Constructionist approaches to critical data literacy: A review

Aayushi Dangol
Sayamindu Dasgupta
adango@uw.edu
sdg1@uw.edu
University of Washington
Seattle, Washington, USA

ABSTRACT

Increased technological capacity to collect and use data has created both new possibilities for benefiting individuals and societies, and critical questions of what is acceptable and just [31]. Because early definitions of data literacy have often excluded aspects of power, equity, empowerment, and emancipation, children's learning experiences have focused more on the potential benefits compared to the critical questions. In this review article, we examine the importance of teaching critical data literacy to children as a key aspect of developing fluency with data. Using constructionist principles [67] as a guiding framework, we synthesize 48 educational research and design approaches that engage youth with data projects. We describe how these projects provide students with information about data's origins and perspectives, and assist them in identifying, analyzing, and presenting data. Finally, we provide design implications and concrete examples on how constructionist approaches can be utilized for teaching critical data literacy.

CCS CONCEPTS

• **Social and professional topics** → **Computing literacy**; • **Security and privacy** → *Human and societal aspects of security and privacy*; • **Human-centered computing** → Interaction design theory, concepts and paradigms.

KEYWORDS

critical data literacy, data literacy, constructionism, meaningful inquiry

ACM Reference Format:
Aayushi Dangol and Sayamindu Dasgupta. 2023. Constructionist approaches to critical data literacy: A review. In *Interaction Design and Children (IDC '23), June 19–23, 2023, Chicago, IL, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3585088.3589367

1 INTRODUCTION

With nearly half of young people reporting being online "almost constantly" [3], their personal data collected through online activities [65] has engendered excitement about tracking and interpreting their data in unprecedented ways. Additionally, the emerging field of data science [6] has also altered the way educators and administrators use data for educational decisions [70]. The combination of these factors has multiple implications for youth. Recognizing the role contemporary youth play in navigating a "datafied" culture [59] there is increasing interest in how to better prepare students in K-12 and undergraduate levels to work with data [33, 55]. The result is a growing number of emerging standards and journal special issues focusing on data science education across different fields [34, 73, 91]. Scholars have suggested that young people need to learn multiple ways of generating, interpreting, and visualizing data [55], especially personal data that individuals generate through everyday activities [65].

All of these arguments are centered around developing data literacy skills: students should be educated to "read, work with, analyze and argue with data" [8]. However, within this paradigm, learners are rarely taught to question what data is used for or how it is shaping society. For example, Code.org's "CS Discoveries" series provides a broad range of computational topics. However, the focus is on learning programming concepts as it only briefly touches on questions about data and society [48]. Furthermore, these arguments do not address the ways in which data can be used to reinforce biases and power structures [48], and the impact that data-driven processes can have on people's lives [41]. Data is not just a neutral recording of facts, but rather it is encoded with specific meanings and stripped of details, contexts, nuances, and a multitude of social, cultural, and psychological factors [48].

More critical perspectives within data literacy (which we term as *critical data literacy*) are needed in the learning processes [41] that can enable learners to critically perceive "the way they exist in the world with which and in which they find themselves" [36, p. 12]. An early advocate of critical questioning in learning contexts was Paulo Freire, an educator and author of *Pedagogy of the Oppressed*, a pioneering text on critical pedagogy. Our work contributes to the emerging literature on critical data literacy with reference to Freire's work [36]. Tygel & Kirsch [85] first used the term to draw parallels between Freirean approaches to literacy education and data literacy models [22]. D'Ignazio & Bhargava [29] also build on Freire's ideas when suggesting approaches to big data literacy. Several scholars have begun exploring innovative ways to teach critical data literacy. One effective strategy for developing critical data literacy in youth is the use of arts-based representational techniques and media. These methods empower learners to interact with and understand the data that is relevant to their personal experiences [7]. This approach not only helps youth learn with data but also encourages them to take a more active role in creating and shaping it [8].

IDC '23, June 19–23, 2023, Chicago, IL, USA

ACM ISBN 979-8-4007-0131-3/23/06.
https://doi.org/10.1145/3585088.3589367

At present, we recognize that there is an emerging need to teach critical data literacy to youth for navigating an increasingly data-driven world. We provide a review of recent and current K-12 and undergraduate data science educational approaches in formal and informal contexts through the lens of constructionism. Taking a constructionist approach may be effective as it seeks to understand youths' conceptions of data and their limitations by examining how they construct meaningful data artifacts in a social context [41]. We translate the three principles of constructionism—the continuity principle, the power principle, and the principle of cultural resonance—for future constructionist integration with youth data science education. Additionally, we also explore the potential benefits of using a constructionist approach to teach critical data literacy and investigate whether this approach can impact desired educational outcomes. Therefore, our paper addresses the following research questions: *In learning experiences that foster critical literacy, what can constructionist pedagogy afford for learners? What are the design approaches and strategies used in prior empirical work on critical data literacy that can be framed in terms of existing constructionist theory and how can they inform future design, practice, and scholarship?*

Our focus is on projects that center perspectives of young people, highlighting the connections between these projects and the three principles of constructionism. Constructionist ideas have guided the evolution of a broad range of learning technologies, especially those around computation and making [45]. By situating constructionism in learning environments, learners shift from passive and declarative knowledge to interactive and practical knowledge [57]. However, to the best of our knowledge, apart from individual studies in specific contexts (e.g., [22] that focuses on the Scratch programming language and online community), constructionist pedagogy and design approaches have not been systematically investigated in terms of its role in fostering critical perspectives about data. Therefore, there are two levels of contributions of this work. One lies in the reflection of a broader goal shared by the educational technology community: to support students in sharing stories, expressing ideas, and arguing for change with data, while assuming a critical perspective. We contribute to this broader goal in that our work synthesizes the findings from a broad range of empirical settings and pedagogical approaches toward a general understanding of what fostering critical data literacy looks like. The second contribution is a set of design implications, along with concrete examples that might be more generally applicable and informative when designing learning environments to foster critical data literacy.

In the following sections, we first discuss related work, and then provide background on critical data literacy and the principles of constructionism. Next, we outline the methods used to conduct our review, followed by our findings. Finally, we offer suggestions for future research and work in this field.

2 RELATED WORK

Several prior reviews have focused on developing and evaluating strategies for promoting critical data literacy in both formal and informal K-12 education environments. Most recently, Louie [58] examined eight K-12 learning interventions that aimed to promote critical data literacy using Gutstein's model of social justice pedagogy [39] and Franklin and Bargagliotti's model of the data inquiry cycle [34]. Lee et al. [54] conducted a systematic review of data science interventions in K-12 education using the data feminism framework described by D'Ignazio and Klein [25]. Additionally, Lee, Wilkerson, and colleagues [50] examined data education through a humanistic lens, considering the personal, cultural, and sociopolitical layers of working with data. Wilkerson & Polman [91] published a special issue exploring data science interventions focusing on the social and environmental contexts of data collection and analysis. Lee & Wilkerson [55] have also conducted a systematic database search on data use by middle and secondary students in science classroom practice. We build on these works while connecting to a framework (viz. constructionism) that has been often used in the design of creative learning tools and experiences that are adjacent (e.g., learning to code).

3 BACKGROUND

Our work is informed by previous research on critical pedagogy, data literacy, and constructionist learning. In this section, we define the key theoretical ideas that form the pillars of this prior research.

3.1 Critical Data Literacy

Critical data literacy encompasses understanding the origins of the data and the perspectives of those who handle it, as well as the ability to identify, analyze, and present data in a way that supports new narratives or clarifies existing ones [44]. Developing a clarified narrative can include students using data to examine and challenge dominant narratives in society, or to create their own counter-narratives that focus on issues they are passionate about [48]. Critical data literacy can be further explained using Tygel & Kirsh [85] and Philip et al. [69] frameworks. Grounded in Paulo Freire's literary method [36], Tygel & Kirsh [85] examine data's ability to empower by breaking critical data literacy down into four parts: data reading, data processing, data communication, and data production. Here, reading refers to understanding how data was generated and its context, processing refers to the transformation of data into information, communication refers to finding the best match between data types to communicate information effectively, and production refers to the ability to create and publish data in an ethical and accessible manner [85].

