Educational Researcher

A Call for a Humanistic Stance toward K-12 Data Science Education

Journal:	*ER*
Manuscript ID	EdR-20-RE-0145.R3
Manuscript Type:	Reviews/Essays

A Call for a Humanistic Stance toward K-12 Data Science Education

EdR-20-RE-0145.R3

Manuscript received: March 31, 2020
Revised: February 18, 2021; May 24, 2021; August 27, 2021
Accepted: August 29, 2021

Victor R. Lee, PhD (corresponding author)
Stanford University
vrlee@stanford.edu

Michelle Hoda Wilkerson, PhD
University of California, Berkeley

Kathryn Lanouette, PhD
William & Mary

KEYWORDS FROM SCHOLARONE:
Substantive Keywords: Computers and Learning, Curriculum, Science Education, Mathematics Education, Technology
Methodology Keywords: Content Analysis, Case Studies

Abstract

There is growing interest in how to better prepare K-12 students to work with data. In this article, we assert that these discussions of teaching and learning must attend to the human dimensions of data work. Specifically, we draw from several established lines of research to argue that practices involving the creation and manipulation of data are shaped by a combination of personal experiences, cultural tools and practices, and political concerns. We demonstrate through two examples how our proposed humanistic stance highlights ways that efforts to make data personally relevant for youth also necessarily implicate cultural and sociopolitical concerns that impact the design and learning opportunities in data-rich learning environments. We offer an interdisciplinary framework based on literature from multiple bodies of educational research to inform design, teaching and research for more effective, responsible, and inclusive student learning experiences with and about data.

Shifting technological infrastructures have expanded how researchers and professionals collect, access, and analyze data in service of education. The emerging field of data science (e.g., Berman et al., 2018) has impacted how data are used in educational decision making (e.g., Piety, 2015), which in turn places new demands on teachers and administrators to use data ethically and effectively (Mandinach et al., 2015). More recently, increased attention to data has also led to growing interest in how educators might support K-12 students in learning about data (Finzer, 2013; Lee & Wilkerson 2018). This is evident in emerging standards and journal special issues that focus on instruction about "big data" and "data science" across domains (e.g., Bargagliotti et al., 2020; Ridgway, 2016; Wilkerson & Polman, 2020). It is also evident in a coordinated effort underway to promote "data science for everyone"[1].

Given this changing landscape, educators have an immediate obligation to consider the nature of students' learning interactions with data. We argue that such interactions are far more complex and wide-reaching than are often presented in curricula and professional development materials. Consider one common distinction that educators make between engaging learners with "first-hand" data that students generate themselves, or "second-hand" data that are provided by a teacher or a curriculum (Hug & McNeill, 2008). Conventional wisdom suggests that engaging students in primary data collection represents a more authentic, personally relevant, and conceptually rich learning experience. However, given how data are currently used in professional practice, some argue that making sense of second-hand data is an important authentic experience in its own right (Duschl, 2008). Recent work suggests that students do engage data collected by others in deeply personal ways. For example, students may interpret such data through lenses connected to their own experiences of race (Philip et al., 2016), place

(Taylor, 2017; Wilkerson & Laina, 2018), and existing data cultures (Van Wart et al., 2020). At the same time, the introduction and use of new automated data collection tools that record data precisely and constantly can undermine some assumed conceptual benefits of collecting first-hand data, such as observing the variability of data as students make their own measurements and errors (Petrosino et al., 2003).

These examples demonstrate that students' experiences of data collection, visualization, analysis, and interpretation are becoming more complex. This complexity reveals the extent to which any educational data activity is in fact a product of manifold individual and social tools and processes. At this critical moment when calls to provide students with more data-intensive learning experiences are still in their formative stages, we argue there needs to be more attentiveness to such tools and processes. Already we are beginning to see the undesirable societal consequences of too hasty an embrace of data (e.g., O'Neil, 2016). These can come in the form of biased data-reliant algorithms, a rush to teach specific marketable skills and programming languages that may not be needed in a few years' time, overly generous claims about how data are used in practice, and a general lack of critical reflection about why students should learn about data.

Too often, the political and industry forces that shape educational reform operate on well-intentioned but inadequately informed models of how teaching and learning work. The assumption that data science is essential because that is where there are currently high-paying jobs, or because it is seemingly an inherently more exciting curricular pursuit for students, ignores the many personal, social, and political factors that shape students' interactions with data. These assumptions can lead to overinvestment in programs that underdeliver because of

inadequate early recognition of the complex personal and social processes, values, constraints, and goals embedded within our educational systems (e.g., Cuban, 2001).

Recognizing that those are impending risks for current data science education zeal (Philip et al., 2013), our goal in this article is to articulate ways in which educators and researchers can deliberately center these human dimensions of student engagement with data—what we call a *humanistic* stance toward data science education. We remain hopeful for what civic possibilities could result from data-intensive learning experiences. At the same time, principled consideration of the human and relational complexity of such experiences must be early and prominent parts of any conversation about data science education for K-12 students. Through a proposed framework and discussion of two cases, we hope to encourage new and more thoughtful ways for approaching the characterization, design, and analysis of student learning with and about data.

