Sandbox Data Science: Culturally Relevant K-12 Computing

Justice T. Walker
jtwalker@utep.edu
University of Texas at El Paso
El Paso, USA

Amanda Barany
amanda.barany@gmail.com
Some School
Philadelphia, USA

Alex Acquah
aacquah@miners.utep.edu
University of Texas at El Paso
El Paso, USA

Sayed Mohsin Reza
skr6024@psu.edu
Pennsylvania State University
Harrisburg, USA

Karen Del Rio Guzman
kadelrioguzman@miners.utep.edu
University of Texas at El Paso
El Paso, USA

Michael Johnson
michaeljohnson.unt@gmail.com
University of North Texas
Denton, USA

Omar Badreddin
o.badreldin@northeastern.edu
Northeastern University
Boston, USA

Alan Barrera
abarrera6@miners.utep.edu
University of Texas at El Paso
El Paso, USA

Course CS1
Programming Language Python
Knowledge Unit Programming Concepts
CS Topics Functions, Data Types, Expressions, Mathematical Reasoning
Resource Type Curriculum Framework

SYNOPSIS

Given an increased focus on computer science education as a valuable context to teach data science—due in part to the potential of computing for accessing, processing, and analyzing digital datasets—there have been steady efforts to develop kindergarten through 12th grade (K-12) curricula that productively engage learners in these academic areas. Bootstrap: Data Science and Exploring Computer Science (ECS) are prominent curricular examples designed to support high school data science access in computing contexts. While these vital efforts have found success bridging computer and data science, there remain growing concerns about how we can ensure that such learning experiences support the demographic and intellectually diverse cohorts of students needed for field innovation, occupational attainment, and public literacy. Challenges to these efforts often persist because existing data sources and activities offered to students are typically shaped by others (e.g., curriculum designers, teachers, etc.) rather than by learners themselves. This results in inquiry-driven questions, processes, and outcomes that can restrict exploration and engagement, as opposed to inherently and authentically linking to learners' diverse personal interests, styles and concerns. Perspectives in culturally responsive computing (CRC) provide viable frames for how to design learning experiences that encourage learner access, empowerment, and personal interests—key features for spurring field diversity through learning. With this imperative and framing in mind, we share our project called "Coding Like a Data Miner" (CLDM), which leverages a social media-based application programming interface (API) to teach learners how to gather, process (or wrangle), analyze and then communicate insights learned from "big data" sets. We describe this design as sandbox data science (SDS)—an approach to computing-based data science that is consistent with CRC perspectives with demonstrated promise in broadening participation and enhancing productivity in computer science education. In this article, we share insights into our rationale and the theoretical perspectives that drive our curricular design. We then provide an overview of the curriculum with case examples of the sorts of pursuits that can be taken up by learners in this context. Finally, we reflect on CLDM and design principles that make SDS a viable approach to broadening computing-based data science participation and productivity. This curriculum and accompanying resources are publicly available for review, use and adaptation at www.abclearninglab.com/cldm.

KEYWORDS

Computer Science Education, Data Science Education, Culturally Relevant Computing, Sandbox Data Science

ACM Reference Format:

Justice Walker, Amanda Barany, Alex Acquah, Sayed Mohsin Reza, Karen Del Rio Guzman, Michael Johnson Omar Badreddin, Alan Barrera. November 2023. Sandbox Data Science: Culturally Relevant K-12 Computing. In ACM *EngageCSEdu*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3631986.

ACM EngageCSEdu, November 2023.