Another effort in developing a theory of critical data literacy for youth is the data framework for democratic participation proposed by Philip et al. [69]. By grounding their framework within sociocultural theories of learning [64, 74], critical pedagogy perspectives [36, 37], as well as Delpit's notion of the "culture of power" [23], Philip and colleagues examine big data as a means to understand equity and social justice [69]. They argue that for students to engage in data work to transform society, youth need to "see themselves as doers and creators of data science, people who can engage with and use data for their own purposes and goals" [69, p. 114-115].

In line with the work of Philip et al. [69], a number of studies have focused on advancing critical data literacy through an equity and justice lens. For example, to make systemic injustices visible in the classroom, Calabrese Barton and Tan [15] highlight the importance of reframing learning as a joint effort for rights, involving students in critical discussions about political injustices, and challenging established power dynamics. D'Ignazio and Klein [25] focus on rethinking of justice and equity through the lens of Data Feminism, advocating the inclusion of dimensions typically excluded from data science, such as care, emotion, relationality, and context [54]. Johnson et al. [44] leverage principles of critical race theory to develop a community of practice centered around work with data, and engage in conversation around data literacy, race, racism, and equity. As these examples show, it is vital that we, educators and designers, approach data critically and immerse ourselves to include aspects of power, equity, and emancipation while teaching data literacy. Consequently, if we do not emphasize critical perspectives while teaching data literacy, there is an imminent risk of amplifying the framing of data-driven and data-mediated approaches as an overwhelmingly positive force for society, while de-emphasizing the harms that these approaches can bring, especially to marginalized communities.

The objective of our paper is therefore to review ways in which critical data literacy can be fostered in youth so they can navigate a world increasingly driven and mediated by data, and to draw from Eubanks [31, p. 131], recognize not only the new possibilities afforded by data, but also question whether these possibilities are acceptable and just within our broader social structures and values.

3.2 Constructionism in fostering Critical Data Literacy

The phrase "to understand is to invent" is attributed to the constructivist educator Jean Piaget. Seymour Papert's constructionism builds on this viewpoint of learners as active builders of knowledge. At the same time, it also emphasizes the importance of learning by making, where learners interact socially to share their thoughts and understandings. In Papert's own words: "Constructionism [. . .] shares constructivism's view of learning as 'building knowledge structures' through progressive internalization of actions." [66, p. 1].

Our foundational approaches are derived from combining theories of constructionism [66] to foster critical data literacy. We take this approach because constructionist principles ensure that critical perspectives are incorporated into learning settings without displacing the focus from developing data literacy skills [24]. Additionally, because constructionism offers learners support for multiple pathways of learning [83], novices and learners who self-identify as "non-technical" can greatly benefit from it [24]. Connecting constructionism with feminist theories also supports the idea of working with personally meaningful and "less detached" forms of knowledge to empower new learners to argue for change with data [83]. For the scope of this paper, the three principles we draw from constructionism are the *continuity principle*, the *power principle*, and *the principle of cultural resonance* [67]. In this section, we define and explain each principle, as well as relate the principles more specifically to the domain of critical data literacy.

3.2.1 Continuity Principle. The continuity principle argues that activities should be connected to some "well-established personal knowledge" through which the participants can inherit "warmth, value, and cognitive competence" [67, p. 54]. Learners have choices in deciding what types of questions can be answered with data and determining whether the data they need has been collected in the first place [24]. The degree of freedom [89] is left to students for exploration. Facilitators help students experiment with ideas, and learners are provided with opportunities to participate in real-world activities that are relevant to them. This approach embraces the view that learning is not minimally guided but optimally guided [81]. In line with feminist approaches, the learning environment is grounded in meaningful contexts [13]. As a result, when learners bring their own context and lived experience into the classroom, they are equipped to ask better questions, uncover missing or bad data, reflect on the data's limitations, and perhaps even challenge the data collection practices [24].

3.2.2 Power Principle. The power principle states that one must engage in work that is personally meaningful and that it cannot be accomplished through other available means [67]. This principle sees knowledge as a source of "personal power"—e.g., the computational concept of a variable being useful for drawing a spiral pattern on the screen, as wished by the learner [67, p. 74]. Based on this principle, students invent methods to describe and explore patterns in data [76], and create new mediums of expression through data that reflect what they know, personally value, and relate to. The ability to create a personally meaningful project also helps build confidence in nontechnical learners and creates a low barrier to entry for engaging in data literacy projects. The artifacts created in this process function as "objects to think with" that allow students to simultaneously visualize, reflect on, and test their data analysis process with others [24].

3.2.3 Principle of cultural resonance. The principle of cultural resonance asserts that the activity must "make sense [within] a larger social context" [67, p. 54], and once students construct the knowledge for themselves, it is important to share with others and receive critique [66]. A concrete style of reasoning results from the ability to manipulate and refine these objects repeatedly [83]. Moreover, shared knowledge is constructed when artifacts are shared in collaborative learning environments [1]. This principle emphasizes the importance of learners actively engaging with the data analysis and visualization process by reusing and remixing code in real-time. It also encourages the inclusion of diverse perspectives during the data collection stage by seeking input from stakeholders [75], and fostering reflection on the role of biases within the context of the individual learner, through critique sessions and other methods. According to Papert, this is an iterative and cumulative learning process that incorporates planning and bricolage [83]. Additionally, this principle foregrounds the envisioning of a better world through technological fluency and interpretation [1]. By relating what they have done to the broader context of how data operates within discourses of power, privilege, and societal inequities, learners can embody and crystallize thought into action.

4 RESEARCH METHODOLOGY

4.1 Procedure

The semi-systematic review approach that we have employed in this study is rooted in the literature discussed by Snyder [79] and is built on the ideas of Wong et al. [92]. The aim of this type of review, as Snyder states, is to "identify and understand all potentially relevant research traditions that have implications for the studied topic and to synthesize these using meta-narratives instead of by measuring effect size" [79, p. 335].

Given the above points, a semi-systematic review approach was chosen for our study mainly due to three reasons: Firstly, the emerging field of critical data literacy is directly connected with several research areas (e.g., science education, statistics education, mathematics education, learning sciences, and computer science). Secondly, we recognize that every author may not use the term "critical data literacy" to describe projects that develop critical perspectives on data. For example, Shapiro [78] teaches students about the ethical implications of data collection and use as part of a data ethics class, while D'Ignazio and Klein [25] center their work on rethinking justice and equity in data science through the lens of Data Feminism. Thirdly, we acknowledge that we do not cover all venues where critical data literacy research is published. This follows a standard practice for semi-systematic literature reviews in that we do not claim to cover all critical data literacy projects, but rather a representative sample of such projects that take a constructionist approach, using a criteria that is explained in the following section.

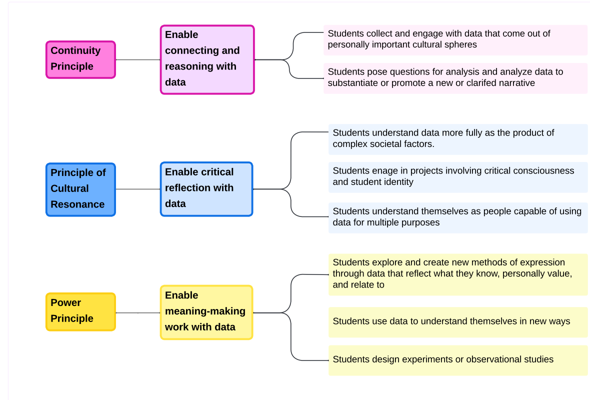

Figure 1: A hierarchy showing how the three constructionist principles relate to the papers that we analyzed for this study

4.2 Paper Selection

To build our corpus, we started by identifying common themes and patterns across critical data literacy and constructionism that would form the foundation of our analysis. The first step in this process involved consulting two authoritative sources on data literacy, as listed in Table 1, where the authors of these sources aimed to be as comprehensive as possible in their scope. We selected these sources based on the relative recency of the papers and the established expertise of the authors. After going through the two authoritative sources, the lead author developed a table that mapped the three constructionist principles, as expressed by Papert [67], to constructionist learning approaches which was further mapped to various skills fostered in critical data literacy. The author team met over a period of two months to iteratively refine an initial set of 11 codes related to the connections between constructionist pedagogy and critical data literacy. Through this process, we arrived at a final set of eight codes on which we reached consensus and resolved any disagreements. To organize our code book and make it easily accessible, we then created a chart (Figure 1) that demonstrates the procedure we used to identify constructionist-based approaches to teach critical data literacy.