A Framework for Attending to the Humanistic Aspects of Data Work

We present a cross-disciplinary three-part framework that represents a synthesis of several lines of ongoing research that have explored how students reason and learn with data across the curriculum. Such scholarship spans statistics education, science education, learning sciences, and new media studies and broadly represents the landscape of cognitive, sociocultural, and political orientations in educational research (e.g., Irgens et al., 2020). We synthesize these works to highlight their specific implications for learning with data, including the significance of:

- Students' personal and direct experiences with data, measurement, and the contexts in which data are collected;

- The cultural and sociotechnical infrastructures and values enacted during a dataset's collection and use (including but not limited to routines, technologies, and norms associated with various classroom, cultural, and disciplinary communities); and
- The enduring political and social narratives that affect the purposes and methods by which datasets are constructed, interpreted, and used as social texts.

We refer to these categories of concern as *layers*, but maintain that they operate simultaneously and in interaction with one another as students engage with a given dataset (Figure 1). For example, while we present personal and sociopolitical considerations as two layers, students' identities are at once personal and co-constructed in conversation with broad social narratives about race, gender, disability status, and more[2]. Similarly, what we consider to be students' personal interests are deeply shaped by their cultural experiences and circumstances (Azevedo, 2011).

<INSERT FIGURE 1 AROUND HERE>

We envision various core data practices—such as collecting, visualizing, analyzing, interpreting, or communicating data—as *constellations* that extend across and thus are shaped by all three layers. Figure 1 illustrates this by depicting nodes existing across layers. Some nodes are more prominent than others in particular layers, so that looking across layers highlights a larger set of interrelationships. As illustrated below, we believe it is analytically profitable to draw attention to each of these explicitly, through the shorthand of layers, in order to gain insight into their influences on students' opportunities to learn, including the way aspects of data activities are often taken for granted. Bringing these layers together foregrounds complex issues and questions that should be asked in both research and design of K-12 data science education.

Personal Layer

The first layer of the framework focuses on the immediate experiences, interests, prior knowledge, and other personal aspects that inform learners' reasoning about a dataset. These include learners' direct experiences of the data collection process, whether those involve developing measures (Lehrer et al., 2007), recording observations (Lehrer & Schauble, 2000), or manipulating some phenomenon as it is being measured through tools such as automated sensors (Thornton & Sokolof, 1997). They also include learners' direct involvement in the context for data collection and analysis. Such activities in the extant literature include, but are not limited to learners posing questions for analysis with data (Arnold, 2007), designing experiments or observational studies (Hardy et al., 2020), and visiting field sites (Manz, 2012). Students also at times are involved in the design of visual representations of data (Lehrer & Schauble, 2004) or inventing methods to describe and explore patterns in data (Schwartz & Martin, 2004). Finally, learners' personal knowledge of a dataset's history or the situation it references also shapes engagement with that dataset (Lee et al., in press). In some cases, that personal knowledge is required because the data are about their families, their communities, or even their own bodies (e.g., Kahn, 2020; Lee & Dubovi, 2020; Van Wart et al., 2020). We describe data that students are directly involved in creating as having higher authorship proximity.

Because high authorship proximity is intuitively thought to strengthen relationships to data, the personal layer is perhaps the most heavily studied in our framework. Decades of research from the statistics and science education communities have focused on how learners' direct experience with collecting data can prepare them to explore statistical patterns including measures of center, variability, trends, and noise (Pfannkuch et al., 2018), to engage meaningfully in data practices such as sampling, measurement, and modeling (Lehrer & English,

2018), and make inferences from data based on their knowledge of the data context (Makar & Rubin, 2009). Similarly, by inventing representations and methods of analysis, students develop understandings of the rules and rationale that underlie conventional treatments of data, and develop flexibility for working with novel data forms and patterns (Lehrer & Schauble, 2004). More broadly, the personal layer is associated with developing a sense of agency and ownership over a dataset and associated data products, and with developing a general understanding of the nature of science.

Importantly, as highlighted by Lee and Wilkerson (2018), proximal experiences with a dataset and its associated context do not always lead to productive outcomes. Studies have suggested that while collecting data about students themselves can engage students' interests, it might also limit students' motivation to reason about broad patterns in a dataset—instead, focusing primarily on themselves and comparing their own cases to those of others (Konold et al, 2015). Hug and McNeill (2008) reported that students were less likely to draw conclusions from their own datasets, in part due to direct knowledge of the limitations of the data they collected. These students also viewed other datasets as authoritative, even if those datasets were subject to the same errors as their own. This, in turn, highlights how students' personal experiences may impact what data sources they deem as trustworthy.

Cultural Layer

A second layer involves the sociotechnical tools, artifacts, and cultural practices that guide and maintain a community of participants in activity. This includes what are often referred to as *disciplinary* practices—the approaches, methods, and instruments developed within science and statistics communities to generate and analyze datasets (Bybee, 2011). It also includes

computational methods and tools emerging from the data science community, which shape what analyses are possible and accessible to young learners (Erickson et al., 2019; Konold, 2007; McNamara, 2018). Finally, the cultural layer includes the norms and procedures that might be developed through classroom consensus as a student community negotiates collective approaches to data generation and analysis (Manz, 2016), as well as students' own repertoires of cultural practices and knowledge, which can serve to inform what they choose to attend to when engaging in reasoning about data (González et al., 2006).

