This approach was designed to maintain student agency while providing the scaffolding necessary for rigorous data analysis, reflecting feminist commitments to collaborative learning and shared expertise rather than traditional hierarchical teacher-student relationships (Tschida *et al.*, 2014).

Participants

The study's participants consisted of 34 middle-grade girls (see Table 4). The demographic breakdown of the participants was as follows: 70% identified as Black/African American ($n = 24$), 12% as Asian/Pacific Islander ($n = 4$), 12% as Hispanic/Latina ($n = 4$) and 5% as White ($n = 2$). Regarding grade levels, 32% of the participants were in the 7th grade ($n = 11$), 50% were in the 8th grade ($n = 17$) and 18% were in the 9th grade ($n = 6$). The age distribution was as follows: 3% of the participants were 11 years old ($n = 1$), 29% were 12 years old ($n = 10$), 50% were 13 years old ($n = 17$) and 18% were 14 years old ($n = 6$). This demographic composition reflects the intentional focus on UUREM girls, with 94% of participants identifying as girls of color, providing an important opportunity to center the

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

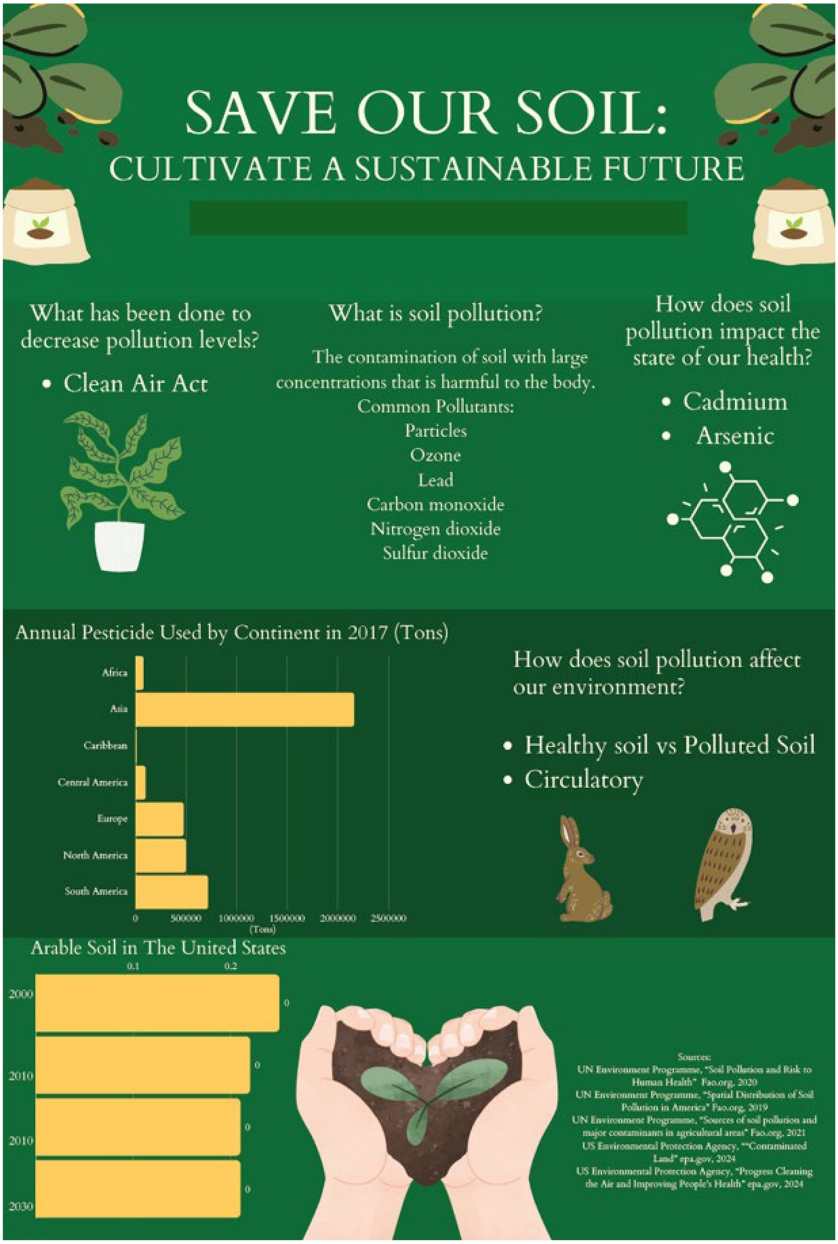

Figure 2. Group 2's final infographic

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

Table 4. Participant demographics

Demographics	Sample (N)	Sample %
Gender		
Female	34	100
Race/Ethnicity		
Black/African American	24	70
Asian/pacific islander	4	12
Hispanic/Latina	4	12
White/Caucasian	2	5
Grade level		
7th	11	32
8th	17	50
9th	6	18
Age		
11	1	3
12	10	29
13	17	50
14	6	18

experiences and perspectives of students who have been historically marginalized in STEM and DSE (Calabrese Barton and Tan, 2020; Ong *et al.*, 2011).

Data collection

For this study, we implemented an exploratory case study using semi-structured interviews (Stake, 1995). We conducted 11 semi-structured interviews (one camper from each of the data science project groups; see Table 5 for an overview of interviewees) to gain qualitative insights into the campers' experiences with the project. The selection of interview participants followed a purposeful sampling (Creswell and Guetterman, 2019) strategy designed to ensure representation across all project groups while maintaining feasibility within the case study approach. From each of the 11 project groups, one participant was

Table 5. Interview participants

Group	Name (pseudonym)	Grade	Race/Ethnicity
1	Raven	8th	Black
2	Anna	8th	White
3	Brielle	9th	Black
4	Tiana	9th	Black
5	Riley	8th	Black
6	Alexis	7th	Black
7	Amani	7th	Black
8	Zuri	8th	Black
9	Makayla	8th	Black
10	Jada	7th	Black
11	Nia	7th	Black

Note(s): Campers' race/ethnicity was self-reported, and all names are pseudonyms.

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

selected by the high school/college-aged mentor assigned to that group. Mentors were asked to identify a group member who had been actively engaged in the project work and who they believed could articulate their group's experience effectively. This sampling approach ensured that we captured perspectives from each distinct project focus area while leveraging the mentors' insights about which students were most reflective about their learning process (Patton, 2015).

While this case study relies primarily on interview data as the main source of evidence, this approach is justified for several reasons: the depth and richness of semi-structured interviews align with feminist research methodologies that prioritize listening to and amplifying marginalized voices (hooks, 1994); interviews allow participants to share their experiences in their own words and from their own perspectives, which is crucial for understanding how UUREM girls make meaning from data science experiences (Tschida *et al.*, 2014); and the exploratory nature of this research, given the limited existing literature on UUREM girls' experiences with CDL, makes in-depth interviews an appropriate starting point for understanding this phenomenon (Patton, 2015).