From there the corpus was expanded by reviewing publications cited within those authoritative sources. Under this approach, we required at least one published document that reviewers and readers could refer to at the time of writing. As a first step, we determined whether a cited publication would be included based on the relevance of the title. When collecting papers, at this stage, we sought projects where students interacted with and made sense of data. This could include projects that integrated issues pertaining to the collection, analysis, interpretation, representation, visualization, and communication of data. This was done iteratively as the list of publications expanded. Papers whose titles were ambiguous or likely to be relevant were also included in this stage.

After building the initial corpus, we refined it by reading the abstracts. While screening the abstracts, we only included projects that integrated explicit attention to issues pertaining to critical data literacy, such as teaching methodologies, design studies, and analyses of computational tools. Projects that support the development of related concepts like statistical reasoning or computational thinking, but do not engage in building critical data literacy, were excluded. For example, papers that focused on teaching statistics with spreadsheet software or articles that discussed general programming education were not considered (e.g., [43]). As a third step, the remaining articles were screened more extensively by reading (parts of) the papers and following it with eligibility checks using the inclusion and exclusion criteria (Table 2).

Since our paper aims to provide an overview of recent and current data science educational approaches from a constructionist perspective, our final selection included papers that enabled connection to data, critical reflection with data, and meaning-making work with data. The publications didn't have to relate to all three of the parameters listed in Figure 1. We have included works that explicitly focus on, or discuss either concrete artifacts for critical data literacy projects, evaluate a solution for building critical data literacy or design a study to explore critical data literacy. Further, we have chosen not to include thesis work, panel announcements and work-in-progress papers.

During the paper selection process, we had to consider the differing perspectives on what constitutes a constructionist activity. While some scholars point out the importance of minimally guided discovery, others argue that some guidance is needed and that constructionism is all about the learner constructing their own understanding through hands-on exploration [5]. To address this, we defined constructionist learning experiences as those that include (i) student-centered discovery, (ii) a learning environment rooted in meaningful contexts, (iii) peer feedback, reflection, and iterative refinement, and (iii) envisioning a better world through technology fluency and interpretation [1]. Similarly, in the course of our review, we had to consider the question of whether a physical artifact was necessary for an activity to be considered constructionist. For example, if students are learning about plant growth at the population level and evaluating data, is it a constructionist project?

Title	Publication venue	Num. of papers identified
A Call for a Humanistic Stance Toward K–12 Data Science Education [56]	Educational Researcher	31
Taking Data Feminism to school: A synthesis and review of pre-collegiate data science education projects [54]	British Journal of Educational Technology	43

Table 1: Overview of the starting corpus with the initial number of papers identified for further screening

Inclusion Criteria	Exclusion Criteria
Articles whose subjects were K-12 or undergraduate students in formal or informal learning environments.	Articles whose participants were professional students enrolled in skill-focused career and technical education courses.
Articles either presenting concrete artifacts for fostering critical data literacy, evaluating strategies for developing critical data literacy, or designing new approaches to explore critical data literacy.	Articles around related concepts like statistical reasoning or computational thinking that did not explicitly aim to research data literacy and/or data science education and/or critical data literacy or articles that discussed general programming education.
Articles on projects enabling connection to, critical reflection on, or meaning-making work with data.	

Table 2: Inclusion and exclusion criteria that we used in the third step of our paper selection process

According to Papert, constructionism involves "building knowledge structures" and adds the idea that "this happens especially felicitously in a context where the learner is consciously engaged in constructing a public entity, whether it's a sand castle on the beach or a theory of the universe." [66, p. 2]. Therefore, for our purposes, we decided to include a project if it supported learners' conceptual understanding without requiring a physical artifact to be constructed. Additionally, we chose to focus on literature whose participants were K-12 and undergraduate students in both formal and informal learning environments, as these programs tend to be grounded within a context of inquiry [44] and critical thinking, which is well-aligned well with the scope of our paper.

This process yielded 119 papers, which were filtered to 90 with the removal of duplicates; 76 papers were screened, with 14 excluded by title and abstract, and a further 14 papers excluded following review of full text after a round of inclusion and exclusion coding, leaving a final total of 48 papers that fit the inclusion criteria.

5 FINDINGS

In this section we provide an in-depth look at how the three principles of constructionism, as described in §3.2, emerge within the designs and practices of critical data literacy instruction.

5.1 Continuity Principle

The continuity principle within critical data literacy connects learning to "well-established personal knowledge" through which the participants can inherit "warmth, value, and cognitive competence" [67, p. 54]. As part of this principle, learners learn to critically reflect with data by taking an active role in collecting, analyzing and investigating data that are personally meaningful to them.

5.1.1 Students collect and engage with data that come out of personally important cultural spheres. A personal connection with data facilitates learners' ability to think critically when collecting and analyzing data, allowing them to determine what data is needed and how to transform that data into meaningful information. For example, in Clegg et al.'s study [19], Division I student-athletes' collected and analyzed their personal data to inform their practice, play, and even daily habits (e.g., sleep, screen time). The continuity principle emerges in this study in two ways. Firstly, by having the students collect and analyze their own personal data, students had to consider the methodologies used as well as the level of detail in the collection process [85]. Understanding that data collection methodologies are designed to highlight some aspects and not others is an important aspect of critically reflecting on data's origins, as it helps students gain an understanding of the problematic non-neutral aspect of data [85]. Secondly, by interpreting the collected data as a metric of their performance, it helped students establish the notion that data was already a part of their of lives [85]. Thus, when students merge their learning environment with their surrounding reality [66], they can understand that knowledge is "situated" and cannot be detached from the contexts in which it is created or realized [1, 11, 74].

Learners can also find themselves embedded within various public datasets. For example, historical census and voting data provide large-scale information on legislation, policy opinions, and local demographics. Additionally, satellites, tide gauges, and other automated devices collect weather, climate, air, and water quality benchmark data multiple times every day [91]. The vast range of public datasets available can enable students to view themselves as people who can use data for purposes that interest them - whether it is for "intellectual, social, civic, ethical [or] aesthetic" purposes [69]. Some examples of student investigations where learners extensively engaged with public datasets include [7, 8, 27, 30, 75, 82, 87, 90]. Other instances that demonstrate how learners find themselves embedded in data include, students engaging with data analysis cycles about self and movement [2, 16, 40, 47, 53, 71]. These studies underscore the importance of learners having access to their own data.

5.1.2 Students pose questions for analysis and analyze data to substantiate or promote a new or clarified narrative. Critical data literacy includes learners identifying and analyzing data to clarify prevailing narratives in our society or to develop their own counter-narratives that surface issues they are passionate about. Consistent generation and engagement with data that examines the different facts and categories considered (what was present versus what was hidden) can bring into light hidden realities and construct novel concepts [85]. This approach to data analysis can be particularly beneficial in educational contexts, as early research suggests that using complex, "messy" public data can be more advantageous for students than textbook-like data from secondhand sources [55]. The multivariate relationships that may be present in a public dataset allow learners to have choices in deciding the type of questions that can be answered with data. For instance, in Wilkerson and Laina's project [90], students used publicly available data to gather insights about their local community's years-long rodent problem. Having a personal connection with data allowed learners to formulate insightful questions, as learners independently developed investigative questions and identified sources of data [90]. The degree of freedom [89] was left to the students, leading to a diverse range of investigations within the same project. As part of data processing [85], a key skill developed by the students was linking data to other sources of information, such as temperature and land area. Another example demonstrating degree of freedom [89], includes Wilkerson et al.'s data storytelling project [24], where students explored patterns of climate change in their personal "special places" anywhere in the world. As in the previous example, the teachers acted as facilitators, guiding learners but allowing them to make decisions about the direction of their project. These projects demonstrate how facilitators can effectively guide students by balancing their investigation of data with the right amount of support and guidance. This approach can be especially helpful if students become overwhelmed searching for meaningful relationships, or if they lose sight of the goals of inquiry as different patterns are revealed [55].

5.2 Principle of Cultural Resonance

The principle of cultural resonance stresses the importance of understanding the social context within which data exists and is collected. It encourages students to think critically about the data they receive, and to consider how it reflects on their own identity and beliefs. Additionally, it helps students understand and apply data to their lives and see themselves as multifaceted users of data, thereby developing a deeper understanding of themselves and their world.