Because culture is embedded in tools and practices, the impact of this layer on students' engagement with data can be significant, but also often implicit and uncriticized. For instance, the use of popular spreadsheet tools common in business, such as Excel, allows students to easily create graphs and calculate summary statistics, but limits students' ability to manipulate data or reflect on analytic processes in the way other scientific analysis packages allow (John & Tony, 1996). More broadly, using digital data analysis tools provides powerful statistics and visualizations, but can limit students' opportunities to explore more artistic visualization methods that emphasize trajectories of experience, outliers, and storytelling (e.g., Lupi & Posavec, 2016). Similarly, the common western scientific practice of positioning scientists (and students) as observers separate from the system under study leads to certain sampling and measurement practices that are taken for granted, such as scooping water samples from the edge of a river. Bang et al. (2012) present an example of "desettling" these cultural divisions between humans and nature, in part by inviting students to wade waist-high into the river and develop a relationship with the surrounding water. In this way, the cultural layer can subtly but substantially shape what is measured and how, what types of patterns can be uncovered and described, and how investigators collect, calibrate, evaluate, and communicate data and findings.

It also shapes whose knowledge and approaches are validated during data work, and what sources of data may be considered trustworthy sources of evidence by students.

Educational researchers have often approached learning experiences involving data as enculturation into using the tools and practices common in western science. These are presented as products of standardization, but that push for standardization is itself cultural. Other research approaches have explored how novel ways of working with data can develop within classroom communities as students build consensus by examining how observations can be structured as data, some data can then be used as evidence, and evidence can be linked with claims that are together eventually transformed into new communally-accepted knowledge (Manz, 2016). This helps illuminate how data practices and tools are developed in communities to tackle specific problems, and how those emergent practices and tools are, in turn, informed by both existing disciplinary approaches and everyday cultural experiences.

Sociopolitical Layer

The sociopolitical layer speaks to the ways in which a given dataset, and the ways data are collected and used more generally, reflect and are shaped by power dynamics. This includes an awareness of the ways in which data are used to, for instance, reproduce anti-Black racism through algorithms trained on biased data (Noble, 2018), or to pathologize the financial behavior of communities of color (Rubel et al., 2016b). It also involves understanding the role of data in corporate and capitalist discourses, along with related issues of consent, privacy, surveillance, and displacement (Vakil, 2018). Attending to the sociopolitical layer of students' interactions with data raises questions about *why* we want students to create and become fluent with data in the first place (e.g., scientific advancement, economic competitiveness, or civic engagement),

how such fluency intersects with critical literacies (e.g., Philip et al., 2016), and whose perspectives and interests a given dataset reflects. Importantly, engaging students with any dataset requires an understanding of how that dataset is expected to operate within broader discourses of power and privilege.

The sociopolitical layer of student engagement with data is the least well studied, though interest has increased in recent years. To illustrate how consequential this layer can be, we turn to a familiar example outside of education: Magazines often publish a "best places to live" list (e.g., U.S. News & World Reports, 2020). These lists often leverage data and analysis criteria that appeal to predominantly white, straight, middle class families, often at the expense of the interests of other populations (e.g., Mock, 2020). Similarly, educational approaches to datasets and analysis risk omitting some students' and communities' perspectives, or dismissing how broader systems of power shape why and how data are used. In one example dealing with persistent social inequities, educators took care to center students' lived experience as part of statistical investigation, but some students did not feel that data lent insight or argumentative power to what they already knew and experienced (Enyedy & Mukhopadhyay, 2007). In another example focused on community-based data science partnerships, Van Wart et al (2020) recount how traditional justice-oriented data "scripts" invoked in educational projects, such as data empowering students to compel policymakers to action, fall short if existing power dynamics are not taken seriously.

From the nascent research in this area, it seems that few educators and researchers fully consider the power and political layer of data activity, or raise youth awareness of such concerns. Data are instead instructionally treated as apolitical and, when collected through accepted normative processes, inherently authoritative. There is some emerging evidence, however, that

explicitly engaging learners with the sociopolitical layer of data can lead them to better understand relationships between patterns, self, and society. For example, activities that intentionally weave data about identity and mobility (and associated issues related to power) with lived experience, interviews, journaling, and other ways of knowing have been shown to engage learners in new ways of reasoning about complex data (e.g., Kahn, 2020). This allows learners to explore how their own actions, the data traces those actions leave behind[3] (Latour, 2007) , and the histories of both impact themselves and broader society (Shapiro et al., 2020).

Synthesis and Guiding Questions

The bodies of research characterized by each layer of the framework highlight important and complementary questions that designers and researchers of data-based educational experiences should consider. Table 1 summarizes the layers and includes key questions and some relevant selected works. Certain scholarly works are intentionally listed across multiple layers of the framework; these in particular highlight how *interactions* between the personal, cultural, and/or sociopolitical layers shape student learning.

[INSERT TABLE 1 AROUND HERE]

Using the Framework to Explore Authorship Practices

The framework we present above can be applied to more deeply analyze students' engagement in a variety of data practices ranging from construction, visualization, manipulation, analysis, interpretation, and/or communication of a given dataset. Below, we illustrate the framework through two cases, with a focus on authorship practices. By authorship practices, we mean students' direct involvement in the design and construction (or reconstruction) of a dataset

as a text - including but not limited to its structure, decisions about what to include in the dataset, methods of quantification or categorizing, sampling, generation and recording of data, and data cleaning. The cases come from our own research because our involvements in these projects allow us to articulate tensions and decisions made related to data authorship practices in ways not easily inferred from other published examples. These two cases nuance simple assumptions about the benefits of data authorship by emphasizing how not only personal, but also cultural and sociopolitical layers shaped design decisions and opportunities to learn in each case. They demonstrate how the questions articulated in Table 1 can be mobilized to understand and inform research and design of data learning experiences.