These interviews provided in-depth perspectives on how the project influenced their engagement with data science and their overall learning experience. This data collection approach allowed us to better assess the qualitative impacts of the data science project on the participants. While semi-structured interviews served as the primary data source, this study also incorporated analysis of student-created infographics as artifacts that provided additional insight into learning processes and outcomes. These visual artifacts served multiple analytical purposes:

- They provided evidence of students' data source selection and evaluation processes through examination of their citations and source diversity.
- They revealed students' approaches to data visualization and audience communication through their design choices and statistical presentation methods.
- They offered a window into students' synthesis and interpretation processes through analysis of how they connected data patterns to broader social issues.

The combination of interview data capturing students' reflective accounts of their experiences with artifact analysis of their actual data science products provided complementary perspectives that enhanced the trustworthiness of our findings and allowed for triangulation of key themes around CDL development.

Data analysis

For our data analysis, we analyzed the transcribed interview data collectively using a deductive coding method (Creswell and Guetterman, 2019; see Table 6 for our codebook). Our coding scheme development was a systematic process that began with an extensive literature review of CDL, reasoning with data and critical approaches to STEM education. We identified key constructs from our theoretical frameworks, particularly sociocultural learning, CDL and data feminism and translated these into operational codes that could capture how these constructs manifested in students' descriptions of their experiences (Miles *et al.*, 2020). The initial codebook was developed collaboratively by the first two authors, with input from the third author, ensuring that codes were both theoretically grounded and empirically relevant to our data. Specifically, our codes were designed to capture the following:

- how students engaged with data as both technical content and social construct, reflecting CDL principles (D'Ignazio and Klein, 2020; Louie, 2022);

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

Table 6. Codebook

Code	Definition	Example	Citation
Engagement with data	Learners' interaction with datasets, research, and information gathering	*Positive*: "We used multiple research sources including national geographic and history websites" *Challenge*: "We were having trouble finding recent data related to soil pollution"	Bargagliotti *et al.* (2020); Bowler *et al.* (2019); Lee *et al.* (2022); Letourneau *et al.* (2020)
Meaning making through social interactions	Collaborative learning and peer engagement	*Positive*: "We all just started, like commenting on each other's parts so that we can help each other out" *Challenge*: "some of us wanted to party, some of us wanted to work"	Bargagliotti *et al.* (2020); Lee *et al.* (2022); Sager *et al.* (2025); Wilkerson and Polman (2020)
Critical examination of data	Evaluating sources, bias, and data quality	*Positive*: "you got to pick the actual research. It can't just say, oh, women are abused, and you gotta be like, the statistics and the facts" *Challenge*: "We had to abandon some tools due to inconsistent results"	Bargagliotti *et al.* (2020); Hautea *et al.* (2017); Lee *et al.* (2021); Louie (2022)
Application and communication of findings	Using technology tools and presenting results	*Positive*: "technology made it easier for us. So, we don't have to take the extra time to like draw graphs" *Challenge*: "We wrote a little too many words. So, we were having to work on like, shorten it down"	Bargagliotti *et al.* (2020); Lee *et al.* (2022); Wilkerson and Laina (2018)
Reflection and iteration	Learning from challenges and improving work	*Positive*: "it came out better than we thought" *Challenge*: "the project was very stressful at times"	Fielding *et al.* (2025); Petrosino *et al.* (2023); Rosenberg (2018)

- how collaborative meaning-making processes unfolded, aligning with sociocultural perspectives of learning (Bang, 2015);
- how students drew upon their lived experiences and cultural knowledge, consistent with feminist pedagogical approaches (hooks, 1994; Tschida *et al.*, 2014); and
- how students navigated challenges and reflected on their learning, reflecting critical thinking and iterative learning processes (Freire, 1970).

To reduce bias, the first two authors (shared first authorship) collaboratively engaged in the coding process. This collaborative approach was particularly important given the study's focus on UUREM girls' experiences, as it allowed for the first-listed author (white man) and second-listed author (black woman) to bring different perspectives to the interpretation of data, helping to ensure that dominant perspectives did not overshadow participants' voices (Milner, 2007). The second author's positionality as a black woman with experience in STEM education was particularly valuable in interpreting participants' experiences and ensuring that cultural and social justice dimensions of their responses were appropriately recognized and analyzed. When disagreements arose, we refined the codebook definitions until reaching a consensus (Wasser and Bresler, 1996). Our consensus-building process involved returning to the theoretical literature and to participants' exact words to ensure that our interpretations remained grounded in both established frameworks and participants' own meaning-making.

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

Positionality

The first-listed author, a white man and learning scientist, served as co-research lead of the camp and developed the data science curriculum. He worked closely with the second-listed author, attending training sessions and meetings and supported the research process throughout the camp. The second-listed author, a black woman pursuing a PhD in STEM education, has a deep connection with the featured STEM summer camp, shaped by her own informal learning experiences and professional work in STEM education. She specializes in middle school STEM education and educational technologies. The third author, a white woman and mathematics education professor, provided mentorship and contributed insights to the research and data analysis. Given the study's focus on UUREM girls' experiences with CDL, the research team's diverse positionalities were both a strength and a consideration throughout the research process. The second author's identity and experiences as a black woman in STEM provided crucial insights for understanding participants' experiences and ensuring that analysis remained attentive to issues of race, gender and social justice. The first and third authors' positions as white researchers required ongoing critical reflection about how their perspectives might influence data collection and analysis, leading to intentional strategies such as having the second author take the lead on interview analysis and ensuring that interpretations remained grounded in participants' own words rather than researchers' assumptions (Milner, 2007; Tschida *et al.*, 2014).

Findings

This analysis of camper interviews and artifacts reveals five key themes that characterize the camp experience, each emerging from and contributing to sociocultural perspectives of learning and data feminism frameworks. The findings demonstrate how authentic data science experiences reflect the socially situated nature of learning and the importance of examining power structures within data practices, highlighting both the transformative potential and inherent challenges of collaborative data science learning for young learners (see Table 7 for the frequency of each code by total number of instances and by participants).

The emergent themes reflect how learning occurs through cultural tools and social interaction, while simultaneously revealing how middle-grade learners navigate and challenge existing power dynamics in data representation and interpretation. These findings illustrate the intersection of collaborative knowledge construction and critical examination of whose voices and experiences are centered in data science.

Progressive skill development through culturally mediated data engagement

This theme demonstrates how learners develop data science competencies through culturally mediated tools and social interactions, with learning occurring through engagement with

Table 7. Frequency of codes by total number of instances and participants

Code	Frequency – total no. of instances	Frequency – no. of participants (out of 11)
Engagement with data	29	11
Meaning making through social interactions	20	11
Critical examination of data	16	9
Application and communication of findings	27	11
Reflection and iteration	23	11

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

personally and culturally relevant problems that challenge traditional power structures in data representation. Learners consistently demonstrated a progression from initial uncertainty to confidence through sustained engagement with real-world datasets. This transformation was particularly evident in how learners approached complex environmental and social issues through data analysis.