5.2.1 Students understand data more fully as the product of complex social factors. It is important to consider the principle of cultural resonance when developing critical data literacy programs, as it allows learners to go beyond the pragmatics of analyzing and reasoning with data. Enyedy et al. [30] exemplify the importance of understanding the limitations of the data [85] with their study of urban students who used GIS maps to study their own communities and demographic trends. The study showed that learners may not always feel that the data lent insight or argumentative power to what they already knew and experienced about past and current inequities. Therefore, it is important to politicize the use of data, and look at it not only from the perspective of a passive user but also in the context of someone who can critique the hidden aspects of data [85]. Incorporating students' lived experiences into classroom projects, as Enyedy et al. [30] demonstrate, can help students better grasp the political and complex nature of data. To critically examine race and power relations, Philip et al. [68] argue that racial literacy is essential for developing critical perspectives towards data. However, only a few educators and researchers fully consider the power and political nuances involved in data collection activities, or educate youth about them [56]. A growing body of evidence suggests that learners can better understand patterns, self, and society when engaging with social, political and economic dimensions of data [56]. For example, Kahn [46] led a project, where students examined how racism, civic and educational opportunities, and financial concerns have impacted people's movement to different parts of the country. Youth conducted interviews with relatives and reviewed library archives related to family migration, comparing the dominant narrative versus the testimonials of people with firsthand experience of what they saw happen [46]. This approach constructs shared knowledge and mitigates limitation of the data by obtaining perspectives from varied stakeholders [54]. A study by Rubel et al. [75] on inspecting local lottery system, for example, encouraged students to speak with different community members (lottery ticket buyers, shopper clerks, etc.) as part of the data collection process. Several other projects also emphasize the importance of assessing individual and local perspectives alongside the inferences made from larger datasets [7, 14, 68]. Therefore, to fully understand the world through data, it is crucial to address central concerns embedded in the raw data by being aware of its interpretive possibilities, recognizing its original context, and engaging in an iterative learning process.

5.2.2 Students engage in projects involving critical consciousness and student identity. As students become more aware of issues surrounding how digital data is gathered and used by various systems, they also become aware about the possible avenues for exclusion generated through certain data-driven algorithms. This was evident in the work of Calabrese Barton et al. [14], who partnered with youth and civic groups to raise valid questions about the aims and potential biases of COVID-19 data. Among other questions, learners asked if the groups they identified with were represented in the data, given that limited data was collected from key populations that were negatively affected, such as indigenous nations in the United States [14]. Moreover, the values, biases, and histories embedded in the dataset, and inequities that may be caused or exacerbated by the biased lenses and framings of the data science algorithms should also be critiqued and examined. This is demonstrated in the study conducted by Cummings et al. [20], where students were encouraged to use strategies, such as looking for evidence of bias or emotion in the information, researching the author/organization's credentials, verifying the information via other sources, etc. before completing an analysis of police traffic stops. At the same time, discussing the social, political, and racial dimensions of data may require talking about the power and privilege that exists within the data set, which determines whose voices get heard [54].

In addition to the biases that are inherent within data, prior research [17, 49, 84] has established that learners exhibit various biases in their data-based reasoning [55]. However, students can demonstrate greater sophistication in analysis and inference making by receiving appropriate support from instructor and classroom experiences. Specifically, the principle of cultural resonance encourages students to provide feedback and critiques to one another, and inspires learners to strive for shared understandings. This appears in the project carried out by Philip et al.[68] where students examined a visualization of Netflix data to identify 'popular' film rentals by neighborhood demographics. It wasn't until a few students dived deeper into the data that the class began to recognize films popular among Black communities [68]. Therefore, once students construct the knowledge for themselves, it is important to share with others and receive additional feedback. This can also equip students to ask better questions, uncover missing or bad data, reflect on the data's limitations and perhaps even challenge the data collection practices [24]. Even when students have proximal experiences with a data set, the ability to see, assess, and use the learning outcomes is still vital as learners are likely to draw limited conclusions from their own data sets and view other data sets as authoritative, in part due to direct knowledge of the shortcomings of the data they collected [42]. By encouraging students to invent, receive critique and iterate over their own understandings and representations of data, learners can develop richer understandings of what is being shown and what is properly inferred [26].

5.2.3 Students understand themselves as multifaceted users of data. The principle of cultural resonance also promotes students' understanding of themselves as people capable of using data for multiple purposes. This arises in a study conducted by Bowler et al. [9], where teenagers in public libraries in Pittsburgh saw themselves as users and owners of data through the use of social media platforms, the creation of mobile media, and the ownership of mobile devices. Furthermore, this principle emphasizes how students can use data to envision a better world. For example, in Taylor and Hall's work [82], youth collected data to identify safe bicycle riding areas in their neighborhood, enabling the mayor's office to better hear youth voices on urban planning decisions. A similar example is a study conducted by Van Wart et al. [87], where students used data in an attempt to persuade city leaders to follow a set of recommendations for a park restoration project. Both of these examples shed light on how students can use data to bring about positive change. While there isn't much research in this area [56], the principle of cultural resonance encourages students to critically reflect on their role as multifaceted data users by examining their relationships with their communities, neighborhoods, and larger societal structures. This includes considering how these groups and individuals are impacted by structures of power, privilege, and inequality that exist at the local, national, and global level.

5.3 Power Principle

The power principle within critical data literacy emphasizes students engaging in meaning-making work with data. In order to process data and understand themselves in new ways, students need to situate their encounters with the world in appropriate cultural contexts to understand "what they are about" [12]. Based on this principle, students also invent methods to describe and explore patterns in data [76] and create new mediums of expression through data that reflect what they know, personally value, and relate to.

5.3.1 Students use data to understand themselves in new ways. Critical data literacy encourages engagement with data in learners' cultural contexts, which help students gain a better understanding of their lived experiences, beliefs, and values. This is important in countering instructional practices that foreground statistical reasoning and present data as objective and devoid of context [35]. In the literature reviewed, the use of wearable technologies such as fitness trackers was employed as a valuable tool for illustrating how learners are situated within their cultural contexts [16], [51], [52], [50]. For example, as part of elementary statistics classroom unit, Lee et al. [51] integrated wearable activity trackers for students to compare favored recess activities to determine which were more physically demanding. In their use of physical activity data, Lee and colleagues [53] advocate for a "quantified self" approach where data across groups of learners are pooled together to highlight patterns and variability. Additionally, Philip et al. [68] also highlight the importance of peer-generated data from mobile technologies as a valuable learning tool for students. This type of data-driven learning encourages students to better understand the complexities of their own cultural identity in relation to that of others and can help learners develop an appreciation for the diversity that exists within their communities. In addition to self tracking devices, log data is also being used to help students engage in self reflection about themselves and their broader community. Scratch Community Blocks is an another attempt to support such reflection [21]. Through these blocks, youth are able to query data about the Scratch user community, such as how popular a particular user is, or which programming blocks are most used. [21]. Thus, by critiquing, questioning, and debating the implications of social data analytics, students have gained a sense of critical data literacy through these experiences [41].

5.3.2 Students explore and create new methods of expression through data. The power principle also encourages students to build artifacts with materials readily available in their classrooms and homes. Data sculptures, which map a variable onto a physical artifact [7], can be an effective way for students to communicate patterns found in their exploration of data. For example, in a study conducted by Matuk et al. [61], a seventh grade student arranged crayons, scissors, and hair elastics to symbolize the relationships found in the Pew dataset about teen's use of social media. Another example includes a group of elementary students mapping air pollution to a more abstract rendering on a loom [7]. Additionally, Stornaiuolo's [80] data stories project allowed high school students to narrate stories about themselves and their world through data T-shirts using personal data they curated and collected. Here too, the artifacts are grounded in real-world experiences of the learners, allowing them to explore and construct their own knowledge. As another method of expression, Bhargava et al. [8] incorporated Freire's Popular Education approach to have students aged 14 to 24 create a "data mural" using local data to depict their community. This process of turning data into art was described by participants as "drawing with lots of information" [8, p. 211].

Furthermore, students' artifacts also become "objects-to-think-with [67, p. 11]". The process of creating and refining their artifacts allows for contemplation and negotiation, demonstrating the creators' ideas through many manifestations [13]. For example, a student may start with one medium, such as clay, but switch to another, such as yarn, when it proves to be more effective in representing their findings [61]. Additionally, the variability of the nature of materials also draws from long histories of more empowering approaches to learning. Teaching data literacy to students with technical backgrounds, such as computer science and mathematics is different from teaching data literacy to non-technical students or novice learners [24]. However, both technical and non-technical learners can express themselves and tell stories with data when using arts-integrated data projects grounded in constructionist theories. The use of interdisciplinary approaches to data literacy can be very beneficial for engaging disengaged students and for supporting deeper learning of critical data perspectives [80]. Additionally, hands-on learning builds confidence for novice learners and creates a low barrier to entry for technical disciplines essential for later learning [93].