ACM ISBN 979-8-4007-0484-0/23/11.
https://doi.org/10.1145/3631986

1 ENGAGEMENT HIGHLIGHTS

In this era of ongoing and exponential technical advancement, digital data permeates most aspects of daily life. From smart watches that gather and track our personal daily health activity data, to social media platforms that leverage our browsing histories to inform algorithms about user perspectives, decision-making and behaviors, the collection and analysis of digital data has revolutionized how society functions. These developments are a part of a broader field defined more recently as "data science" [1] that explores data collection and analysis techniques at scales not possible decades before. Concomitant with these developments have been calls from teachers, researchers, and policymakers for education initiatives that will not only address a growing demand for data science professionals, but also prepare future innovators, data scientists, and informed citizens to advance the field and steward its impact [2, 3, 4]. The response has been a series of curricular design efforts intended to bring data science to existing academic disciplines (e.g., mathematics, physics, engineering, etc.) to both advance priorities within those fields and better understand the interdisciplinary nature of learning in these areas [5, 6]. Examples include activities where learners engage with a math or physics-based dataset, and then use data science techniques to gain insights into features or phenomena in that context. While promising, these existing efforts often fail to engage with one key feature of contemporary data science due to its technical complexity: “big data” [7, 8], or the massive data sources that are typically generated through automated processes. One promising solution to this issue lies in the integration of data science and computer science education, with its emphasis on computational thinking [9]. Many of the actions needed to engage productively with data science, such as planning and enacting data collection, processing data into analyzable form, and then under-standing and communicating data sets, can be enacted with relative ease through the application of computer programming. Features of computational thinking and practice such as pattern recognition, decomposition, abstraction, and algorithm design also hold value for conceptualizing data science processes, especially at scale [10].

One challenge with using computer science education as a context for teaching data science involves the creation and curation of learning experiences that promote diversity and inclusion, which are essential for the intellectual diversity that inspires field innovation. In practice, teachers may struggle to accommodate a wide variety of student interests, cultural histories, and geopolitical viewpoints, especially when using existing curricular models. Several earlier studies in computer science education, for example, have provided pre-curated data sources on which students can apply data science and computer programming techniques. While not free of value, this approach ultimately restricts the scope of student engagement, and the kinds of questions learners are able to explore. Due to the lack of learner involvement in data collection and analysis, personalization and relevance in their educational experiences are limited at best. In data science, these issues are further complicated by the fact that learners are often left out when creating datasets, choosing analytical approaches, and engaging with lines of inquiry, thus disconnecting learning experiences from learners and their diverse epistemological (e.g., inquiry) styles. The result is a body of evidence that endorses practice using data sources managed by others rather than by students themselves, restricting the scope and questions in education research and practice. In sum, while computer science education offers several key affordances and opportunities, there remains a set of accompanying equity issues that persist in computing education and have for many decades [11]. In fact, the literature is replete with evidence suggesting that instructional designs that do not account for learners’ diverse social and cultural experiences can have adverse and lasting impacts on learning outcomes [12, 13], including choices regarding field participation [14]. This has been shown to contribute to field attrition and severe underrepresentation among learners who are traditionally at risk of marginalization in these area [15, 16].

In many ways, data science is currently positioned at a critical juncture for research and practice due to new curricular design implementations that circumvent some of these persistent issues in computing. One potential solution lies in the application of best practices observed in other areas of computing education: theories in culturally relevant computing (CRC). CRC theories have precedence as foundational starting points for the intentional design of learning experiences that support learners’ diverse cultural histories, personal interests, and social and political concerns [17, 18]. To address the issues of access and engagement in CS teaching and learning, the use of relevancy as a guiding design concept has been suggested by Ladson-Billings [17] as a crucial component of effective education, and one that is closely tied to issues of equity and social justice. By valuing and centering students’ personal, social, and cultural knowledge and experiences into the learning process, educators can create more meaningful and impactful educational experiences that promote both academic achievement and positive social outcomes. For underrepresented students, educators can in-crease engagement and learning outcomes by creating curricula that fundamentally link to learners’ individual interests, cultural backgrounds, and sociopolitical environment.

In our work, we apply the idea of relevancy through what we call sandbox data science (SDS). SDS is enabled through freely accessible Application Processing Interfaces (APIs) that can be used to gather or “scrape” data from websites or online sources using automated tools or scripts. For us, SDS has emerged as a promising solution to equip learners with tools to conduct their own explorations, addressing the issue of limited personal engagement in pre-college data science curricula. In this way, social media platforms can serve as a massive, diverse, and flexible library of data that learners can use to explore a wide range of inquiries and construct new knowledge. Similar to the constructionist perspectives and open-ended activities in sandbox science using Scratch [19, 20] and electronic textiles, or E-textiles [21] an emphasis on relevancy in curricular design allows for the varied pursuits and problem-solving challenges that can spur computational thinking with diverse learners [9].