Personalizing Data using Wearable Activity Trackers

Over many years, Lee has developed a program of research and development to engage elementary school students in data analysis through their collection of physical activity data during recess using commercially-manufactured wearable devices (Lee, 2019). By design, the personal layer was a primary focus in how instructional activities were planned and enacted. Using wearable devices for the purpose of supporting student data analysis positioned students as simultaneously the individuals who were obtaining data (in the form of the number of steps taken), and the agents to which the data referenced. Often, this led to students examining data that they 'authored' about their school days activities.

In developing this line of work, Lee and colleagues actively considered and reflected on how these encounters with personal data were being structured by multiple interacting cultural aspects as well. Obtaining data from students, especially during recess, enmeshed the data creation and analysis activities into the local school culture, particularly as experienced by

students. School culture established common routines and activities even in recess that ranged from processes of obtaining play equipment to who participated in which activities on the playground. This led to in-depth and novel student-initiated investigations of common recess activities (Lee et al, 2015). Results from assessment of student learning of elementary statistics content from this approach have been encouraging (Lee et al, in press).

However, in thinking about cultural groups that are implicated in this approach, not only is students' school culture involved but so is a broader "technoathletic" culture (Lee & Drake, 2014), in which records of physical activities tied to fitness and sport are collected and used for purposes ranging from fitness goal setting to establishing social positioning within athletic and wellness communities. Commercial wearable fitness devices for personal data collection had recently emerged as a new genre of consumer technology, and such tools provided the ability for students to easily collect and access data in a form that was familiar to them. However, participants in technoathletic cultures are typically well-resourced adults specifically interested in fitness and wellness. This shaped how the collected data were made available to students: recorded in time increments that reflect how adults structure their activities (e.g., by the hour or half hour), and indexed against an assumed "ideal" of 10,000 steps taken per day, which has been promoted as a wellness standard in the United States. That ideal is actually arbitrary and believed to be sourced to a device marketing campaign from the 1960s (Tudor-Locke & Bassett, 2004).

Use of commercial devices made uniformity and bulk access to data tractable, but it also implicated dependence on proprietary online services coupled with those devices. By design, the data generated through the wearables could only be accessed through these online services. Because these services mediated data access, new questions were introduced for researchers,

teachers, and students to consider at the sociopolitical layer, including issues of data provenance, algorithmic secrecy, and representation (Drake et al, 2017). Moreover, these devices recorded information that students and parents might want to keep private, and retained that information to be used internally to further the company's business interests. To manage this, Lee's team took additional precautionary steps such as using temporary de-identified accounts. The company also had full control over defining *what*, exactly, was being measured: the way that "steps" were defined and counted by the devices was wholly determined by a proprietary algorithm that students could not see nor modify. While this design choice did motivate students to raise and investigate questions about what would be registered as a step, there were a number of occasions when students felt that what the commercial devices recorded and what the students did physically did not correspond with one another. This was exacerbated by ableist assumptions designed into the technology that had become apparent when students with injuries or who used mobility devices were limited in their ability to participate in activities.

Personalizing Public Data through Representation and Transformation

Wilkerson, Lanouette, and colleagues have been studying middle school youths' use of publicly sourced scientific datasets to explore and share stories about how key issues such as nutrition and climate change impact themselves and their communities (Lopez et al., 2021). A major conjecture motivating the project was that by exploring their personal connections to public datasets, students would gain insight into the cultural and sociopolitical layers that shape who and what is counted, measured, and recorded in datasets. Students were encouraged to use data transformation and visualization to explore hidden disparities and highlight their own perspectives within these datasets. In one activity, students supplemented a nutrition dataset

focused on commercial products with cost information and home meals to explore intersections of nutrition, access, representation, and marketing. In another, they explored a dataset of climate indicators by first focusing on places around the world special to them to highlight how the causes and impacts of climate change are unfairly distributed among countries and regions.

Like the case above, these activities were designed to be concretely grounded in students' personal experience. But whereas in the previous case the *cultural* entangling of personal experiences are key, in this case personal knowledge is leveraged to highlight the *sociopolitical* context and history of datasets and their construction and use. By utilizing storytelling conventions, students were expected to humanize the patterns found in data, and to consider how data are mobilized within broader sociopolitical discourses. And students' engagement in authorship practices with data were limited to visualizing, transforming, supplementing, and writing *about* patterns in data, rather than engaging in the design and collection of the core dataset itself.

At the same time, these activities take disciplinary cultural technologies for granted. For instance, much of students' work with data was done within the *Common Online Data Analysis Platform* ("*CODAP*"). This tool was designed to emphasize a core set of actions valued in data science practice, including creating scatterplots and performing certain manipulations such as filtering a large dataset. This emphasis has also shaped the ways in which students were to engage with and personalize particular datasets (e.g., through filtering, grouping, or adding records). *CODAP* does enable students to add multimedia such as images and text to their data "document," supporting the storytelling aspect of the work (Wilkerson et al., 2021). However, in general, it supports a very specific approach to visualizing, analyzing, and transforming data in a space where many options exist (e.g., Storniaulou, 2020).

There were also a number of tradeoffs stemming from the research group's reliance on existing public datasets. The decision to use public datasets was motivated by a desire to center sociopolitical concerns—existing datasets operate as social texts, constructed within a political milieu. However, while students' personal connections to these datasets offered compelling insights into the sociopolitical contexts of data, they did not engage students with specific elements of data construction such as measurement and sampling. Indeed, these details are difficult to find for public datasets, and the measurements used (percent daily value, parts per million, etc.) often require scaffolding to interpret. Additionally, by centering existing datasets as objects of inquiry, students were positioned as reactive to those datasets, rather than as designers and authors of data in their own right.