5.3.3 Students design experiments and observational studies. Students can further engage with data in sophisticated and meaningful ways by designing their own experiments. As opposed to step-by-step instructional classroom activities, the power principle encourages students to generate different kinds of solutions with no single "right" answer. Designing experiments with data is first and foremost a means of helping students understand what kinds of questions can be answered (or at least partially answered) using data, as well as figuring out if the data one needs is actually collected [24]. Then, once learners understand what kinds of questions may be answered with data, students can proceed with using data to understand real-world phenomena, using data as evidence, arguing from data, and negotiating the meaning of data in various contexts [55]. For example, in Makar & Rubin's project [60], students used data to investigate plant growth in different conditions and examined whether the plant heights in each condition differ from each other after a set period of time. There are also a growing number of simulations and games that support students in designing experiments to make meaning of data. One example is the River City/EcoMUVE project [62] where a mysterious health issue or ecological issue was simulated in three-dimensional, multi-user virtual world. Students were then encouraged to collect data such as interviews with virtual denizens, scientific sample collection from rivers and lakes, observations, and so on to solve the scientific mystery [62]. Another way that simulations have been explored is through embedding fictional events such as earthquakes into classroom space [63]. In Moher's project [63], students designed a study to locate earthquake epicenters using simulated seismographic data displayed on devices in different classroom locations. When children learn to characterize general phenomena that are worth learning about, they develop a basis for thinking about the world and how it functions. These examples illustrate how children's learning can be more meaningful when they use data to develop a basis for thinking about themselves and the world.

6 DISCUSSION

The discussion in the literature review lends us insights for developing a critical understanding of data in youth. Based on the reviewed literature on data-focused constructionist projects, the learning environment or experience can be planned such that learners can (i) connect and reason with data (ii) participate in projects that enable the construction of knowledge about the varied purposes of data, and (iii) develop new methods of expression that reflect what they know, personally value and relate to. Taking this into consideration, designers of learning environments that integrate constructionism with critical data science education should carefully consider if and how the three principles, viz., continuity principle, power principle, and the principle of cultural resonance are realized in the technical and pedagogical design. In addition, the insights gained from this study revealed several trends as well as opportunities for further research that can be used to inform the design of future learning environments. In this section, we highlight these key findings of our work, and discuss the resultant three design opportunities.

6.1 Exploring intermediate possibilities that combine student agency and structure

Given that critical data literacy is explicitly concerned with the perspectives and personal experiences of the learner, a constructionist approach is particularly well-suited for enabling students to make meaning with data. However, in school-classrooms contexts, the content of education is determined by the unified curricula, textbooks, and exams. Due to the high emphasis placed in structure, students do not have high agency in implementing different ways of doing and learning. This makes it challenging to propose open-ended activities that are relevant and engaging to every student, and conflicts with a constructionist approach of student led discovery. However, in constructionist classrooms the role of the teacher and student is not constrained to a particular structure. The teacher may facilitate student learning by providing resources, optimal guidance, and support, while the student takes responsibility for their own learning.

In the course of this literature review, we found little prior work that addressed supporting teachers in implementing constructionist approaches to teach data literacy, as well as in supporting and holding students accountable for their own learning in such environments. The absence of this literature suggests that there is vast potential for further exploration into how educators and students adapt to constructionist data literacy activities in a formal educational environment. Inspired by Karen Brennan's dissertation work [10], we suggest that structure and agency need not be in opposition to each other when designing learning environments. Educators can facilitate constructionist data literacy activities, by first providing scaffolding activities that allow students to acquire data literacy skills and build confidence in their ability to use data. This can take the form of employing simple, gratifying projects. As students progress, teachers can give them ample opportunities to trust their ideas and explore possibilities surrounding their projects by experimenting on their own. For example, students can be encouraged to understand different kinds of data-driven questions based on what is meaningful to them and their communities. At this point, teachers can ask students to share their thought processes and probe them to identify linkages between their individual lives, culture, and values that may be contributing to the evolution of their design process. This practice is inspired by Eleanor Duckworth's research. In her book *Tell Me More* [28], she highlights the importance of the facilitator's involvement as an encouraging listener to sustain the learner's interest in discovering something new. As students share their ideas, it allows them to pause, think, and reflect on their work. Resnick [72] highlights reflection as a critical part of the creative learning process where children engage in discussions to evaluate their design and thinking process. Schön [77] suggests this concept, calling it "reflect in action" (while doing something) and "reflect on action" (after you have done it). In utilizing these conceptual frameworks, instructors can allow students to think creatively and take risks, while also providing structure. Thus, educators and students can both benefit from this approach, resulting in more engaging learning experiences.

6.2 Acknowledging varied backgrounds of students

As part of building critical data literacy, in the literature reviewed we found examples of projects that encouraged learners from diverse backgrounds to think critically about the non-neutral aspects of data, to identify what data is meaningful and what might be biased or incomplete, or if information needs to be corroborated with additional sources. Although some data-driven constructionist projects have been successful in engaging learners with diverse backgrounds, broader efforts are still needed to create an inclusive and empowering learning environment that facilitates an open exchange of ideas between students. For example, previous research has shown that homogeneous and culturally dominant students can control classroom interactions and drive the whole class toward their thinking, while marginalized voices not interested in popular topics can be ignored [38]. Additionally, students might have widely varying degrees of background knowledge and extracurricular experience to draw upon, and what a "basic" discovery is for one student might not be the case for others. To ensure that all students are given equitable opportunities to participate in discussions and share their ideas, we propose that practitioners work towards creating opportunities for students to share their individual backgrounds, experiences, and perspectives with the class. This can be implemented by making space for discourse in classrooms, where learners can engage in conversations with their own or other peoples' artifacts and obtain feedback. Educators should also strive to create a safe space for students to voice their perspectives, to ask questions and to challenge assumptions. A study by Everson and colleagues [32] found that building trust and overcoming broader inequities outside the classroom were essential to making space for critical conversations. In their study students quickly engaged in counter-narratives about computing and linked them to their personal lives once trust was built [48]. Another way to build understanding of the student body is to create opportunities for students to explore personally relevant data and discuss the implications of the data with their classmates. In a culturally heterogeneous environment, it is important to remember that data can be easily misinterpreted and can create false impressions or stereotypes about certain individuals or populations. Therefore, practitioners should also take into account the potential for bias in data and carefully consider context and accuracy to ensure that all students can engage in meaningful dialogue.

6.3 Decentralized curation of local datasets

Many studies in our corpus focused on students collecting data about themselves using wearable activity trackers, self-surveying, or asking questions to their peers. Another means of working with data was through publicly available historical, census, voting, and satellite data. While the collection of publicly available datasets is vast, in some cases the critical use of data will occasionally be hindered by missing data or poorly organized data from students' local communities. In some cases, researchers have sought to fill this gap by using online map services (eg, Google Maps) [82] and geographic information systems [34], as it allowed learners to build and challenge how those data were interpreted based on their own personal experiences living in those locales [54]. Building on prior work and seeking inspiration from the CORGIS project [4], which in addition to providing public data supports the integration of new datasets with libraries for Java, Python, and Racket, we propose a decentralized curation of local datasets. To ensure that all voices are heard and that all perspectives are included in the conversation, Lee and colleagues [54] suggest that data work should be a collaborative effort rather than one person taking on the task alone [54], and propose inviting multiple stakeholders to participate in relating their experiences [54]. Open Data portals that are run at geographically local levels (e.g., by municipal authorities) can be leveraged to some extent, but they are not available everywhere, and their quality varies [86]. Furthermore, they are not necessarily designed for educational activities. A more focused and curated effort, targeting educational use, and supporting local-level participation and contribution may be beneficial.