In the next section, we describe how relevancy as a design principle informed the development of key activities in our curriculum (activity design). We then provide an illustrative case example of how this approach might be enacted in the classroom and conclude with reflections on what this might mean for computing-based and culturally relevant data science teaching and learning.

2 ACTIVITY DESIGN

Our activity design consists of 5 phases: data gathering, preprocessing, analysis, visualization, and communication.

2.1 Data Gathering

Data gathering is an important first phase in all data science. In SDS, this can be understood as the defining moment when learners (1) choose a culturally relevant or responsive topic to explore, (2) identify a strategy (writing code and setting parameters) for collecting relevant data from a context that can enable their inquiry, and (3) enact their strategy to export a personalized dataset. While this is not a new concept, most areas in STEM education that deploy this approach tend to do so on limited scales; personalized data sets are often generated through local means (e.g., personal observations) rather than at scale. We overcome these issues using computational data mining techniques (shown in Figure 1) in which learners run computing scripts (e.g., python code) to extract information directly from internet sources (e.g., Twitter, YouTube, etc.). After students identify a meaningful inquiry, whether through explicit instruction, implicit scaffolding, or other means, they proceed to gather data by creating credentials to access social media platforms and collect their data. This involves learners creating credentials to access the social media platform, writing code to authenticate themselves, setting parameters for search queries (e.g., types of data, output formats, etc.) and then exporting that data for analysis. This step in our SDS approach allows learners to collect or scrape large-scale current and historical data sets along their personal interests using topical keywords or hashtags (e.g., #BlackHistoryMonth, #covid-19 #ElPaso, #chihuahuas, etc.). This process is illustrated stepwise in Figure 1.

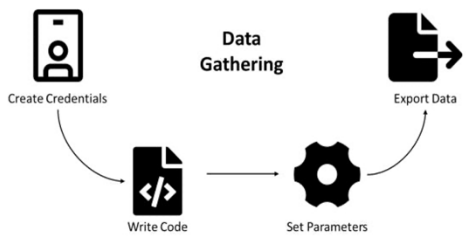

Figure 1: SDS data gathering involves creating user credentials to access web-based platforms, writing code to verify themselves, setting data set parameters (e.g., output data, for-mats, etc.), and exporting that data for additional processing, analysis, and reporting.

2.2 Data Cleaning

Often characterized as "wrangling," the next step, data preprocessing, is used to clean, transform, and reductively organize the collected raw data in a useful and efficient format that can support later quantitative or qualitative data analysis. In SDS, the data pre-processing phase consists of three steps: (1) data cleaning, (2) data transformation, and (3) data reduction. Data cleaning can consist of a wide range of techniques (depending on learner style) to first understand the material nature of their data and then problem solve through any formatting or output issues that might complicate future analysis. Datasets mined from online sources are commonly filled with variations and degrees of "messiness," which affords learners the opportunity to engage with and evaluate raw data closely to understand its structure and identify structural issues. Learners can then develop personalized computational and heuristic strategies/styles to remove incorrect or unnecessary information (cleaning the data) and then organize it into structures suitable for analysis (data transformation). For example, a student interested in tweets that use scientific evidence might first remove tweets that were repeated during the scraping process (cleaning) and later format the tweets so that they are in chronological order (data transformation).

Finally, learners can reduce data structures into forms that contain ordered information consistent with desired attributes. In social media data, metadata about the dataset of tweets could include the number of likes given to each tweet, the number or type of emojis used, content sentiment (e.g., tweets that include the word "evidence"), et cetera. This process of data reduction also involves reducing data "noise" (e.g., content SPAM, output irregularities, etc.). All of this is conducted using programming code to execute functions on datasets that would otherwise be too massive to carry out manually. This process is illustrated stepwise in Figure 2.