Raven's (Group 1, Grade 8) experience with soil pollution research exemplified this trajectory: "We're looking at the ways that flooding can affect soil." This quote, while initially seeming simple, illustrates depth when examined alongside their final infographic

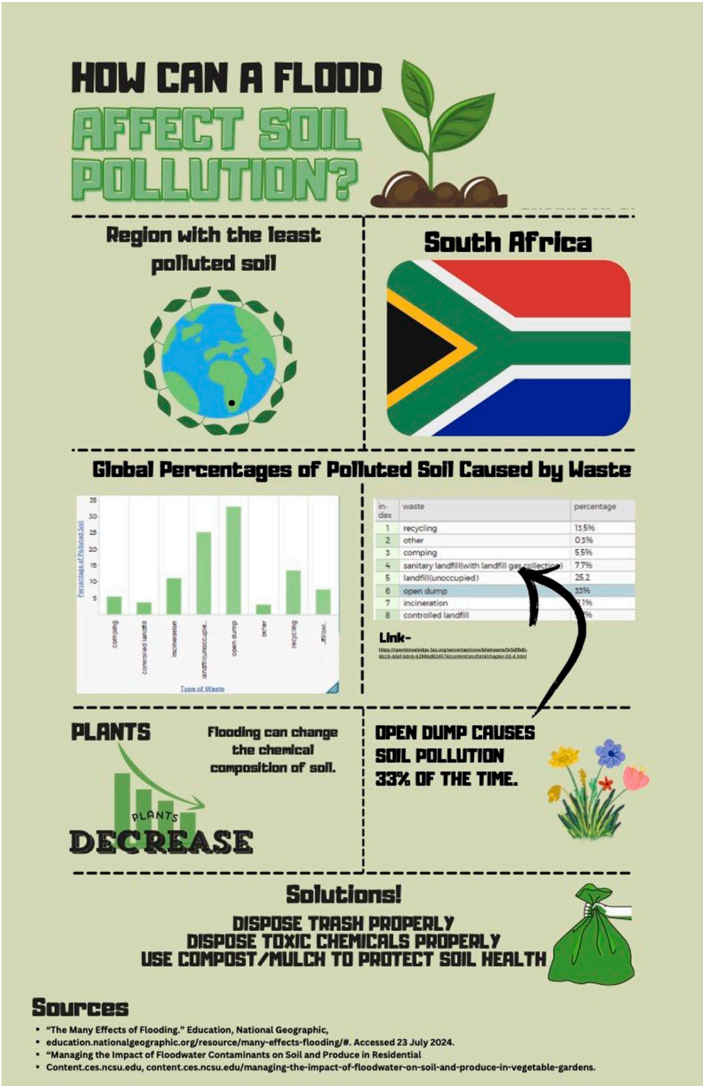

Figure 3. Group 1's produced infographic on soil pollution

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

(see Figure 3), which demonstrates an understanding of flood-soil relationships. Their group created a comprehensive analysis showing that "Flooding can change the chemical composition of soil," supported by sources from National Geographic Education and North Carolina State University Extension. The infographic reveals their engagement extended to understanding regional variations, identifying "South Africa" as the "Region with the least polluted soil" and connecting waste disposal practices to soil contamination with the specific statistic that "open dump causes soil pollution 33% of the time."

Anna's (Group 2, Grade 8) systematic approach was evident in her data collection process: "We found a lot of .gov websites that were accessible, and we verified them." This methodical practice is substantiated by their final infographic (see Figure 1), which exclusively cites authoritative sources, including multiple United Nations Environment Programme reports from FAO.org (Food and Agriculture Organization of the United Nations; 2019, 2020, 2021) and US Environmental Protection Agency (EPA) publications from EPA.gov (2024). Their source list demonstrates not only the verification process Anna described but also temporal awareness, incorporating both historical data (2017 pesticide use) and current policy information (2024 EPA updates).

The progression observed here reflects how cultural tools and artifacts mediate learning, as learners moved from surface-level data collection to deep pattern recognition through engagement with issues affecting their communities and identities. The depth of engagement varied by project complexity, with Anna's group pursuing multi-variable analysis. Their infographic integrates multiple data types: continental pesticide usage statistics (showing Asia using nearly 2.5 million tons annually), temporal trends in arable soil changes in the USA (spanning 2000–2030) and health impact categorizations across different body systems. This integration demonstrates their ability to synthesize diverse datasets into coherent environmental narratives.

Learners showed remarkable persistence when facing data availability challenges. Anna noted, "We were having trouble finding recent data related to soil pollution," yet their final product successfully incorporates current information. Their infographic includes 2024 EPA data on contaminated land and recent progress reports on air quality improvements, indicating they continued searching until finding current sources the groups deemed reputable, spanning the timeframe they needed.

Collaborative knowledge construction through social meaning-making

This theme demonstrates how knowledge construction occurs through social interaction and shared cultural practices, while also revealing how collaborative processes can both reproduce and challenge existing power dynamics within group work. The collaborative meaning-making processes observed reflect the distributed nature of cognition and the importance of multiple perspectives in data interpretation, particularly when examining issues of social justice and equity. Social interactions emerged as fundamental to learning outcomes and project quality, with learners developing an understanding of collaborative benefits beyond simple task division. The data revealed that peer collaboration served multiple functions: efficiency enhancement, quality control and knowledge construction.

Anna (Group 2, Grade 8) articulated the cognitive benefits of collaboration: "When we collaborate, we can bounce ideas off of each other." and emphasized quality assurance aspects: "We can help fix mistakes that maybe the other person didn't catch." This peer review function exemplifies how social scaffolding occurs through collaborative interaction, while also highlighting power dynamics as some learners took on quality control roles over others' work. Additionally, this function proved crucial for maintaining project standards, as

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

Raven (Group 1, Grade 8) described: "We all just started, like commenting on each other's parts so that we can help each other out."

The challenges learners encountered reveal how collaborative work can reproduce hierarchies and unequal participation, particularly around issues of academic preparation and cultural capital, while also providing opportunities for peer support and shared authority. However, collaboration also presented challenges, particularly around uneven contribution quality and differing standards for evidence-based reasoning. Brielle's (Group 3, Grade 9) experience with a teammate's infographic illustrated this tension, as she had to address multiple issues that emerged during their collaborative review process. Using criteria developed through class discussions about evidence quality (claims must be supported by cited sources, statistics must include contextual information about sample and methodology and interpretive statements must distinguish between data patterns and personal opinions), Brielle identified work that lacked adequate source citation, contained factual inaccuracies and included what she termed "excessive bias."

Brielle explained:

> I had to go back and I'm still going back again, and changing a lot of stuff she [groupmate] had said, because the sentences [were] either wrong or it's like a bias or some sort […] hers was a lot more biased and opinionated [and] didn't have any facts.