7 LIMITATIONS

In this review, we chose two sources that we deemed authoritative on data literacy to build the corpus. However, the process of determining these sources as authoritative is inherently subjective and influenced by our own positionality. Additionally, these two sources [54, 56] had the same lead author, which may introduce a bias in perspective on the state of data literacy, resulting in some blind spots. As a result, we acknowledge that the selection of papers reviewed may be subject to bias and may have limited the inclusivity of the corpus and the resulting findings. Furthermore, since what constitutes a constructionist activity is an open question, our determination of constructionist learning approaches may not cover the whole breadth of constructionism. Another potential limitation of this study is the ambiguity surrounding what constitutes "data". The definition of data may vary among individuals, which may result in differing perspectives on the state of data literacy. The other point to consider is that besides numerical measurements, data can also include qualitative information like images, videos, and verbal comments. Although we considered both quantitative and qualitative data, the sample we reviewed skewed more toward quantitative data.

Although our paper delves into constructionist approaches for cultivating critical data literacy, it does not incorporate the complex interdependence of cultural, power, and equity considerations in the design and implementation of constructionist projects [88]. The selection of activities to sample is also limited in that we did not include papers related to artificial intelligence (AI) education in our review, although there has been nascent research on youth collecting data, building AI models and evaluating them. For example, Zimmermann et al. [94] investigated how youth with no programming experience can incorporate ML classifiers into athletic practice by building models of their own physical activity. Despite the importance of exploration, making, and play in AI education projects, their emphasis on learning technical concepts and processes was out of scope for this paper. Additionally, since we required a published document for reviewers and readers to refer to at the time of writing, we were unable to include projects that did not have an academic publication. One example of such projects is St. Clair's Data Games and Data Science Games project [18], which invites students to build visualizations, organize, and manipulate their gameplay using log data in real-time. A final limitation of our study is that it is primarily focused on projects implemented in an informal learning context and didn't capture how teachers used similar activities in a classroom context. That said, we offer our work in the hopes that future scholars will be able to use this as a starting point to further investigate current teacher practices in teaching critical data literacy.

8 CONCLUSION

It has been exciting to see the broad range of empirical settings, pedagogical and design innovations that has been employed to support learners in not only learning about the possibilities offered by data, but also question them. Our work contributes to this broader effort by synthesizing and distilling some of the principles and strategies that occur repeatedly across these works, and connect them to an existing theoretical framework that has been often applied in adjacent contexts. We hope that these contributions will be of use to future scholars and practitioners who want to support learners become active and informed participants in a data-driven and mediated world.

ACKNOWLEDGMENTS

We are grateful to the anonymous reviewers of this paper for their valuable feedback and suggestions.

9 SELECTION AND PARTICIPATION OF CHILDREN

No children participated in this work.

REFERENCES

[1] Edith Ackermann. 2001. Piaget's constructivism, Papert's constructionism: What's the difference. *Future of learning group publication* 5, 3 (2001), 438.

[2] Andria Agesilaou and Eleni A Kyza. 2021. Empowering Students to be Data Literate: The Design and Implementation of a Learning Environment to Foster Critical Data Literacy. In *Proceedings of the 15th International Conference of the Learning Sciences-ICLS 2021.* International Society of the Learning Sciences.

[3] Monica Anderson and Jingjing Jiang. 2018. *Teens, Social Media & Technology 2018.* Technical Report. Pew Research Center. 10 pages.

[4] Austin Cory Bart, Ryan Whitcomb, Dennis Kafura, Clifford A. Shaffer, and Eli Tilevich. 2017. Computing with CORGIS: Diverse, Real-world Datasets for Introductory Computing. *ACM Inroads* 8, 2 (May 2017), 66–72. https://doi.org/10.1145/3095781.3017708

[5] Tim Bell and Michael Lodi. 2019. Constructing Computational Thinking Without Using Computers. *Constructivist Foundations* 14, 3 (2019), 342–351. https://constructivist.info/14/3/342 Number: 3 Publisher: Alexander Riegler.

[6] Francine Berman, Rob Rutenbar, Brent Hailpern, Henrik Christensen, Susan Davidson, Deborah Estrin, Michael Franklin, Margaret Martonosi, Padma Raghavan, Victoria Stodden, and Alexander S. Szalay. 2018. Realizing the potential of data science. *Commun. ACM* 61, 4 (March 2018), 67–72. https://doi.org/10.1145/3188721

[7] Rahul Bhargava and Catherine D'Ignazio. 2017. Data sculptures as a playful and low-tech introduction to working with data. (2017).

[8] Rahul Bhargava, Ricardo Kadouaki, Emily Bhargava, Guilherme Castro, and Catherine D'Ignazio. 2016. Data Murals: Using the Arts to Build Data Literacy. *The Journal of Community Informatics* 12, 3 (Oct. 2016). https://doi.org/10.15353/joci.v12i3.3285

[9] Leanne Bowler, Amelia Acker, Wei Jeng, and Yu Chi. 2017. "It lives all around us": Aspects of data literacy in teen's lives. *Proceedings of the Association for Information Science and Technology* 54, 1 (Jan. 2017), 27–35. https://doi.org/10.1002/pra2.2017.14505401004

[10] Karen Ann Brennan. 2013. *Best of both worlds: Issues of structure and agency in computational creation, in and out of school.* Ph. D. Dissertation. Massachusetts Institute of Technology.

[11] John Seely Brown, Allan Collins, and Paul Duguid. 1989. Situated Cognition and the Culture of Learning. *Educational Researcher* 18, 1 (Jan. 1989), 32–42. https://doi.org/10.3102/0013189X018001032

[12] Jerome S. Bruner. 2003. *The culture of education* (7th print ed.). Harvard Univ. Press, Cambridge, Mass.

[13] Deirdre Butler. 2007. A constructionist view of what it means to be digitally literate – Learning in 21st century classrooms. *Nordic Journal of Digital Literacy* 2, 2 (July 2007), 61–77. https://doi.org/10.18261/ISSN1891-943X-2007-02-02

[14] Angela Calabrese Barton, Day Greenberg, Chandler Turner, Devon Riter, Melissa Perez, Tammy Tasker, Denise Jones, Leslie Rupert Herrenkohl, and Elizabeth A. Davis. 2021. Youth Critical Data Practices in the COVID-19 Multipandemic. *AERA Open* 7 (Jan. 2021), 233285842110416. https://doi.org/10.1177/23328584211041631

[15] Angela Calabrese Barton and Edna Tan. 2020. Beyond Equity as Inclusion: A Framework of "Rightful Presence" for Guiding Justice-Oriented Studies in Teaching and Learning. *Educational Researcher* 49, 6 (Aug. 2020), 433–440. https://doi.org/10.3102/0013189X20927363

[16] Cynthia Carter Ching, Mary K. Stewart, Danielle E. Hagood, and Roxanne Naseem Rashedi. 2016. Representing and Reconciling Personal Data and Experience in a Wearable Technology Gaming Project. *IEEE Transactions on Learning Technologies* 9, 4 (Oct. 2016), 342–353. https://doi.org/10.1109/TLT.2016.2602265

[17] Clark A. Chinn and Betina A. Malhotra. 2002. Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education* 86, 2 (March 2002), 175–218. https://doi.org/10.1002/sce.10001

[18] St Clair. 2016. Data science games. *N. (2016). Data science games. @Concord* 20, 2 (2016).

[19] Tamara Clegg, Daniel M. Greene, Nate Beard, and Jasmine Brunson. 2020. Data Everyday: Data Literacy Practices in a Division I College Sports Context. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376153

[20] Danielle Cummings, Marcus Anthony, Crystal Watson, Ahmad Watson, and Sherle Boone. 2021. Combating Social Injustice and Misinformation to Engage Minority Youth in Computing Sciences. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education.* ACM, Virtual Event USA, 1006–1012. https://doi.org/10.1145/3408877.3432452

[21] Sayamindu Dasgupta and Benjamin Mako Hill. 2017. Scratch Community Blocks: Supporting Children as Data Scientists. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* ACM, Denver Colorado USA, 3620–3631. https://doi.org/10.1145/3025453.3025847

[22] Sayamindu Dasgupta and Benjamin Mako Hill. 2020. Designing for Critical Algorithmic Literacies. http://arxiv.org/abs/2008.01719 arXiv:2008.01719 [cs].

[23] Lisa D. Delpit. 2006. *Other people's children: cultural conflict in the classroom.* New Press : Distributed by W.W. Norton, New York.

[24] Catherine D'Ignazio and Rahul Bhargava. 2018. Creative data literacy: A constructionist approach to teaching information visualization. (2018).

[25] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data feminism.* The MIT Press, Cambridge, Massachusetts.