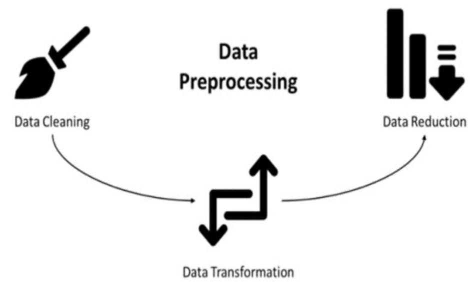

Figure 2: SDS data preprocessing involves learners deploying their own styles and strategies to data cleaning (i.e., removing erroneous information), data transformation (i.e., organizing into structures suitable for analysis), and data reduction (i.e., removing "noise" such as output irregularities or SPAM).

2.3 Data Analysis

Data analysis is a process of changing and processing raw data to extract information that is relevant to a students' line of inquiry. In SDS, this is an iterative process that can include both quantitative and qualitative data sources. The analysis strategies (consistent with SDS) are governed mostly by available data and the questions proposed by learners—in this way analyses are always couched in learner priorities rather than dictated by others. In this phase, relevant data are identified, and analysis conducted.

For qualitative data learners can, using computer code, generate categorical data (e.g., gender, language, presence or absence of a sentiment, etc.) used to inform learner sensemaking or understanding. For instance, learners can generate counts, content weights or even code data to glean content sentiment or innumerable other insights about content. In SDS, learners are drivers of analyses and understandings drawn from primary data sources. By contrast, learners can conduct analysis, also using computer code, with numerical data such as descriptive statistics (e.g., social media content like count mean, medians, modes, etc.) and inferential statistics (e.g., correlations, analysis of variance, etc.) to draw relevant understanding and conclusions (e.g., social media content popularity, demographic-based engagement, etc.). This phase is illustrated stepwise in Figure 3.

2.4 Data Visualization

The fourth phase of our curriculum involves data visualization, or representing data in forms that can readily tell a story about insights gleaned. In SDS, this means that learners can not only choose what representations are used to visualize data but can also shape how narratives about data are conveyed. Like prior steps, learner interests and heuristic styles are supported throughout engagement as contexts and ideas are centered around topics selected. This phase consists of three steps: (1) identifying tools (e.g., appropriate code packages, libraries, etc.), (2) ascertaining data dimensions to be harnessed and shaped by tools, and (3) creating a visualization that abstracts and/or conveys a succinct idea or narrative about the data analyzed. Using python libraries, learners can deploy code to generate seemingly innumerable ranges of visualizations (e.g., word clouds, column charts, line charts, pie charts, scatter charts, etc.) to showcase their insights. This phase is illustrated stepwise in Figure 4.

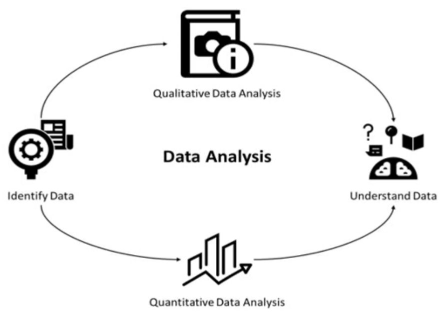

Figure 3: SDS data analysis can involve qualitative and quantitative data sources that learners analyze in ways only limited by the data sets they collect as well as their personal interests. SDS sources are primary and used to guide learner sense-making and understanding.

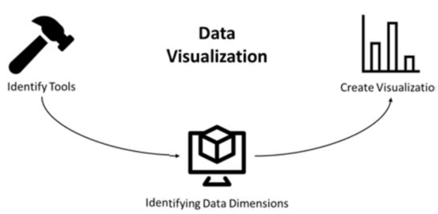

Figure 4: SDS data visualization program involves (1) identifying and leveraging code libraries, (2) assessing data dimensions/material nature and (3) using computational tools to generate a communicative visualization consistent with insights gleaned.

2.5 Data Communication

Phase five of SDS involves data communication phase and consists of activities in which learners come together to construct narratives designed for specific and relevant audiences. Next, learners record or archive data and information such that it is preserved for future use and then that modality is used to report on findings. In other words, SDS gives learners the freedom to select communication modalities (e.g., websites, research papers, video-based public service announcements, portfolios, etc.), preserve data and relevant information in an archivable format (e.g., digital file, etc.) and then use it to report and share with others. Learner artifacts or products also serve to reflect learning outcomes, computational thinking/practice mastery, etc. This phase is illustrated stepwise in Figure 5.