By "excessive bias," Brielle referred to interpretive claims presented as facts without supporting evidence, such as broad generalizations about reproductive health care that were not substantiated by the data sources they had identified. This peer review process, while challenging, demonstrated how collaborative environments can foster the development of disciplinary standards for evidence evaluation when students are supported in articulating explicit criteria for data quality and interpretation.

The emergence of specialized roles reflects how learners negotiate authority and expertise within collaborative spaces, with some assuming leadership positions while others contribute specialized knowledge, revealing both opportunities for shared power and potential reproduction of traditional hierarchies. Learners developed role-based collaboration strategies, with clear specialization emerging organically. Riley (Group 5, Grade 8) explained: "I did mainly all the research, because I'm the researcher," while Tiana (Group 4, Grade 9) took on management responsibilities: "We used roles. But each of us have a different part of the group. I'm the Group Manager." This natural division of labor appeared to enhance both individual ownership and collective efficiency.

Critical data literacy through power-conscious analysis

This theme reveals how learners developed CDL by examining not only source reliability but also questioning whose voices and perspectives are represented in data, whose experiences are made visible or invisible and how data can be used to challenge or reinforce existing power structures. The critical examination demonstrated by learners reflects growing awareness of data as socially constructed rather than neutral or objective. Learners demonstrated an understanding of source evaluation and bias recognition, moving beyond simple fact-checking to nuanced analysis of information quality and reliability. This critical examination manifested in multiple dimensions: source verification, bias identification and evidence-based reasoning.

Source reliability emerged as a central concern, with learners developing practical heuristics for evaluation. Raven explained their approach: "For research, we use websites that are well known, so we know that they're unbiased," while Anna (Group 2, Grade 8) emphasized verification processes: "We had to make sure that they were verified and not

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

liars." These evaluation strategies reflect learners' growing understanding that data sources carry cultural and institutional authority, with some voices privileged over others in determining what counts as credible or trustworthy information. This attention to source credibility extended to domain-specific awareness, with multiple learners prioritizing government and organizational websites over commercial sources.

The development of evidence-based reasoning skills demonstrates how learners learned to interrogate not only the quality of evidence but also whose experiences and perspectives were centered or marginalized in data representation. Another example of critical thinking appeared in Brielle's (Group 3, Grade 9) analysis of teammate contributions, where she identified work that "wasn't based off any research" and articulated evidence standards: "You got to pick the actual research. It can't just say, oh, women are abused, and you gotta be like, the statistics and the facts." This comment reveals understanding of the distinction between opinion and evidence-based claims, a fundamental data literacy skill. Learners also demonstrated awareness of data limitations and contemporary relevance, with several explicitly seeking recent data to ensure the current applicability of their findings.

Culturally responsive communication and public engagement

This theme demonstrates how learners developed skills in translating complex data analysis into culturally relevant and accessible public discourse, while navigating tensions between academic rigor and community accessibility. The communication processes revealed both opportunities for amplifying marginalized voices and challenges in maintaining nuanced analysis when creating public-facing content. The integration of digital tools emerged as transformative for both analytical capabilities and communication effectiveness. Learners strategically leveraged multiple platforms, such as CODAP, Canva and Google Sheets, to enhance different aspects of their work, demonstrating an understanding of tool-specific strengths.

The strategic tool selection demonstrates how learners navigated multiple cultural contexts and communication expectations, adapting their approaches based on audience needs while maintaining their analytical integrity. The campers developed workflows that optimized different platforms for specific analytical and communication functions. Anna (Group 2, Grade 8) described her group's strategic approach:

> We used CODAP for our base graphs to get the statistical analysis right, then we reformatted everything in Canva for visual appeal because we knew our audience needed something more engaging than just raw output.

This workflow demonstrates understanding that analytical tools (CODAP) and communication tools (Canva) serve complementary but distinct purposes in the research process.

However, analysis of student products reveals that this tool-switching sometimes led to analytical shortcuts. Brielle's group (Group 3, Grade 9) on reproductive rights infographic (Figure 4), while visually compelling with its clean layout and strategic use of color-coded statistics (13%, 60%, 1%), appears to rely primarily on external data sources rather than their own CODAP analysis. Their final product showcases visual communication but provides limited evidence of the statistical exploration process that was central to the learning objectives.

The audience awareness demonstrated by learners reflects their understanding of how different communities access and interpret information, revealing sophisticated knowledge of cultural communication patterns and power dynamics in public discourse. Learners demonstrated growing awareness of audience needs in shaping their communication strategies. Brielle recognized that "people have like, short attention spans, especially when

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

Figure 4. Group 3's produced infographic

you're throwing numbers at them," leading her group to prioritize visual elements that provided a "clearer perspective" than presenting raw statistical results. This audience awareness influenced both tool selection and content organization, with groups gravitating toward infographic formats (as instructed) and simplified data presentations.

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

This tension reveals the challenges of democratizing data analysis while maintaining methodological transparency, highlighting how communication choices can either empower or obscure public understanding of complex social issues. The tension between accessibility and rigor became apparent in the groups' final products. While Group 3 successfully created public-facing content with clear calls to action ("Please vote locally in your county, city, and state for effective rights"), their statistical claims lack the contextual information necessary for evaluation. The prominently displayed "13% of those under the age of 18 of the sample group, were found to be pregnant at a early age from a study done in 2022 by the UNICEF" demonstrates this challenge – the statistic is presented for impact but without sufficient methodological context about sample size, geographic scope, or operational definitions.

Despite overall positive responses, technology integration presented significant challenges that affected both learning processes and outcomes. Several learners reported initial confusion with CODAP's interface, with Makayla (Group 9, Grade 8) noting, "At first I couldn't figure out how to get the variables to show up right, so I just kept clicking things until something worked," indicating that tool mastery sometimes became a barrier to statistical learning rather than a facilitator.

The evidence suggests that while digital tools successfully enhanced learners' ability to create engaging communications, the emphasis on visual appeal sometimes overshadowed the analytical process that was central to learning objectives. Groups that produced the most visually sophisticated final products (as evidenced in Figures 3 and 4) showed less evidence of iterative statistical exploration in their process documentation.

The pattern of critical analysis in this study deserves additional context, particularly regarding our examination of the reproductive rights infographic (Figure 4). This project received more detailed critique not due to topic sensitivity, but because it exemplified common challenges students faced when transitioning from data analysis to public communication. The group's infographic demonstrated sophisticated visual design and clear calls to action yet also illustrated the tension between accessibility and methodological transparency that emerged across multiple projects. Rather than representing a deficit in this particular group's work, this analysis highlights a systemic challenge in DSE: how to support students in maintaining analytical rigor while creating engaging public communications. This pattern was evident across projects but was most pronounced in topics where students felt strong personal connections, suggesting the need for additional scaffolding around evidence-based argumentation when students are working on issues of high personal relevance.