[26] Andrea A. diSessa. 2004. Metarepresentation: Native Competence and Targets for Instruction. *Cognition and Instruction* 22, 3 (Sept. 2004), 293–331. https://doi.org/10.1207/s1532690xci2203_2

[27] Avantika R. Diwadkar, Susan Yoon, Jooeun Shim, Michael Gonzalez, Ryan Urbanowicz, and Blanca E. Himes. 2021. Integrating Biomedical Informatics Training into Existing High School Curricula. *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science* 2021 (2021), 190–199.

[28] Eleanor Duckworth (Ed.). 2001. *"Tell me more": listening to learners explain.* Teachers College Press, New York.

[29] Catherine D'Ignazio and Rahul Bhargava. 2015. Approaches to Building Big Data Literacy. In *In Proceedings of the Bloomberg data for good exchange conference (p. 6).* 6.

[30] Noel Enyedy and Shiuli Mukhopadhyay. 2007. They Don't Show Nothing I Didn't Know: Emergent Tensions Between Culturally Relevant Pedagogy and Mathematics Pedagogy. *Journal of the Learning Sciences* 16, 2 (March 2007), 139–174. https://doi.org/10.1080/10508400701193671

[31] Virginia Eubanks. 2012. *Digital dead end fighting for social justice in the information age.* https://ieeexplore.ieee.org/book/6267527 OCLC: 1196098268.

[32] Jayne Everson, F. Megumi Kivuva, and Amy J. Ko. 2022. "A Key to Reducing Inequities in Like, AI, is by Reducing Inequities Everywhere First": Emerging Critical Consciousness in a Co-Constructed Secondary CS Classroom. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*. ACM, Providence RI USA, 209–215. https://doi.org/10.1145/3478431.3499395

[33] William Finzer. 2013. The Data Science Education Dilemma. *Technology Innovations in Statistics Education* 7, 2 (2013). https://doi.org/10.5070/T572013891

[34] Christine Franklin and Anna Bargagliotti. 2020. Introducing GAISE II: A Guideline for Precollege Statistics and Data Science Education. *Harvard Data Science Review* 2, 4 (Oct. 2020). https://doi.org/10.1162/99608f92.246107bb

[35] Christine A Franklin, Gary D Kader, Anna E Bargagliotti, and Richard L Scheaffer. 2015. *Statistical Education of Teachers.* Technical Report. American Statistical Association. 88 pages. https://www.amstat.org/asa/files/pdfs/EDU-SET.pdf

[36] Paulo Freire. 2020. Pedagogy of the oppressed. In *Toward a Sociology of Education*. Routledge, 374–386.

[37] Henry A. Giroux. 2001. *Theory and resistance in education: towards a pedagogy for the opposition* (rev. and expanded ed ed.). Bergin & Garvey, Westport, Conn.

[38] Seema Gupta. 2011. Constructivism as a paradigm for teaching and learning. *International Journal of Physical and Social Sciences* 1, 1 (2011), 23–47.

[39] Eric Gutstein. 2003. Teaching and Learning Mathematics for Social Justice in an Urban, Latino School. *Journal for Research in Mathematics Education* 34, 1 (Jan. 2003), 37. https://doi.org/10.2307/30034699

[40] Lisa Hardy, Colin Dixon, and Sherry Hsi. 2020. From Data Collectors to Data Producers: Shifting Students' Relationship to Data. *Journal of the Learning Sciences* 29, 1 (Jan. 2020), 104–126. https://doi.org/10.1080/10508406.2019.1678164

[41] Samantha Hautea, Sayamindu Dasgupta, and Benjamin Mako Hill. 2017. Youth Perspectives on Critical Data Literacies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 919–930. https://doi.org/10.1145/3025453.3025823

[42] Barbara Hug and Katherine L. McNeill. 2008. Use of First-hand and Second-hand Data in Science: Does data type influence classroom conversations? *International Journal of Science Education* 30, 13 (Oct. 2008), 1725–1751. https://doi.org/10.1080/09500690701506945

[43] C. Nash John and K. Quon Tony. 1996. Issues in Teaching Statistical Thinking with Spreadsheets. *Journal of Statistics Education* 4, 1 (March 1996), 3. https://doi.org/10.1080/10691898.1996.11910503

[44] Britney Johnson, Ben Rydal Shapiro, Betsy DiSalvo, Annabel Rothschild, and Carl DiSalvo. 2021. Exploring Approaches to Data Literacy Through a Critical Race Theory Perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–15. https://doi.org/10.1145/3411764.3445141

[45] Yasmin B. Kafai. 1996. *Constructionism in Practice: Designing, Thinking, and Learning in a Digital World.* Routledge.

[46] Jennifer Kahn. 2020. Learning at the Intersection of Self and Society: The Family Geobiography as a Context for Data Science Education. *Journal of the Learning Sciences* 29, 1 (Jan. 2020), 57–80. https://doi.org/10.1080/10508406.2019.1693377

[47] Seokbin Kang, Leyla Norooz, Vanessa Oguamanam, Angelisa C. Plane, Tamara L. Clegg, and Jon E. Froehlich. 2016. SharedPhys: Live Physiological Sensing, Whole-Body Interaction, and Large-Screen Visualizations to Support Shared Inquiry Experiences. In *Proceedings of the The 15th International Conference on Interaction Design and Children*. ACM, Manchester United Kingdom, 275–287. https://doi.org/10.1145/2930674.2930710

[48] AJ KO, Anne Beitlers, Brett Wortzman, Matt Davidson, Alannah Oleson, Mara Kirdani-Ryan, and Stefania Druga. 2021. Critically Conscious Computing: Methods for Secondary Education. *CONSULTATO: 10* (2021).

[49] Deanna Kuhn. 1989. Children and adults as intuitive scientists. *Psychological Review* 96, 4 (1989), 674–689. https://doi.org/10.1037/0033-295X.96.4.674

[50] Victor R. Lee, Joel Drake, Ryan Cain, and Jeffrey Thayne. 2021. Remembering What Produced the Data: Individual and Social Reconstruction in the Context of a *Quantified Self* Elementary Data and Statistics Unit. *Cognition and Instruction* 39, 4 (Oct. 2021), 367–408. https://doi.org/10.1080/07370008.2021.1936529

[51] Victor R. Lee, Joel Drake, and Kylie Williamson. 2015. Let's Get Physical: K-12 Students Using Wearable Devices to Obtain and Learn About Data from Physical Activities. *TechTrends* 59, 4 (July 2015), 46–53. https://doi.org/10.1007/s11528-015-0870-x

[52] Victor R. Lee, Joel R. Drake, Ryan Cain, and Jeffrey Thayne. 2015. Opportunistic uses of the traditional school day through student examination of Fitbit activity tracker data. In *Proceedings of the 14th International Conference on Interaction Design and Children*. ACM, Boston Massachusetts, 209–218. https://doi.org/10.1145/2771839.2771861

[53] Victor R. Lee, Joel R. Drake, and Jeffrey L. Thayne. 2016. Appropriating Quantified Self Technologies to Support Elementary Statistical Teaching and Learning. *IEEE Transactions on Learning Technologies* 9, 4 (Oct. 2016), 354–365. https://doi.org/10.1109/TLT.2016.2597142

[54] Victor R. Lee, Daniel R. Pimentel, Rahul Bhargava, and Catherine D'Ignazio. 2022. Taking data feminism to school: A synthesis and review of pre-collegiate data science education projects. *British Journal of Educational Technology* 53, 5 (Sept. 2022), 1096–1113. https://doi.org/10.1111/bjet.13251

[55] Victor R Lee and Michelle H Wilkerson. 2018. *Data Use by Middle and Secondary Students in the Digital Age: A Status Report and Future Prospects.* Technical Report. National Academies of Sciences, Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6-12. Washington, D.C. 43 pages.

[56] Victor R. Lee, Michelle Hoda Wilkerson, and Kathryn Lanouette. 2021. A Call for a Humanistic Stance Toward K–12 Data Science Education. *Educational Researcher* 50, 9 (Dec. 2021), 664–672. https://doi.org/10.3102/0013189X211048810

[57] Ilya Levin and Dina Tsybulsky. 2017. The Constructionist Learning Approach in the Digital Age. *Creative Education* 08, 15 (2017), 2463–2475. https://doi.org/10.4236/ce.2017.815169

[58] Josephine Louie. 2022. *Critical data literacy: Creating a more just world with data.* Technical Report. Workshop on Foundations of Data Science for Students in Grades K-12. 16 pages.