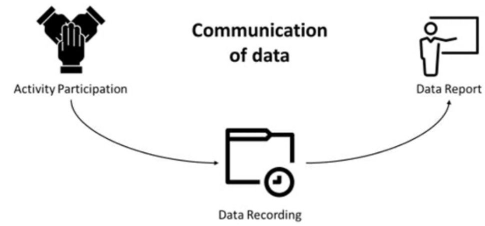

Figure 5: SDS data communication is when learners construct narratives meant to communicate findings to targeted audiences. In addition, this phase involves preserving data in a format that is savable and then reporting on insights using a wide range of modalities—to communicate information and demonstrate learning.

3 Implementation

Here, we highlight an illustrative example (context and method adapted from [22, 23, 24, 25]) of a pursuit we took to data mine Twitter, a popular social media platform. We explored a contemporary issue laden with sociopolitical implications— controversies related to the use of COVID-19 vaccines. This example is meant to illustrate the sort of inquiry learners might carry out using an SDS approach. Consistent with the framework we describe as five phases, we describe the steps we took to gather, preprocess, analyze, visualize, and then communicate ideas regarding public discourse about the COVID-19 pandemic. In this context, we characterized watershed moments—when public government agencies were announcing important breakthroughs about the design of a COVID vaccine.

3.1 Gathering

To generate the initial set of Tweets for the study, we used the Twitter Application Programming Interface (API), and Python 3.8's Tweepy library to download aggregated user data. We used Tweepy's api.search function to identify all Tweets that matched the search string "to:CDCgov." This query identifies all Tweets posted to the platform that are sent to the United States Centers for Disease Control and Prevention up to a limit of seven days before the query was made (i.e., August 6th and August 13th, 2021). The CDC was selected out of convenience because we were knew that this agency had significant engagement from the public about COVID-19 and vaccinations. This made it possible for us to examine Tweets around a uniform topic.

3.2 Cleaning

We then executed python code that selected out data related to usernames of the person or group who posted the Tweet, as well as the full body text of the Tweet itself. The final dataset consisted of about 1479 response Tweets organized into an excel spreadsheet that we could use to review and determine next steps for analysis.

3.3 Analysis

We then conducted analysis of how frequently tweets we mined or scraped from the platform included what we called platform specific features: ampersigns (the @ symbol commonly used to enlist others), emojis (to express sentiments), retweets (to amplify ideas), hashtags (to tag trending topics), and external URLs (used to direct readers to other sites). Our analysis found that of 705 intelligible tweets: 288 included ampersigns, 251 included URLs to other topics, 128 included retweets, 66 included hashtags and 54 included emojis.

3.4 Visualization

Using this data, we also made a network map to understand how some of the objects we described above were connected or occurred in relation to one another. This generated Figure 6. Tweets that used words associated with scientific evidence were more likely to use a wider variety of digital tools to express ideas (such as retweets, ellipses, and URLs). Tweets that did not use these words were more likely to lead with URLs, as well as retweets and emojis.

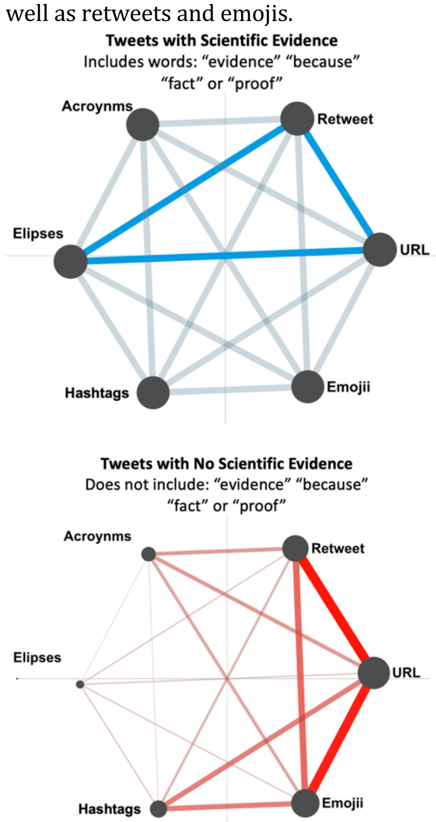

Figure 6: Network maps visualizing the differences in the use of platform-specific features for tweets that included words related to scientific evidence (blue, top) and tweets that did not include scientific evidence (red, bottom).