Opportunities and Considerations Across the Cases

Through use of this framework, we have interrogated students' authorship activities with data and identified how constellations of personal, cultural, and sociopolitical factors shaped what students were able and encouraged to do in both cases. In the first case, elementary students were ostensibly authors of the data that they used during instruction. They were able to both directly experience the generation of data points and to develop statistical and inferential reasoning through work with a broader dataset that had been created through cultural interactions with peers, but were constrained by commercial and ableist assumptions embedded within the activities. In the second case, middle school students' authorship activities involved transforming existing data sets, in the process learning how sociopolitical forces shape the structure and content of the dataset. They had opportunities to explore how datasets can be contextualized but had less exposure to data collection and were positioned as reactive rather than proactive data

authors. Across both cases, the researcher and educator teams worked to make data personally relevant, and yet the opportunities to learn and implications across different layers of the framework were still quite different.

The framework also highlights how considering the personal, cultural, and sociopolitical layers of data engagements can expose otherwise missed or taken-for-granted features of activities. It highlights how different activities may complement one another to provide students more robust insights into data that can extend across tools, disciplines, or experiences. We argue that such analysis will only become increasingly necessary as emerging technologies further complicate the landscape of how data are collected, experienced, structured, and shared. As it stands, students can already engage in authorship practices in a variety of ways—as authors of simulations that create data, users of probeware and other automated sensors, collectors of qualitative data for later quantification, users of online virtual or remote labs, or participants in a distributed system in which data is aggregated across students in the same classroom, or across the globe (Lee & Wilkerson, 2018). Rather than a technocentric treatment of these developments as new categories of data engagement, we argue they should be seen through our multi-layered humanistic stance that is focused on students' opportunities to learn.

Conclusion

In this article, we aimed to respond to changing landscapes of data in society at large and in K-12 contexts in particular. If the recent past is any indication, attention toward data in education—and specifically, about how to best teach and help students learn with and about data—will only increase. Students are already participating in a world where ever more of their daily and future professional activities will involve the collection and analysis of some form of data. They would

be well served if we develop learning experiences that encourage thoughtful and critical participation in practices of data creation, interpretation, analysis, argumentation, and critique, particularly as formalized notions of widespread K-12 data science education gain traction (Lee & Delaney, in press).

What we offer through this article is a framework for educators, designers and researchers to thoughtfully and systematically consider students' humanistic entanglements with data. While our framework is necessarily general and encompassing, it can provide a starting point for the development of more elaborated descriptive frameworks as the field of data science education evolves and new dimensions and interactions become apparent. We advocate for more deliberate acknowledgment and study of our relationships to data being simultaneously shaped by forces that are personal, cultural, and socio-political. Given the three layers we articulate here, discussions about relevance, authenticity, and access become more complicated, complications that we argue are key points of engagement for researchers, educators and students alike. Our cases illustrate that making data "personal" or positioning students as authors is more complicated than intuition might suggest. In particular, first/second-hand distinctions and a focus on particular types of tools or data become less notable here, where varying degrees of closeness and relevance are not inherently tied to physical proximity or production of the data itself.

Such relationships with (and uses of) data are not trivial—they have broader consequences with respect to questions of epistemology and influence in the world beyond classroom walls. When we accept or value particular forms of data - particularly as represented through instruction - we reinforce or increase the influence it has beyond the immediate setting. For instance, accepting that student achievement performance data is a reflection of school effectiveness continues to drive how we design, evaluate, and fund schools (O'Neil, 2016).

Similarly, counting books in a child's home biases against other forms of literacies occurring within families, with such data serving to highlight deficiencies in non-dominant communities and narrowing what forms of literacies are valued and sustained. If education researchers are to work toward any notion of education serving to increase students 'data literacy', 'data acumen', or ability to work in 'data science', we contend that these forces are ones that we, and the students that we serve, must acknowledge.

Ultimately, we foresee that independent of this article, interest in teaching with and about data will continue to grow rapidly. There has been some base literature to inform how we can support that work effectively, and we encourage its use (e.g., Ben-Zvi et al., 2017). At the same time, given growing awareness of how data are intertwined in how we participate in society, more research, theorizing, synthesis, and local innovations are necessary (Wilkerson & Polman, 2020). Through the arguments and examples provided here, we hope to promote deeper awareness of how we engage with data, whether it be through authorship or other forms of data practice (e.g., visualization, critique). Interdisciplinary and humanistic stances that build upon what we provide here could then help us design and implement data-intensive learning experiences more accountable and valuable for all.

Notes

1. See datascience4everyone.org

2. Indeed, a students' identity is further inscribed in the cultural layer as they consider their positionality within a given classroom or disciplinary community, e.g., as a data scientist.

3. e.g. digital records of physical locations, online activities, and clicks and pauses within games, apps, and social media

References

Arnold, P. (2007). What about the P in the PPDAC cycle? An initial look at posing questions for statistical investigation. *Education*, *55*.

