

Student Participation in Sociocritical Data Literacy: Shapes, Trends, and Future Directions from a Middle School Science Unit

Collette Roberto, Xinyu Wei, Edward Rivero, Michelle Wilkerson

roberto_c@berkeley.edu, weixinyu0820@berkeley.edu, edward_rivero@berkeley.edu, mwilkerson@berkeley.edu
University of California, Berkeley

Abstract: Data science has increasingly integrated sociocritical theories and approaches, helping youth to not only learn data science but also relate data to their everyday and sociohistorical lives. Our project, Writing Data Stories, furthers these efforts by exploring sociocritical data literacies in a large-scale classroom enactment. We examine trends in middle school science student groups' (n=11) *data participation* and *sociocritical participation*, showing how these forms of participation ebb and flow across a 21-day unit. We then present focal group case studies to further unpack how participation shifted over time and suggest what factors contributed to these shifts. We found that data participation was affected by the tools at students' disposal, and sociocritical participation was shaped by the questions groups asked of each other and the data. These findings suggest that special attention to tools and guiding questions is critical when designing for sociocritical data literacy in middle school science contexts.

Introduction

Against the backdrop of an increasingly online and data-driven educational landscape, data science education has become an increasingly important field in its own right (Wilkerson & Polman, 2020). In order to be its most impactful and relevant to students, data science is also seeking to incorporate more critical and humanizing approaches to data (Lee et al., 2021). Writing Data Stories leverages sociocritical literacy (Gutiérrez, 2008; Irgens et al., 2020) to provide rich data science learning, especially in non-dominant youth communities, through integrating scientific inquiry units into the K-12 science classroom. In these inquiry units, students are empowered to explore, visualize, and alter a dataset's contents, structure, or values using the Common Online Data Analysis Platform (CODAP) (Finzer & Damelin, 2016).

A key aspect of the project is supporting students to leverage both everyday and disciplinary genres to make sense of, contextualize, and reauthor datasets (Wilkerson & Polman, 2020). For example, in one of our classroom units, youth were supported to contrast advertising genres used to convince people to consume in the grocery stores with genres used to convince people in science, such as claims, evidence, and reasoning. Students then were invited to take their ideas a step further in exploring a cereal dataset to investigate which factors that might trigger diabetes to a more significant extent, such as sugar, fat, and cholesterol, and try to construct a convincing argument for what is the ideal cereal the school cafeteria should provide to students as their breakfast.

Theoretical Framework

Our work leverages notions of sociocritical literacy (Gutiérrez, 2008), a historicizing literacy that privileges students' sociohistorical lives to point out contradictions and tensions in the relationship between school-and-non-school-based knowledge. Other studies in data science have leveraged similar approaches that have seen success supporting students to critique and reauthor data in ways that invite new sociohistorical understandings of family, migration, and place. For instance, Kahn (2020) examined how students transformed datasets to write intergenerational family geobiographies, making new forms of data learning possible for students and families. Also, Van Wart and colleagues (2020) showed how remapping and envisioning youth's local neighborhoods troubled their relationship to normative data scripts and their relevance or hindrance to community improvement goals.

Building on these studies as well as sociocritical literacy, our study contributes a syncretic approach (Gutiérrez, 2008), which aims to reorganize the relationship between disciplinary knowledge and everyday knowledge as both valuable and needed in the process of knowledge building. We designed syncretic units spanning 21 days of in-school instruction, investigating nutrition, diabetes, and data literacy in a middle school on the west coast of the United States. At the same time, we set the goal to support youth to become agentive data practitioners (Wilkerson & Polman, 2020), primarily through their participation in data moves (Erickson et al., 2019) that inspire students to structure, calculate, filter, and transform datasets. Our preliminary findings showed how middle school students leveraged their everyday knowledge to create new forms of data literacy in the classroom while exploring the nutrition datasets on CODAP.

Research Question

We were interested in how a syncretic design worked in the classroom context and whether youth's everyday knowledge came into a meaningful conversation to create new forms of data literacy, expanding the scientific disciplinary learning of nondominant youth. The research question is: How did youth's everyday knowledge mediate the development of meaningful conversations around data in a middle school science classroom?

Methodology

Various artifacts were collected for the qualitative data analysis, including classroom videotapes, student worksheets, student end-of-day reflections, and screen recordings of the students' Chromebook activity. A codebook was iteratively created to generate themes related to the research question, and there were six overarching subcodes under the two top-level codes, *sociocritical participation* and *data participation*. Based on their prior seating arrangement, 41 students were assigned into 11 small groups (from group A to K) and participated in this study simultaneously. Student participation has been analyzed using six subcodes to investigate how and when they engaged in the unit to develop their sociocritical data literacy across 21 days.