3.5 Communication

In the data communication phase, we described how our analysis informs our understanding of the way information is being used and shared around COVID-19 vaccines, and how platform tools like ampersigns, hashtags and emojis are used to emphasize or express ideas. We also see that information expressed is sometimes scientific, but not always (examples shown in Figure 7). When it is not scientific, we saw that the ideas remain popular, which is reflected in how many likes the tweet received. This raises concern and caution about how people may be at risk of being misinformed about vaccines on Twitter and other media outlets.

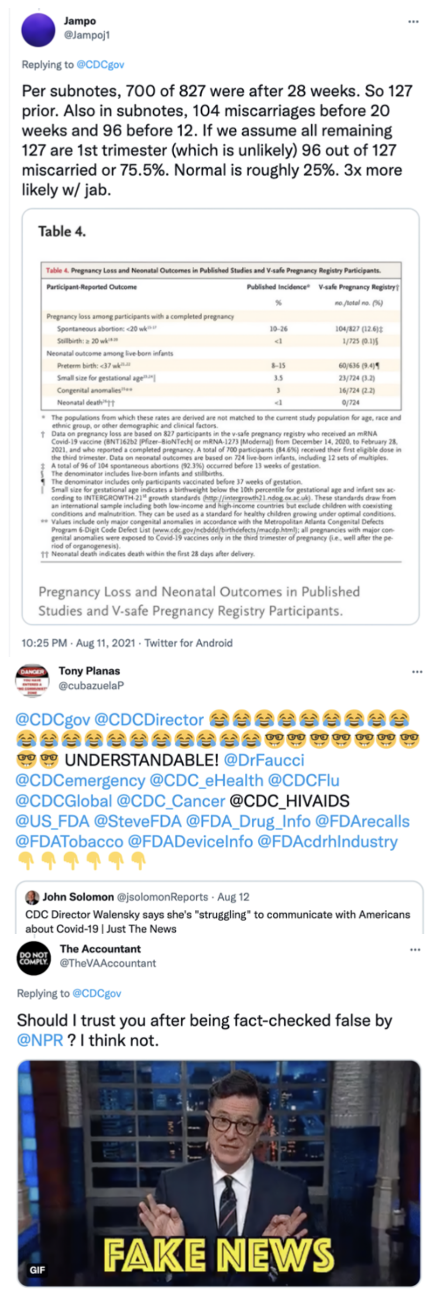

Figure 7: seven (top to bottom). Tweets about COVID-19 vaccine use features to communicate lots of information that is sometimes scientific but sometimes based on personal experiences. Ampersigns and URLs are very common in this topic, which implies that this is an effective way to enlist others or point readers to other sources, which may or may not be scientific. Retweets, hashtags, and emojis seem less common and probably help amplify or express emotion about an idea.

4 KEY IDEAS

Our work to bring SDS to pre-college groups is firmly grounded in our desire to empower learners in ways that have shown significant promise in other forms of learning with computing. In our case, we use computers as a tool for learners to access troves of data using accessible social media platforms which are increasingly ubiquitous among youth and adults alike. This creates a kind of "sandbox" in which learners can explore their own interests. The phases we outline are similar to other forms of active learning (i.e., free or open inquiry) that emphasize agency in pursuit. We present how this can be accomplished in data science that leverages computing. In SDS, learners have significantly more agency in their efforts and are free to do so using strategies, heuristics, styles, and other prior social and cultural resources that they bring to learning environments. This approach honors learners in ways we know are not only culturally responsive, but also empowering. We believe that the insights they gain are liberating as learners pursue ideas that are meaningful to them. In an effort to define our approach as a guiding framework for learning designs that aim to understand or deploy SDS, we highlight the following key features:

4.1 Designing for Multiple Interests and Scales

Because learners have myriad and sometimes competing values, needs and priorities, it is important to design experiences that are relevant to learners and across scales (e.g., personally, socially, and globally relevant). This is a central feature in SDS and enabled because data sources need to be sufficiently robust as to enable learners to explore expansively.