Azevedo, F. (2011). Lines of practice: A practice-centered theory of interest relationships. *Cognition & Instruction, 29*(2), 147-184. doi: 10.1080/07370008.2011.556834

Bang, M., Warren, B., Rosebery, A. S., & Medin, D. (2012). Desettling expectations in science education. *Human Development*, 55(5-6), 302-318. https://doi.org/10.1159/000345322

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

Berman, F., Rutenbar, R., Hailpern, B., Christensen, H., Davidson, S., Estrin, D., . . . et al. (2018). Realizing the potential of data science. *Communications of the ACM,* 61(4), 67–72. doi:10.1145/3188721

Ben-Zvi, D., Makar, K., & Garfield, J. (Eds.). (2017). *International handbook of research in statistics education*. Cham: Springer.

Bybee, R. W. (2011). Scientific and engineering practices in K-12 classrooms: Understanding a framework for K-12 science education. *Science and Children*, 49(4), 10-16.

Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.

Drake, J., Cain, R., & Lee, V. R. (2017). From wearing to wondering: Treating wearable activity trackers as tools for inquiry. In I. Levin & D. Tsybulsky (Eds.), *Optimizing STEM Education With Advanced ICTs and Simulations* (pp. 1-29). IGI Global.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, *32*(1), 268-291. https://doi.org/10.3102/0091732X07309371

Enyedy, N., & Mukhopadhyay, S. (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), 139-174.

Erickson, T., Wilkerson, M. H., Finzer, W., & Reichsman, F. (2019). Data moves. *Technology Innovations in Statistics Education, 12*(1).

Finzer, W. (2013). The data science education dilemma. *Technology Innovations in Statistics Education*, *7*(2). Retrieved from https://escholarship.org/uc/item/7gv0q9dc

González, N., Moll, L. C., & Amanti, C. (Eds.). (2006). *Funds of knowledge: Theorizing practices in households, communities, and classrooms*. Routledge.

Hardy, L., Dixon, C., & Hsi, S. (2020). From data collectors to data producers: Shifting students' relationship to data. *Journal of the Learning Sciences*, *29*(1), 104-126. DOI: 10.1080/10508406.2019.1678164

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

Irgens, G. A., Simon, K., Wise, A., Philip, T., Olivares, M. C., Van Wart, S., . . . Kahn, J. (2020). Data Literacies and Social Justice: Exploring Critical Data Literacies through Sociocultural Perspectives. In Gresalfi, M. and Horn, I. S. (Eds.), *The Interdisciplinarity of the Learning Sciences. In M. Gresalfi & I. S. Horn (Eds.), The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020, Volume 1* (pp. 406-413). Nashville, TN: International Society of the Learning Sciences.

John, C. N., & Tony, K. Q. (1996). Issues in teaching statistical thinking with spreadsheets. *Journal of Statistics Education, 4*(1). doi: 10.1080/10691898.1996.11910503

Kahn, J. (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), 57-80.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics*, *88*(3), 305-325.

Latour, B. (2007). Beware, your imagination leaves digital traces. *Times Higher Literary Supplement, 6*(4), 129-131.

Lee, V. R. (2019). On researching activity tracking to support learning: a retrospective. *Information and Learning Sciences, 120*(1/2), 133-154. https://doi.org/10.1108/ILS-06-2018-0048

Lee, V. R., & Delaney, V. (in press). Identifying the content, lesson structure, and data use within pre-collegiate data science curricula. *Journal of Science Education and Technology*. https://doi.org/10.1007/s10956-021-09932-1

Lee, V. R., & Drake, J. (2012). Physical activity data use by technoathletes: Examples of collection, inscription, and identification. In J. van Aalst, K. Thompson, M. J. Jacobson, & P. Reimann (Eds.), *The Future of Learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012)* (Vol. 2, pp. 321-325). International Society of the Learning Sciences.

Lee, V. R., Drake, J., Cain, R., & Thayne, J. (2015). Opportunistic Uses of the Traditional School Day Through Student Examination of Fitbit Activity Tracker Data. In M. U. Bers & G.

Revelle (Eds.), *Proceedings of the 14th International Conference on Interaction Design and Children* (pp. 209-218). ACM. https://doi.org/10.1145/2771839.2771861

Lee, V. R., Drake, J., Cain, R., & Thayne, J. (in press). Remembering what produced the data: Reflective reconstruction in the context of a 'quantified self' elementary data and statistics unit. *Cognition & Instruction*. https://doi.org/10.1080/07370008.2021.1936529

Lee, V. R., & Dubovi, I. (2020). At home with data: Family engagements with data involved in Type 1 Diabetes management. *Journal of the Learning Sciences, 20*(1), 11-31. doi:10.1080/10508406.2019.1666011

Lee, V. R., & Wilkerson, M. (2018). *Data use by middle and secondary students in the digital age: A status report and future prospects*. Commissioned Paper for the National Academies of Sciences, Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6-12. Washington, D.C.

Lehrer, R., Kim, Mj. & Schauble, L. Supporting the Development of Conceptions of Statistics by Engaging Students in Measuring and Modeling Variability. *International Journal Computer Math Learning* 12, 195–216 (2007). https://doi.org/10.1007/s10758-007-9122-2

Lehrer, R., & English, L. (2018). Introducing Children to Modeling Variability. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 229-260). Cham: Springer International Publishing.

Lehrer, R., & Schauble, L. (2000). Inventing data structures for representational purposes: Elementary grade students' classification models. *Mathematical Thinking and Learning, 2*(1&2), 51-74.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, *41*(3), 635-679.