Identity development through reflective data practice

This theme demonstrates how learners developed both analytical competencies and personal agency through reflective engagement with data, connecting individual growth to broader social consciousness and community impact. The metacognitive development observed reflects how learners positioned themselves as capable agents of social change through data-informed action. Campers demonstrated metacognitive awareness, recognizing their own learning progression and developing appreciation for challenge-based growth. This theme encompassed both immediate skill development and broader personal transformation through engagement with complex, real-world problems including soil pollution, reproductive rights, child obesity, flood risk, food access, air pollution and wildfire data analysis.

The explicit recognition of developmental stages reveals how learners developed confidence in their analytical abilities while simultaneously recognizing their capacity to contribute meaningful perspectives to complex social problems. The progression from confusion to mastery was explicitly recognized by campers through their ability to articulate distinct developmental stages and emotional shifts. Raven (Group 1, Grade 8) reflected that

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

she was, initially "worried or stressed out" but realized "It was really simple" and noted that "it came out better than we thought." This demonstrates both emotional self-awareness and recognition of competence development over time. Tiana (Group 4, Grade 9) provided clear temporal markers for her learning journey: "Third day, we know what we're doing," while connecting skill development to personal impact: "It just makes me feel smarter."

The metacognitive awareness extended beyond individual recognition to understanding the benefits of collaborative learning. Anna (Group 2, Grade 8) articulated how teamwork enhanced error detection: "We can help fix mistakes that maybe the other person didn't catch," while Brielle (Group 3, Grade 9) recognized efficiency gains through collaboration: "Would have been like, three times longer if I had went and done all that same time by myself."

The reframing of challenge as productive rather than threatening demonstrates how learners developed resilience and confidence in tackling complex social problems, viewing difficulty as an opportunity for growth and deeper understanding rather than a barrier to participation. Campers developed an understanding of challenge as valuable for learning, reframing difficulty as productive rather than threatening. Tiana explicitly noted: "it's a good challenge," while Riley (Group 5, Grade 8) captured the productive tension inherent in meaningful learning: "allowed me to expand more on data science" while acknowledging the demanding nature of the work. This reframing of challenge as beneficial rather than problematic suggests internalization of growth-oriented learning principles.

Despite initial stress levels, with Alexis (Group 6, Grade 7) describing the experience as "very stressing" and multiple campers reporting confusion with project requirements, learners persisted and found value in the challenge. The progression from initial overwhelm to competence recognition demonstrates resilience and adaptive learning strategies.

The connection between individual skill development and community impact reveals how learners developed critical consciousness, recognizing their potential to use data analysis as a tool for social justice and community empowerment rather than simply academic achievement. Campers connected their analytical work to broader community impact. Tiana focused on prevention strategies that could have "scalability vision for broader impact," while Riley recognized that findings could be applied "worldwide instead of just one specific state." Alexis made direct community connections through "social media outreach for food desert awareness," demonstrating understanding that authentic research should lead to action.

Campers also demonstrated awareness of quality control processes. Anna described how "we all just started, like commenting on each other's parts so that we can help each other out," while Brielle actively revised work to meet evidence-based standards. This peer feedback system shows metacognitive awareness of collaborative learning strategies and quality assurance processes essential for authentic research work.

These five themes reveal how authentic data science experiences can support the development of both analytical competencies and critical consciousness when grounded in sociocultural learning principles and data feminist approaches that center equity, justice and community empowerment, where learners navigated data science concepts through collaborative, technology-enhanced investigation of personally meaningful social and environmental issues. While challenges around initial instruction clarity, tool mastery and collaborative quality control emerged, learners demonstrated remarkable resilience and growth, ultimately developing both technical skills and critical thinking capabilities essential for 21st-century civic engagement.

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

Discussion

The findings from this study highlight how learners develop CDL through authentic, collaborative engagement with real-world data sets, revealing the complex interplay between individual cognitive development and social learning processes (Bargagliotti *et al.*, 2020; Lee *et al.*, 2022). These results extend our understanding of DSE by demonstrating how situated learning experiences can foster both technical competencies and critical thinking capabilities essential for civic engagement in an increasingly data-driven society (Wilkerson and Polman, 2020).

Situated learning and authentic engagement

The progression from initial uncertainty to confidence observed across all camp participants aligns strongly with situated perspectives of learning's emphasis on authentic practice within meaningful contexts (Bang, 2015; Greeno, 2006). This finding builds upon decades of informal STEM learning research demonstrating that learning embedded in personally and socially relevant contexts enhances engagement and retention (Bell *et al.*, 2009; Falk and Dierking, 2010). Learners' engagement with environmental and social issues, from Raven's soil pollution research to Brielle's demographic analysis, demonstrates how authentic problems provide the contextual scaffolding necessary for meaningful learning. What extends beyond established informal STEM practices is how learners developed sophisticated epistemological reasoning about data quality and bias within these authentic contexts, moving beyond engagement to develop a critical awareness of how power structures and social contexts shape data collection, interpretation and use (D'Ignazio and Bhargava, 2020). The learners' persistence in overcoming data availability challenges, particularly Anna's continued search for current soil pollution data spanning "2020–2024," exemplifies what Brown *et al.* (1989) describe as the development of "robust knowledge" through situated practice. Rather than abandoning their inquiry when faced with obstacles, campers developed the research tenacity characteristic of authentic scientific practice. This resilience demonstrates how authentic contexts foster not just content learning but also the development of scientific dispositions and epistemological beliefs about the nature of inquiry (Calabrese Barton and Tan, 2020). Further, this suggests that situated learning contexts not only teach technical skills but also cultivate the dispositional aspects of CDL, such as the willingness to persist through challenges and maintain standards of evidence (Louie, 2022). Critically, learners' navigation of data scarcity revealed their emerging understanding of "data gaps," or the recognizing that absence of data often reflects systemic inequities rather than lack of problems worth studying (D'Ignazio and Bhargava, 2020).

Social construction of data literacy

The emergence of meaning-making through social interactions validates sociocultural perspectives of learning that position knowledge construction as fundamentally collaborative (Bang, 2015; Vygotsky, 1978). Building on extensive research in collaborative learning environments (Bell *et al.*, 2009; Rogoff *et al.*, 2016), our findings reveal how groups negotiated meaning not only among themselves but also with facilitators, creating multi-layered communities of practice that extended beyond peer-to-peer interaction. Learners' recognition that collaboration enables them to "bounce ideas off each other" and "help fix mistakes" demonstrates *communities of practice*, where learning occurs through participation in shared endeavors (Lave and Wenger, 1991).