4.2 Designing for Authentic Practice

SDS emphasizes authentic practice, which moves beyond simulated engagement by replicating activities conducted by others. Instead, we emphasize an approach where data science inquiries are open-ended and often unexplored. This means that while engagement can be messy and uncertain, it is also both productive and worthwhile, as it reflects real contexts and real life.

4.3 Designing for Diversity

Recognizing that data science and computing are disciplines that have historically been rife with issues of diversity, equity, and inclusion—we argue that SDS transcends these issues by honoring learners and their diverse heuristic styles. In other words, SDS goes beyond equity approaches that emphasize access and empowerment by making room for learners to engage data sets flexibly along their learning progressions, and this allows for personal agency.

4.4 Designing for Literacies

We recognize that equipping learners with the tools to collect evidence and shape narratives can empower and support field innovation and occupational attainment. There are several literacies, including technical literacy, methodological literacy, and sociopolitical literacy that may benefit from support and merit consideration for further exploration. This is because learners encounter sociopolitical issues that require a form of literacy beyond technical knowledge, which is referred to as sociopolitical literacy. While this paper extensively discusses this area, it is important to acknowledge that ethics in this context exist in a spectrum, encompassing various ethical perspectives. Developing the ability to navigate and incorporate these diverse ethics is a strength in (SDS). This means that learners can pursue their areas of interest while adhering to their own ethical principles.

Creating a framework that facilitates the reconciliation of these diverse perspectives is essential, particularly in a heterogeneous learning environment where individuals hold varying points of view. This aspect of SDS is significant. In terms of civic engagement, it is also imperative to support learner literacies about data science. In SDS, this is fundamental and reflected in activities that encourage learners to engage with data that might sometimes be subjective or in competition. Such chances—to us—mean that learners will have opportunities to gain firsthand insights about why competing viewpoints are dynamic or sometimes ephemeral. It is important to note that the accessibility of data sources can vary over time (e.g., access privileges can change-twitter or the forms of data can vary based on the platform, etc.) and collection methods may need to adapt over time. We view this as an important strength in SDS, because it reflects the most up-to-date practices and challenges in data science collection and as a result learning is situated in authentic practice. Embracing and intentionally engaging students with these accessibility challenges can provide students with a realistic understanding of data science in a dynamic context that models real world practice.

CLDM represents our enactment of SDS to further pre-college educational practice in computing and data science. Our approach is distinct because of its CRC emphasis on learner agency as a means of supporting computing and data science education that is simultaneously productive and agentic.

REFERENCES

[1] Francine Berman, Rob Rutenbar, Brent Hailpern, Henrik Christensen, Susan Davidson, Deborah Estrin, Michael Franklin, Margaret Martonosi, Padma Ragha-van, Victoria Stodden, et al. Realizing the Potential of Data Science. Communications of the ACM, 61(4):67–72, 2018.

[2] William Finzer. The Data Science Education Dilemma. Technology Innovations in Statistics Education, 7(2), 2013.