Lopez, M. L., Roberto, C., Rivero, E., Wilkerson, M. H., Bakal, M., & Gutiérrez, K. (2021). Curricular reorganization in the third space: A case of consequential reasoning around data. In E. de Vries, Y. Hod, & J. Ahn (Eds.), *Proceedings of the 2021 Annual Meeting of the International Society for the Learning Sciences (ISLS 2021) (Vol. 1, pp. 466-473)*. Bochum, Germany: ISLS.

Lupi, G., & Posavec, S. (2016). *Dear data.* Chronicle Books.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1).

Mandinach, E. B., Parton, B. M., Gummer, E. S., & Anderson, R. (2015). Ethical and appropriate data use requires data literacy. *Phi Delta Kappan*, *96*(5), 25–28. https://doi.org/10.1177/0031721715569465

Manz, E. (2012). Understanding the codevelopment of modeling practice and ecological knowledge. *Science Education*, *96*(6), 1071-1105.

Manz, E. (2016). Examining evidence construction as the transformation of the material world into community knowledge. *Journal of Research in Science Teaching*, *53*(7), 1113-1140. https://doi.org/10.1002/tea.21264

McNamara, A. (2018). Key attributes of a modern statistical computing tool. *The American Statistician*. *73*(4), 375-384. doi: 10.1080/00031305.2018.1482784

Mock, B. (2020, January 9). What "livability"looks like for black women. *Bloomberg City Lab*. https://www.bloomberg.com/news/articles/2020-01-09/the-best-and-worst-cities-for-black-women

Noble, S. (2018). *Algorithms of Oppression*. New York University Press.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pfannkuch, M., Ben-Zvi, D., & Budget, B. (2018). Innovations in statistical modeling to connect data, chance and context. *ZDM: The International Journal on Mathematics Education,* 50 (7), 1113–1123.

Philip, T. M., Olivares-Pasillas, M. C., & 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*, *34*(4), 361-388. https://doi.org/10.1080/07370008.2016.1210418

Philip, T. M., Schuler-Brown, S., & Way, W. (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), 103-120.

Piety, P. J. (2015). *Assessing the educational data movement*. Teachers College Press.

Ridgway, J. (2016). Implications of the Data Revolution for Statistics Education. *International Statistical Review*, 84(3), 528-549. doi:https://doi.org/10.1111/insr.12110

Rubel, L. H., Lim, V. Y., Hall-Wieckert, M., & Sullivan, M. (2016a). Teaching Mathematics for Spatial Justice: An Investigation of the Lottery. *Cognition and Instruction, 34*(1), 1-26. doi:10.1080/07370008.2015.1118691

Rubel, L. H., Hall-Wieckert, M., & Lim, V. Y. (2016b). Teaching mathematics for spatial justice: Beyond a victory narrative. *Harvard Educational Review, 86*(4), 556-579. doi: 10.17763/1943-5045-86.4.556

Schwartz, D. L., & Martin, T. (2004). Inventing to Prepare for Future Learning: The Hidden Efficiency of Encouraging Original Student Production in Statistics Instruction. *Cognition and Instruction,* 22(2), 129-184. doi:10.1207/s1532690xci2202_1

Shapiro, B. R., Meng, A., O'Donnell, C., Lou, C., Zhao, E., Dankwa, B., & Hostetler, A. (2020, April). 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* (pp. 1-13).

Stornaiuolo, A. (2020). Authoring Data Stories in a Media Makerspace: Adolescents Developing Critical Data Literacies. *Journal of the Learning Sciences,* 20(1), 81-103. doi:10.1080/10508406.2019.1689365

Taylor, K. H. (2017). Learning along lines: Locative literacies for reading and writing the city. *Journal of the Learning Sciences, 26*(4), 533-574. doi: 10.1080/10508406.2017.1307198

Thornton, R. K., & Sokoloff, D. R. (1997, March). Realtime physics: active learning laboratory. In *AIP Conference Proceedings* (Vol. 399, No. 1, pp. 1101-1118). American Institute of Physics.

Tudor-Locke, C., & Bassett, D. R. (2004). How many steps/day are enough? *Sports Medicine*, *34*(1), 1-8.

US News & World Reports. (2020.). *150 Best Places to Live in the US in 2020-2021*. https://realestate.usnews.com/places/rankings/best-places-to-live

Vakil, S. (2018). Ethics, identity, and political vision: Toward a justice-centered approach to equity in computer science education. *Harvard Educational Review*, *88*(1), 26-52. https://doi.org/10.17763/1943-5045-88.1.26

Van Wart, S., Lanouette, K., & Parikh, T. S. (2020). Scripts and counterscripts in community-based data science: Participatory digital mapping and the pursuit of a Third Space. *Journal of the Learning Sciences, 20*(1), 127-153. doi:10.1080/10508406.2019.1693378

Wilkerson, M., Finzer, W., Erickson, T., & Hernandez, D. (2021). Reflective data storytelling for youth: The CODAP Story Builder. In *Proceedings of the 20th ACM SIGCHI Interaction Design and Children Conference (IDC '21) (pp. 503-507)*. Athens, Greece, 24-30 June. doi: 10.1145/3459990.3465177

Wilkerson, M. H., & Laina, V. (2018). Middle school students' reasoning about data and context through storytelling with repurposed local data. *ZDM, 50*(7), 1223-1235. doi:10.1007/s11858-018-0974-9

Wilkerson, M. H., & Polman, J. L. (2020). Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences, 20*(1), 1-10. doi:10.1080/10508406.2019.1705664

Author bios:

Victor R. Lee, PhD, is Associate Professor in the Graduate School of Education at Stanford University, 485 Lasuen Mall, Stanford, CA 94061; vrlee@stanford.edu. His research focuses on future possibilities of STEM education with emphasis on learning with data across contexts, maker education, elementary computer science education, and new technology supports for teachers.