Particularly significant is how learners naturally developed role-based collaboration strategies, with Riley identifying as "the researcher" and Tiana assuming "Group Manager" responsibilities. This organic specialization reflects legitimate peripheral participation

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

(Lave and Wenger, 1991), where learners gradually assumed expert roles within their collaborative groups. What represents a novel contribution is the emergence of spontaneous quality control mechanisms through peer review, as Raven described, "we all just started commenting on each other's parts." This development of epistemic standards through social interaction extends beyond typical collaborative learning outcomes, suggesting that authentic data science inquiry catalyzes the development of disciplinary practices for maintaining rigor and accuracy without explicit instruction.

However, the challenges around uneven contribution quality, evidenced by Brielle's need to "go back and change a lot of stuff," highlight the complex negotiation required in collaborative learning environments. This tension between efficiency and quality control suggests that while social interaction enhances learning, it requires careful structuring to maximize its benefits while minimizing its drawbacks (Rogoff *et al.*, 2016). To address this limitation during the study, facilitators implemented daily reflection protocols and peer feedback structures that supported learners in developing constructive critique skills while maintaining collaborative relationships.

Critical data literacy in practice

The campers' approach to source evaluation and bias recognition demonstrates the development of CDL (Louie, 2022). Through the lens of data feminism (D'Ignazio and Bhargava, 2020), learners' practical heuristics for source evaluation reveal sophisticated understanding of how power and authority shape data credibility. Learners' practical heuristics for source evaluation, such as Raven's preference for "websites that are well known" and Anna's emphasis on verification processes, reveal emerging understanding of data provenance and reliability. Rather than simply accepting traditional markers of authority, learners developed nuanced criteria that considered both institutional credibility and methodological transparency, reflecting the principle of "examining power" in data sources (D'Ignazio and Bhargava, 2020). Most notably, Brielle's articulation of evidence standards, "You got to pick the actual research. It can't just say, oh, women are abused, and you gotta be like, the statistics and the facts," demonstrates an understanding of the distinction between opinion and evidence-based claims. This finding extends Louie's (2022) framework of CDL by showing how learners can develop nuanced appreciation for evidence quality through authentic inquiry experiences (Lee *et al.*, 2022). This epistemological reasoning represents an advance beyond typical informal STEM outcomes, as learners developed not just facility with data analysis tools but critical consciousness about the social construction and potential misuse of statistical evidence.

The campers' attention to data recency and relevance further demonstrates CDL in action. Rather than accepting any available data, campers actively sought current information, recognizing that temporal relevance affects data utility. This behavior exemplifies the principle of "considering context" by recognizing that data's meaning and utility are inherently situated in time and social circumstances (D'Ignazio and Bhargava, 2020).

Reasoning with data through tool-mediated activity

The strategic use of digital tools (e.g. CODAP, Canva, Google Sheets) illustrates how reasoning with data emerges through "tool-mediated activity" (Vygotsky, 1978). Learners' recognition that technology provides both efficiency ("We don't have to take the extra time to draw graphs") and accuracy benefits demonstrates developing understanding of computational thinking principles.

Particularly sophisticated was Anna's workflow description: "Used CODAP for base graphs, reformatted in Canva for visual appeal," which reveals strategic tool selection based

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

on specific affordances. This behavior aligns with the argument that learners develop statistical reasoning through engagement with appropriate technological tools that make data patterns visible and manipulable (Wilkerson and Laina, 2018). What extends beyond existing research is how learners developed metacognitive awareness of how different tools shape both analytical processes and audience interpretation, recognizing that visualization choices carry implicit arguments about data's meaning and significance, reflecting the principle that "data visualization is never neutral" (D'Ignazio and Bhargava, 2020).

The initial confusion learners experienced with CODAP, followed by gradual mastery, is consistent with research on learning with new representational technologies (Barron *et al.*, 1998; NASEM, 2022). Rather than viewing these challenges as obstacles, the findings suggest they represent necessary phases in developing fluency with data science tools (NASEM, 2022). To support learners through this progression, facilitators provided structured scaffolding that normalized confusion as productive struggle while maintaining focus on analytical reasoning rather than technical proficiency.

Metacognitive development and identity formation

The remarkable metacognitive awareness demonstrated by campers – Raven's recognition of progression from "worried or stressed out" to "it was really simple," and Tiana's articulation of developmental stages. This self-awareness, documented through structured reflection protocols implemented throughout the camp, appears crucial for sustained engagement with challenging data science concepts and represents a key mechanism by which informal STEM experiences can foster long-term interest and persistence (Petrosino *et al.*, 2023).

Learners' developing identity as capable researchers and analysts, evidenced by Anna's reflection that the project "helped me learn more about soil pollution" and Tiana's feeling that "it just makes me feel smarter," suggests the formation of *figured worlds* or mental models of practice that include both technical competencies and identity positions (Urrieta, 2007). This identity development extends existing research on STEM identity formation by demonstrating how engagement with authentic data science problems can simultaneously develop technical competence and critical consciousness, fostering "data activism" or the capacity to use data literacy skills for social critique and change (D'Ignazio and Bhargava, 2020). This identity development may be particularly crucial for sustaining interest in STEM fields among diverse learner populations (Milton *et al.*, 2023).

Limitations

This study is not without its limitations. The relatively small sample size and the short duration of the camp may restrict the generalizability of the findings. To address sample size limitations, we used purposeful sampling to ensure diverse perspectives and used multiple data sources (interviews, artifacts) to triangulate findings and enhance trustworthiness. Additionally, the single-gender environment of the camp may also influence the outcomes. While this setting can foster a supportive community for participants, it may not reflect the dynamics present in mixed-gender educational contexts. Overall, these considerations highlight the need for further exploration in diverse settings and over extended periods to draw more comprehensive conclusions regarding the efficacy of such educational initiatives. This study focused exclusively on student perspectives and learning outcomes. Future research should examine facilitator practices and classroom interactions to provide a more complete picture of the learning ecosystem. While we collected rich data on student experiences and final products, direct observation of collaborative processes would strengthen understanding of peer learning dynamics.

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

Implications and conclusion

Based on our findings, several key implications emerge for DSE design. First, authentic inquiry contexts provide unique opportunities for developing CDL that extends beyond technical skills to include enhanced epistemological reasoning about data's social construction and potential for both insight and harm. Second, collaborative learning environments require explicit scaffolding around evidence evaluation and peer review processes, as learners may lack the skills to effectively critique each other's work without structured guidance on research standards. Third, the sequencing of digital tools needs careful consideration to maintain focus on analytical processes rather than just communication outputs – educators may need to require documentation of statistical exploration before students move to visualization and presentation platforms. For example, tools such as Canva, while engaging, can shift student attention away from reasoning with the data if introduced before analytical work is documented. Fourth, instructors should prepare students for and normalize the progression from initial confusion to competence, reframing productive struggle as an essential component of learning rather than a sign of inadequate instruction. Fifth, using authentic, community-relevant problems rather than abstract data sets appears crucial for sustaining motivation through the challenging analytical work required for meaningful DSE. Sixth, integrating data feminism principles throughout curriculum design can help learners develop critical consciousness about data's role in perpetuating or challenging social inequities. Finally, building structured metacognitive reflection opportunities into the curriculum can help learners recognize their skill development and transfer strategies to new contexts, amplifying the learning that occurs through collaborative investigation of real-world issues.