[59] Deborah Lupton and Ben Williamson. 2017. The datafied child: The dataveillance of children and implications for their rights. *New Media & Society* 19, 5 (May 2017), 780–794. https://doi.org/10.1177/1461444816686328

[60] Katie Makar and Andee Rubin. 2014. Informal statistical inference revisited. In *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9*, Vol. 9. 1–6.

[61] Camillia Matuk, Kayla DesPortes, Anna Amato, Megan Silander, Ralph Vacca, Veena Vasudevan, and Peter J Woods. 2021. Challenges and opportunities in teaching and learning data literacy through art. In *Proceedings of the 15th International Conference of the Learning Sciences-ICLS 2021*. International Society of the Learning Sciences.

[62] Shari Metcalf, Amy Kamarainen, M. Tutwiler, Tina Grotzer, and Chris Dede. 2011. Ecosystem Science Learning via Multi-User Virtual Environments. *IJGCMS* 3 (Jan. 2011), 86–90. https://doi.org/10.4018/jgcms.2011010107

[63] Tom Moher. 2006. Embedded phenomena: supporting science learning with classroom-sized distributed simulations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Montréal Québec Canada, 691–700. https://doi.org/10.1145/1124772.1124875

[64] Na'ilah Suad Nasir and Victoria M. Hand. 2006. Exploring Sociocultural Perspectives on Race, Culture, and Learning. *Review of Educational Research* 76, 4 (Dec. 2006), 449–475. https://doi.org/10.3102/00346543076004449

[65] Luci Pangrazio and Neil Selwyn. 2018. "It's Not Like It's Life or Death or Whatever": Young People's Understandings of Social Media Data. *Social Media + Society* 4, 3 (July 2018), 205630511878780. https://doi.org/10.1177/2056305118787808

[66] Seymour Papert and Idit Harel. 1991. Situating constructionism. *Constructionism* 36, 2 (1991), 1–11.

[67] Seymour A. Papert. 2020. *Mindstorms.* Hachette UK. OCLC: 1239803576.

[68] Thomas M. Philip, Maria C. Olivares-Pasillas, and Janet Rocha. 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* 34, 4 (Oct. 2016), 361–388. https://doi.org/10.1080/07370008.2016.1210418

[69] Thomas M. Philip, Sarah Schuler-Brown, and Winmar Way. 2013. A Framework for Learning About Big Data with Mobile Technologies for Democratic Participation: Possibilities, Limitations, and Unanticipated Obstacles. *Technology, Knowledge and Learning* 18, 3 (Oct. 2013), 103–120. https://doi.org/10.1007/s10758-013-9202-4

[70] Philip J. Piety. 2013. *Assessing the educational data movement.* Teachers College Press, New York.

[71] Kyrill Potapov, Asimina Vasalou, Victor Lee, and Paul Marshall. 2021. What do Teens Make of Personal Informatics?: Young People's Responses to Self-Tracking Practices for Self-Determined Motives. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–10. https://doi.org/10.1145/3411764.3445239

[72] Mitchel Resnick. 2007. All I really need to know (about creative thinking) I learned (by studying how children learn) in kindergarten. In *Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition - C&C '07*. ACM Press, Washington, DC, USA, 1–6. https://doi.org/10.1145/1254960.1254961

[73] Jim Ridgway. 2016. Implications of the Data Revolution for Statistics Education: The Data Revolution and Statistics Education. *International Statistical Review* 84, 3 (Dec. 2016), 528–549. https://doi.org/10.1111/insr.12110

[74] Barbara Rogoff. 2003. *The cultural nature of human development.* Oxford University Press, Oxford [UK] ;a New York.

[75] Laurie H. Rubel, Maren Hall-Wieckert, and Vivian Y. Lim. 2016. Teaching Mathematics for Spatial Justice: Beyond a Victory Narrative. *Harvard Educational Review* 86, 4 (Dec. 2016), 556–579. https://doi.org/10.17763/1943-5045-86.4.556

[76] Daniel L. Schwartz and Taylor Martin. 2004. Inventing to Prepare for Future Learning: The Hidden Efficiency of Encouraging Original Student Production in Statistics Instruction. *Cognition and Instruction* 22, 2 (June 2004), 129–184. https://doi.org/10.1207/s1532690xci2202_1

[77] Donald A. Schön. 1990. *Educating the reflective practitioner: toward a new design for teaching and learning in the professions.* Jossey-Bass, San Francisco.

[78] Ben Rydal Shapiro, Amanda Meng, Cody O'Donnell, Charlotte Lou, Edwin Zhao, Bianca Dankwa, and Andrew Hostetler. 2020. Re-Shape: A Method to Teach Data Ethics for Data Science Education. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376251

[79] Hannah Snyder. 2019. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research* 104 (Nov. 2019), 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

[80] Amy Stornaiuolo. 2020. Authoring Data Stories in a Media Makerspace: Adolescents Developing Critical Data Literacies. *Journal of the Learning Sciences* 29, 1 (Jan. 2020), 81–103. https://doi.org/10.1080/10508406.2019.1689365

[81] Keith S Taber. 2012. Constructivism as educational theory: Contingency in learning, and optimally guided instruction. In *Educational theory*. Nova, 39–61.

[82] Katie Headrick Taylor and Rogers Hall. 2013. Counter-Mapping the Neighborhood on Bicycles: Mobilizing Youth to Reimagine the City. *Technology, Knowledge and Learning* 18, 1-2 (July 2013), 65–93. https://doi.org/10.1007/s10758-013-9201-5

[83] Sherry Turkle and Seymour Papert. 1992. Epistemological Pluralism: Styles and voices within the computer culture. *Humanistic Mathematics Network Journal* 1, 7 (1992), 8.

[84] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science* 185, 4157 (Sept. 1974), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

[85] Alan Freihof Tygel and Rosana Kirsch. 2016. Contributions of Paulo Freire for a Critical Data Literacy: a Popular Education Approach. *The Journal of Community Informatics* 12, 3 (Oct. 2016). https://doi.org/10.15353/joci.v12i3.3279

[86] Jurgen Umbrich, Sebastian Neumaier, and Axel Polleres. 2015. Quality Assessment and Evolution of Open Data Portals. In *2015 3rd International Conference on Future Internet of Things and Cloud*. IEEE, Rome, Italy, 404–411. https://doi.org/10.1109/FiCloud.2015.82

[87] Sarah Van Wart, Kathryn Lanouette, and Tapan S. Parikh. 2020. Scripts and Counterscripts in Community-Based Data Science: Participatory Digital Mapping and the Pursuit of a Third Space. *Journal of the Learning Sciences* 29, 1 (Jan. 2020), 127–153. https://doi.org/10.1080/10508406.2019.1693378

[88] Shirin Vossoughi, Paula K. Hooper, and Meg Escudé. 2016. Making Through the Lens of Culture and Power: Toward Transformative Visions for Educational Equity. *Harvard Educational Review* 86, 2 (June 2016), 206–232. https://doi.org/10.17763/0017-8055.86.2.206

[89] L. S. Vygotsky. 1980. *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

[90] Michelle Hoda Wilkerson and Vasiliki Laina. 2018. Middle school students' reasoning about data and context through storytelling with repurposed local data. *ZDM* 50, 7 (Dec. 2018), 1223–1235. https://doi.org/10.1007/s11858-018-0974-9

[91] Michelle Hoda Wilkerson and Joseph L. Polman. 2020. Situating Data Science: Exploring How Relationships to Data Shape Learning. *Journal of the Learning Sciences* 29, 1 (Jan. 2020), 1–10. https://doi.org/10.1080/10508406.2019.1705664

[92] Geoff Wong, Trish Greenhalgh, Gill Westhorp, Jeanette Buckingham, and Ray Pawson. 2013. RAMESES publication standards: meta-narrative reviews. *BMC Medicine* 11, 1 (Dec. 2013), 20. https://doi.org/10.1186/1741-7015-11-20

[93] Barry J. Zimmerman. 2000. Attaining Self-Regulation. In *Handbook of Self-Regulation*. Elsevier, 13–39. https://doi.org/10.1016/B978-012109890-2/50031-7

[94] Abigail Zimmermann-Niefield, Makenna Turner, Bridget Murphy, Shaun K. Kane, and R. Benjamin Shapiro. 2019. Youth Learning Machine Learning through Building Models of Athletic Moves. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. ACM, Boise ID USA, 121–132. https://doi.org/10.1145/3311927.3323139