For the sake of space, we only elaborate on the subcodes that were most germane to our case studies. Under the code of *sociocritical participation*, syncretism is defined as using everyday and disciplinary resources to expand both everyday and disciplinary learning. Both had to be expanded to be syncretism, which we call our two criteria of syncretism. Here is a hypothetical example of student activities that would meet these two criteria of syncretism. Say for instance that students are engaged in a task naming food factors that are most important to them. If this group uses everyday ideas such as taste *and* disciplinary ideas such as sugar content on a Nutrition Facts label to choose their factors, they are expanding not only their reasoning about data but also expanding their personal connections to data that they might see in their everyday lives.

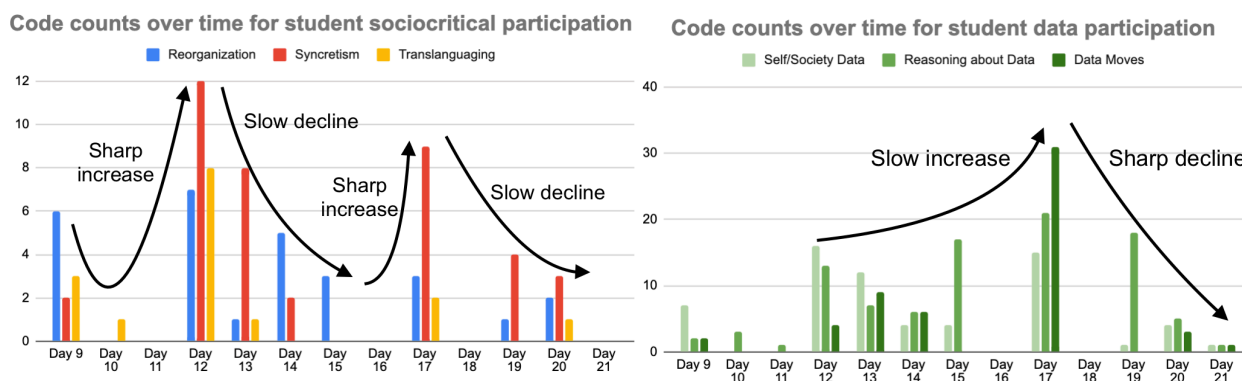
Next, two sub-codes of *data participation* need to be defined as these are examined in the cases. Reasoning about data is defined as students engaging in disciplinary data knowledge and practices already valued in the school context. For example, this code was used whenever students talked about patterns, center/trend/spread, case versus aggregate, construction of datasets. Data moves are coined as students filtering, grouping, sorting, aggregating, or otherwise transforming the data toward new purposes. For example, this code was used when students filtered out cereals on the dataset that were too high in sugar.

Findings

Figure 1 and Figure 2 present subcode counts and chronological trends in groups' *sociocritical participation* and *data participation*. Interesting trends are the spikes in all sociocritical participation in Figure 1 (Day 9, 12, and 17) and the spike in all data participation in Figure 2 (Day 17), which then experience a slow decline over the next few days.

Figure 1 and 2

Focal student groups (n=11) *sociocritical participation* and *data participation* across 21 days



Analysis of key student group cases

Next, we conducted case studies of key groups that seemed to have similar trends in participation to the overall trajectories in participation pictured in one or both the Figures. Analyzing these representative cases can illuminate the possible factors that contribute to these trends, especially those unique to the classroom context. We considered a trend robust if the pattern was present for individual group trajectories for at least two groups.

Case 1: Syncretism decline from the Day 12 spike

Group K was one of two groups (Group J, Group K) whose participation matched the decline in syncretism from Day 12 into Day 15. Students in this group had pseudonyms Jonathan, Diana, and Yazmin. On Day 12, students were asked to discuss what nutrition factors in food were most important to them:

Diana: We have to find two or more important facts.

Dahlia: I think it's "fiber, sugar, and calories" (group agrees).

Diana: Where can you find this information?

Yazmin: The informational label. It'd be...it's on the cereal box huh!

This exchange met both the criteria for syncretism. Reasoning about data was expanded when this group demonstrated an emerging ability to pinpoint data sources in their everyday life (a cereal box that Yazmin was imagining). Personal connections to data was expanded when this group articulated which nutritional factors they were most interested in and began brainstorming where to look up this information.

On the following day, this group saw a decline in syncretism because their activities did not seem to expand reasoning about data even though they made personal connections to data. On day 13, students input foods they ate into CODAP, and group K began discussing how many calories were in each group member's favorite foods. For instance, Diana looked up the calories on her favorite snack, "Takis," and seemed shocked by how much sugar and how many calories are in it. In these moments, although students were learning more about their favorite foods using the language of nutrition facts, their reasoning about data did not seem to expand any further than entering the data into CODAP.