This study demonstrates that when learners engage with personally meaningful data science problems through collaborative, technology-enhanced inquiry, they develop both technical competencies and critical thinking capabilities essential for civic participation in a data-driven society. The findings advance informal STEM learning research by revealing how authentic data science inquiry can simultaneously develop technical competence, critical consciousness and collaborative practices essential for democratic participation in increasingly data-driven societies. The findings support situated and sociocultural perspectives of learning while extending our understanding of how CDL develops through authentic practice (Bang, 2015; Greeno, 2006; Bell *et al.*, 2009; Falk and Dierking, 2010; Wilkerson and Polman, 2020).

By demonstrating how learners developed sophisticated epistemological reasoning about data quality, bias and social construction within collaborative inquiry contexts, this study extends beyond existing informal STEM best practices to reveal new possibilities for fostering critical data citizenship. These insights suggest promising directions for DSE that prioritize meaningful engagement over technical skill acquisition alone, recognizing that learning occurs through participation in authentic communities of practice (Milton *et al.*, 2023; Lave and Wenger, 1991). Most significantly, this research demonstrates how integrating data feminism principles within informal STEM learning environments can foster both analytical competence and critical consciousness, preparing young people not just to consume and produce data, but to interrogate its social implications and use it as a tool for social justice. Future research should examine how these approaches scale across diverse educational contexts and how sustained engagement with CDL practices influences long-term civic participation and social action outcomes.

Author contribution

Marc T. Sager and Saki L. Milton are shared first author.

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

References

Bang, M. (2015), "Culture, learning, and development and the natural world: the influences of situative perspectives", *Educational Psychologist*, Vol. 50 No. 3, pp. 220-233, doi: 10.1080/00461520.2015.1075402.

Bang, M. and Medin, D. (2010), "Cultural processes in science education: supporting the navigation of multiple epistemologies", *Science Education*, Vol. 94 No. 6, pp. 1008-1026, doi: 10.1002/sce.20392.

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L. and Spangler, D.A. (2020), *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II)*, (2nd ed.), American Statistical Association, Alexandria, VA.

Barron, B.J.S., Schwartz, D.L., Vye, N.J., Moore, A., Petrosino, A., Zech, L. and Bransford, J.D. and The Cognition and Technology Group at Vanderbilt (1998), "Doing with understanding: lessons from research on problem- and Project-Based learning", *Journal of the Learning Sciences*, Vol. 7 No. 3, pp. 271-311, doi: 10.1207/s15327809jls0703&4_2.

Bell, P., Lewenstein, B. V., Shouse, A. W. and Feder, M. A. (Eds), (2009), *Learning Science in Informal Environments: People, Places, and Pursuits*, National Academies Press, Washington, DC.

Benjamin, R. (2019), *Race after Technology: Abolitionist Tools for the New Jim Code*, Polity Press, Cambridge.

Bowler, L., Acker, A. and Chi, Y. (2019), "Perspectives on youth data literacy at the public library: teen services staff speak Out", *Journal of Research on Libraries and Young Adults*, Vol. 10 No. 2.

Broussard, M. (2018), *Artificial Unintelligence: How Computers Misunderstand the World*, MIT Press, Cambridge, MA.

Browne, S. (2015), *Dark Matters: On the Surveillance of Blackness*, Duke University Press, Durham, NC.

Brown, J.S., Collins, A. and Duguid, P. (1989), "Situated cognition and the culture of learning", *Educational Researcher*, Vol. 18 No. 1, pp. 32-42, doi: 10.3102/0013189X018001032.

Calabrese Barton, A. and Tan, E. (2020), "Beyond equity as inclusion: a framework of "rightful presence" for guiding justice-oriented studies in teaching and learning", *Educational Researcher*, Vol. 49 No. 6, pp. 433-440.

Calabrese Barton, A., Greenberg, D., Turner, C., Riter, D., Perez, M., Tasker, T., Jones, D., Herrenkohl, L.R. and Davis, E.A. (2021), "Youth critical data practices in the COVID-19 multipandemic", *AERA Open*, Vol. 7, doi: 10.1177/23328584211041631.

Creswell, J.W. and Guetterman, T.C. (2019), *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, (6th ed.), Pearson, New York, NY.

Criado-Perez, C. (2019), *Invisible Women: Data Bias in a World Designed for Men*, Abrams Press, New York, NY.

D'Ignazio, C. and Bhargava, R. (2020), "Data visualization literacy: a feminist starting point", In Engebretsen, M. and Kennedy, H. (Eds), *Data Visualization in Society*, Amsterdam University Press, Amsterdam, NL, pp. 207-222.

D'Ignazio, C. and Klein, L.F. (2020), *Data Feminism*, MIT Press, Cambridge, MA.

Esmonde, I. and Booker, A. N. (Eds.). (2017), *Power and Privilege in the Learning Sciences: Critical and Sociocultural Theories of Learning*, (1st ed.), Routledge, New York, NY.

Falk, J.H. and Dierking, L.D. (2010), *The Museum Experience Revisited*, Left Coast Press, Walnut Creek, CA.

Fielding, J., Makar, K. and Ben-Zvi, D. (2025), "Developing students' reasoning with data and data-ing", *ZDM – Mathematics Education*, Vol. 57 No. 1, pp. 1-18, doi: 10.1007/s11858-025-01671-6.

Freire, P. (1970), *Pedagogy of the Oppressed*, In Ramos, M.B., Trans., Herder and Herder, Freiburg im Breisgau, DE.

Greeno, J. (2006), "Learning in activity", In R. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences*, Cambridge University Press, Cambridge, pp. 79-96.

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

Hautea, S., Dasgupta, S. and Hill, B.M. (2017), “Youth perspectives on critical data literacies”, *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI ‘17)*. *Association for Computing Machinery, New York, NY*, pp. 919-930.

Hooks, B. (1994), *Teaching to Transgress: Education as the Practice of Freedom*, Routledge, New York, NY.

Ireland, D.T., Freeman, K.E., Winston-Proctor, C.E., DeLaine, K.D., McDonald Lowe, S. and Woodson, K.M. (2018), (“Un) hidden figures: a synthesis of research examining the intersectional experiences of black women and girls in STEM education”, *Review of Research in Education*, Vol. 42 No. 1, pp. 226-254.

Kanari, Z. and Millar, R. (2004), “Reasoning from data: how students collect and interpret data in science investigations”, *Journal of Research in Science Teaching*, Vol. 41 No. 7, pp. 748-769, doi: 10.1002/tea.20020.

Kazak, S., Fujita, T. and Turmo, M.P. (2023), “Students’ informal statistical inferences through data modeling with a large multivariate dataset”, *Mathematical Thinking and Learning*, Vol. 25 No. 1, pp. 23-43, doi: 10.1080/10986065.2021.1922857.