Michelle Hoda Wilkerson, PhD, is an Associate Professor in the Graduate School of Education at the University of California, Berkeley, 2121 Berkeley Way West, Berkeley, CA, 94720; mwilkers@berkeley.edu. Her research focuses on students' development of computational literacies, including but not limited to their understanding and use of data, simulations, and visualizations and the role these artifacts play in science and society.

Kathryn Lanouette, PhD, is an Assistant Professor in the School of Education at William & Mary, 301 Monticello Avenue, Williamsburg, VA, 23185; lanouette@wm.edu. Her research examines the ways in which children learn science, focusing on the interplay of data, place-based pedagogies, and emerging technologies to support expansive and equitable learning opportunities.

Acknowledgements/funding information:

Authors Lee and Wilkerson contributed equally, jointly share lead authorship, and are listed in alphabetical order. The authors thank Thad Domina and the anonymous reviewers for detailed and constructive feedback that greatly improved the framework and writing. Work reported in this article was supported by National Science Foundation Grants 1054280, 1541676, and 1900606. The opinions expressed herein are those of the authors and do not necessarily reflect those of the National Science Foundation.

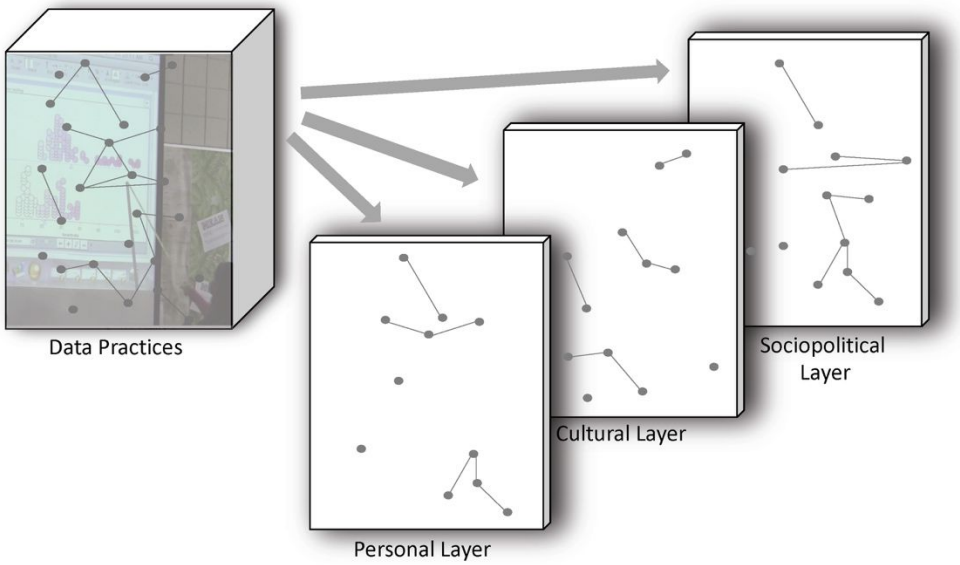

Figure 1. Figure illustrating data practices as constellations of factors that extend across personal, cultural, and sociopolitical layers. This figure intentionally illustrates overlaps across layers, as aspects of data practice within one layer are often linked to others.

Table 1. Layers of mediation within data practices

Layer	Description	Example Questions	Sample Representative Works
Personal	Direct experiences with the design, generation, and analysis of data as well as with the phenomena and contexts from which data are created.	• How might students' existing knowledge and perception of a situation inform their analysis of data? • How might students' direct involvement in generating a data set impact students' understanding of its statistical, computational, and evidentiary features? • How do students' personal interests, affective responses, and identities shape their interactions with data?	Kahn (2020); Lee & Dubovi (2020); Lehrer & Schauble (2004); Makar & Rubin (2009); Stornaiuolo (2020), Wilkerson & Laina (2018); Van Wart et al. (2020)
Cultural	Routines, technologies, and values developed within various classroom, cultural, and disciplinary communities that shape how a dataset is constructed and used.	• What is the role of sociotechnical tools such as data collection devices, analysis packages, or visualization technologies in shaping student learning with data? • How do cultural routines and values impact the ways that students conceptualize and interpret data? • How might the interplay of situation-specific needs and disciplinary practices lead students to treat and interpret data in certain ways? • What epistemic norms are being implicated in data learning activities?	Lee (2014); Bang et al. (2012); Kahn (2020); Konold (2007); Lehrer & Schauble (2004); Manz (2016); McNamara (2018)
Sociopolitical	Enduring political and social forces that affect the purposes and methods by which datasets are constructed, interpreted, and used and who is empowered to do or resist such work.	• How do students come to learn about the ways political and social ideologies are encoded within datasets and their use? • To what extent are the benefits and limitations of using data examined as a form of powered discourse? • How do enacted data-driven practices and cultures serve to differentially empower or marginalize communities? • What degrees of access and visibility are provided with respect to data and their use, and how explicitly are those understood?	Bang et al. (2017); Kahn (2020); Noble (2018); O'Neil (2016); Philip et al. (2013); Rubel et al. (2016); Van Wart et al. (2020)