[3] Dichev and Darina Dicheva. Towards Data science Literacy. Procedia Computer Science, 108:2151–2160, 2017.
[4] Angelina Chen. High School Data Science Review: Why Data Science Education Should be Reformed. Harvard Data Science Review, 2(4), 2020.
[5] Jiang and Jennifer Kahn. Data Wrangling Practices and Collaborative Interactions with Aggregated Data. International journal of computer-supported collaborative learning, 15:257–281, 2020.
[6] Michelle Hoda Wilkerson and Joseph L Polman. Situating data science: Exploring How Relationships to Data Shape Learning. Journal of the Learning Sciences, 29(1):1– 10, 2020.
[7] Davy Cielen and Arno Meysman. Introducing Data Science: Big data, Machine learning, and more, using Python Tools. Simon and Schuster, 2016.
[8] Birte Heinemann, Simone Opel, Lea Budde, Carsten Schulte, Daniel Frischemeier, Rolf Biehler, Susanne Podworny, and Thomas Wassong. Drafting a Data Science Curriculum for Secondary Schools. In Proceedings of the 18th Koli calling International Conference on Computing Education Research, pages 1–5, 2018.
[9] Jeannette M Wing. Computational thinking. Communications of the ACM, 49(3):33–35, 2006.
[10] Koby Mike, Noa Ragonis, Rinat B Rosenberg-Kima, and Orit Hazzan. Computational Thinking in the Era of Data Science. Communications of the ACM, 65(8):33–35, 2022.
[11] Rachel Estrella, Joanna Goode, Jennifer Jellison Holme, and Kimberly Nao. Stuck in the Shallow End. 2008.
[12] Victor R Lee, Michelle Hoda Wilkerson, and Kathryn Lanouette. A Call for a Humanistic Stance Toward K–12 Data Science Education. Educational Researcher, 50(9):664–672, 2021.
[13] Victor R Lee, Daniel R Pimentel, Rahul Bhargava, and Catherine D'Ignazio. Taking Data Feminism to School: A Synthesis and Review of Pre-collegiate Data Science Education Projects. British Journal of Educational Technology, 53(5):1096–1113, 2022.
[14] David C Webb, Alexander Repenning, and Kyu Han Koh. Toward an Emergent Theory of Broadening Participation in Computer Science Education. In Proceedings of the 43rd ACM Technical Symposium on Computer Science Education, pages 173–178, 2012.
[15] Joan Peckham, Lisa L Harlow, David A Stuart, Barbara Silver, Helen Mederer, and Peter D Stephenson. Broadening Participation in Computing: Issues and Challenges. ACM SIGCSE Bulletin, 39(3):9–13, 2007.
[16] Peggy Doerschuk, Jiangjiang Liu, and Judith Mann. Inspired Broadening Participation in Computing: Most Successful Strategies and Lessons Learned. In 2010 IEEE Frontiers in Education Conference (FIE), pages T2H–1. IEEE, 2010.
[17] Gloria Ladson-Billings. Toward a Theory of Culturally Relevant Pedagogy. American Educational Research Journal, 32(3):465–491, 1995.
[18] Tia C Madkins, Alexis Martin, Jean Ryoo, Kimberly A Scott, Joanna Goode, Allison Scott, and Frieda McAlear. Culturally relevant computer science pedagogy: From Theory to Practice. In 2019 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT), pages 1–4. IEEE, 2019.
[19] Mitchel Resnick, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, Amon Millner, Eric Rosenbaum, Jay Silver, Brian Silverman, et al. Scratch: Programming for All. Communications of the ACM, 52(11):60–67, 2009.
[20] Amanda Barany, Sayed Reza, Michael Johnson, Alan Barrera, Omar Badreddin, Crystal Fuentes, and Justice T. Walker. Towards the Design of a Culturally Relevant Curriculum for Equitable Data Mining-based CS Education. In 2023 International Conference of the Learning Sciences Annual Meeting, Montreal, Canada, 2023.
[21] Gayithri Jayathirtha and Yasmin B Kafai. Interactive Stitch Sampler: A Synthesis of a Decade of Research on Using Electronic Textiles to Answer the Who, Where, How, and What for K–12 Computer Science Education. ACM Transactions on Computing Education (TOCE), 20(4):1–29, 2020.
[22] Justice Walker, Amanda Barany, Alan Barrera, Stefan Slater, Omar Badreddin, Sayed Mohsin Reza, and Michael Johnson. Social Media Discourse Instruments as Tools Socioscientific Argumentation. 2023.
[23] Walker, Stefan Slater, and Yasmin Kafai. A Scaled Analysis of How Minecraft Gamers Leverage YouTube Comment Boxes to Participate and Collaborate. 2019.
[24] Justice T. Walker, Amanda Barany, Alan Barrera, Sayed Reza, Karen Guzman Del Rio, Alex Acquah, Omar Badreddin, and Michael Johnson. Sandbox data science: Culturally Relevant K-12 Computing. In 2023 IEEE Frontiers in Education Conference (FIE). IEEE, 2023.
[25] Michael Johnson, Amanda Barany, Alan Barrera, Alex Acquah, and Justice Walker. Lessons Learned from Online Codesign: Exploring reflections of Connected Participatory Strategies for Computing-based Data Science Curricular Design. In Connected Learning Summit, Irvine, CA, 2023.