Lave, J. and Wenger, E. (1991), *Situated Learning: Legitimate Peripheral Participation*, Cambridge University Press, Cambridge.

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.

Lee, H.S., Mojica, G.F., Thrasher, E.P. and Baumgartner, P. (2022), “Investigating data like a data scientist: key practices and processes”, *Statistics Education Research Journal*, Vol. 21 No. 2, p. 3.

Letourneau, S.M., Liu, C.J., Donnelly, K., Meza, D., Uzzo, S. and Culp, K.M. (2020), “Museum makers: family explorations of data science through making and exhibit design”, *Curator: The Museum Journal*, Vol. 63 No. 1, pp. 131-145.

Louie, J. (2022), *Critical Data Literacy: Creating a More Just World with Data*, National Academies of Sciences, Engineering, and Medicine, Washington, DC.

McClain, K. and Cobb, P. (2001), “Supporting students’ ability to reason about data”, *Educational Studies in Mathematics*, Vol. 45 Nos 1-3, pp. 103-129, doi: 10.1023/A:1013874514650.

Miles, M.B., Huberman, M.A. and Saldana, J. (2020), “Qualitative data analysis”, 4th ed., Sage Publications.

Milner, H.R. (2007), “Race, culture, and researcher positionality: working through dangers seen, unseen, and unforeseen”, *Educational Researcher*, Vol. 36 No. 7, pp. 388-400, doi: 10.3102/0013189X07309471.

Milton, S., Sager, M.T. and Walkington, C. (2023), “Understanding racially minoritized girls’ perceptions of their STEM identities, abilities, and sense of belonging in a summer camp”, *Education Sciences*, Vol. 13 No. 12, p. 1183, doi: 10.3390/educsci13121183.

Milton, S., Sager, M.T. and Walkington, C. (2025), “*Incorporating critical data literacy into a STEM summer camp*”, *Proceedings of the 19th International Conference of the Learning Sciences – ICLS 2025*, International Society of the Learning Sciences: Helsinki, Finland.

National Academies of Sciences, Engineering, and Medicine (NASEM) (2022), *Foundations of Data Science for Students in Grades K-12*, National Academies Press, Washington, DC.

Noble, S.U. (2018), *Algorithms of Oppression: How Search Engines Reinforce Racism*, NYU Press, New York, NY.

Ong, M., Wright, C., Espinosa, L. and Orfield, G. (2011), “Inside the double bind: a synthesis of research on undergraduate and graduate women of color in STEM”, *Harvard Educational Review*, Vol. 81 No. 2, pp. 172-209, doi: 10.17763/haer.81.2.t022245n7x4752v2.

Pangrazio, L. and Selwyn, N. (2023), *Critical Data Literacies: Rethinking Data and Everyday Life*, MIT Press, Cambridge, MA.

Patton, M.Q. (2015), *Qualitative Research and Evaluation Methods: Integrating Theory and Practice*, (4th ed.), SAGE Publications, Thousand Oaks, CA.

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026

Petrosino, A.J., Walkington, C. and Ekberg, D. (2023), *Frameworks for Integrated Project-Based Instruction in STEM Disciplines*, Information Age Press, Charlotte, NC.

Philip, T.M. and Azevedo, F.S. (2017), "Everyday science learning and equity: mapping the contested terrain", *Science Education*, Vol. 101 No. 4, pp. 526-532, doi: 10.1002/sce.21286.

Philip, T.M., Olivares-Pasillas, M.C. and Rocha, J. (2016), "Becoming racially literate about data and Data-Literate about race: data visualizations in the classroom as a site of Racial-Ideological Micro-Contestations", *Cognition and Instruction*, Vol. 34 No. 4, pp. 361-388, doi: 10.1080/07370008.2016.1210418.

Rogoff, B. (2003), *The Cultural Nature of Human Development*, Oxford University Press, Oxford.

Rogoff, B., Callanan, M., Gutiérrez, K.D. and Erickson, F. (2016), "The organization of informal learning", *Review of Research in Education*, Vol. 40 No. 1, pp. 356-401, doi: 10.3102/0091732X16680994.

Rosenberg, J.M. (2018), *Understanding Work with Data in Summer STEM Programs through an Experience Sampling Method Approach*, Michigan State University, East Lansing, MI.

Sager, M.T., Miller, S. and Drozda, Z. (2026), "Data science education in informal learning environments: a review of the literature", *Journal of Statistics and Data Science Education*, pp. 1-28, doi: 10.1080/26939169.2026.2618201.

Sager, M.T., Wieselmann, J.R. and Petrosino, A.J. (2025), "From classroom to community: evaluating data science practices in education and social justice", *Education Sciences*, Vol. 15 No. 7, p. 878, doi: 10.3390/educsci15070878.

Stake, R.E. (1995), *The Art of Case Study Research*, SAGE Publications, Thousand Oaks, CA.

Stornaiuolo, A. (2020), "Authoring data stories in a media makerspace: adolescents developing critical data literacies", *Journal of the Learning Sciences*, Vol. 29 No. 1, pp. 81-103, doi: 10.1080/10508406.2019.1689365.

Tschida, C.M., Ryan, C.L. and Ticknor, A.S. (2014), "Building on windows and mirrors: encouraging the disruption of "single stories" through children's literature", *Journal of Children's Literature*, Vol. 40 No. 1, pp. 28-39.

The Concord Consortium (2026), "Common online data analysis platform (CODAP)", available at: https://codap.concord.org/

Urrieta, L. (2007), "Figured worlds and education: an introduction to the special issue", *The Urban Review*, Vol. 39 No. 2, pp. 107-116, doi: 10.1007/s11256-007-0051-0.

Vygotsky, L.S. (1978), *Mind in Society: The Development of Higher Psychological Processes*, Harvard University Press, Cambridge, MA.

Wasser, J.D. and Bresler, L. (1996), "Working in the interpretive zone: conceptualizing collaboration in qualitative research teams", *Educational Researcher*, Vol. 25 No. 5, pp. 5-15.

Wilkerson, M.H. and Polman, J.L. (2020), "Situating data science: exploring how relationships to data shape learning", *Journal of the Learning Sciences*, Vol. 29 No. 1, pp. 1-10.

Wilkerson, M.H. and Laina, V. (2018), "Middle school students' reasoning about data and context through storytelling with repurposed local data", *ZDM*, Vol. 50 No. 7, pp. 1223-1235, doi: 10.1007/s11858-018-0974-9.

Corresponding authors
Marc T. Sager can be contacted at: msager@smu.edu; Saki L. Milton can be contacted at: slmilton@smu.edu

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

Downloaded from ftp.nowpublishers.com/ils/article-pdf/127/5-6/315/11362887/ils-07-2025-0102en.pdf by guest on 14 August 2026