Case 2: Increase into Day 17 of all data participation and syncretism

Group E was one of two groups (Group E, Group J) whose participation matched the increase of data participation and syncretism leading to Day 17. Students in this group had pseudonyms Natalie, Lolita, Alex, and Bryson.

On Day 15, Group E seemed to expand their personal connections to data as related to their everyday lives, but did not necessarily expand their reasoning about data. In the following excerpt, students evaluated a cereal using both nutrition factors and their everyday experiences, creating an opening for syncretism that was not pursued by the group:

Natalie: Oh, this one has no fat and no fiber.

Alex: Which one?

Natalie: Rice krispies.

Bryson: [examining the Nutrition Facts printout] No total fat, no fiber.

Natalie: Doesn't rice krispies taste like nothing? Like... not the bar, but the cereal?

This moment was not quite syncretism because although Natalie was beginning to make personal connections to the data with her question, the group's reasoning about data after this moment remained solely focused on comparing cereals based on nutritional factors. Thus, they did not meet the criteria for syncretism that both everyday and disciplinary participation be expanded. Had they leveraged Natalie's everyday experience to start comparing cereals based on taste and nutrition, these moments might have better aligned with syncretism.

On Day 17, Group E finally did see moments we categorized as full examples of syncretism, and they experienced a substantial increase in data moves. These data moves included adding a new case card for cocoa pebbles (Figure 3), adding new attributes to the dataset such as company, price, and taste, inputting values according to their everyday experience (Taste: 10/10), and leveraging other sources of information beyond CODAP (Google Search, NutritionX website - Figure 4) for other factors (Price, Company). We attribute the increase in data moves to the fact that Day 17 was students' first time using computers, exploring CODAP, and exploring the cereal dataset.

Using these data moves, Group E added a new cereal into the dataset and characterized it using information from both everyday and disciplinary sources. As far as everyday sources, they interviewed each other to decide that the Cocoa Pebbles was neither hot nor cold, but room temperature, that it had an excellent taste, and that the shelf was unknown. As far as disciplinary sources, they searched on Google for typical prices and searched on NutritionX.com for nutritional information (Figure 4).

Figure 3 and 4

Students added a new cereal case on CODAP and searched Cocoa Pebbles on NutritionX.com.

food name	cocoa pebbles
taste	10/10
serving size	0.8 cups cups
calories	120 cal
sugars	10g g
fiber	0 g
sodium	170 mg mg
fat	1 g g
protein	1g g
hot or cold	room temperature
shelf	unknown
company	post
weight of one serving	29g oz
cost	2.99



Cocoa Pebbles

Nutrition Facts

Cocoa Pebbles

Serving Size: cup (29g)

Amount Per Serving	Calories from Fat 9
Calories 120	
% Daily Value*	
Total Fat 1g	2%
Saturated Fat 1g	5%
Trans Fat 0g	
Cholesterol 0mg	0%
Sodium 170mg	7%
Potassium 45mg	1%
Total Carbohydrates 25g	8%
Dietary Fiber 0g	0%
Sugars 10g	

How long would it take to burn off 120 KCal?

Walking (3mph)	32 minutes
Running (6mph)	12 minutes
Bicycling (10mph)	17 minutes

Values estimated based on person weighing 140 lbs. [Login](#) to personalize.

[Find on Amazon](#)

These moments met both the criteria for syncretism. Reasoning about data was expanded when students began to see their personal experiences as valid, inputting ideas like Taste = 10/10 into CODAP. Personal connections to data expanded when students seemed interested enough in their new attribute of cost to continue inputting costs for other cereals in the dataset.

Discussion

This work begins to suggest how sociocritical data literacy takes shape, ebbs, and flows in a middle school classroom context. For instance, Figures 1 and 2 suggest that data and sociocritical participation seem to mutually reinforce one another, such as the spike in both syncretism and data moves on Day 17. Across our focal groups, it was not surprising that data participation was affected by the presence of tools (i.e., CODAP, search engines, and the cereal dataset). However, it was interesting to see how students' sociocritical participation, namely syncretism, seemed largely to be supported or not by the questions students asked of each other and the data. For instance, in Case 1, syncretism emerged when Diana asked where sources of information could be found. Similarly, in Case 2, syncretism emerged when members of Group E interviewed each other about Cocoa Pebbles. These findings are significant because they suggest that additional attention to tools and guiding questions is needed for a syncretic approach to support sociocritical data literacy in a large-scale classroom context. A limitation in this work is that we centered syncretism, but more is needed toward *sociocritical* data literacy. Future work should support analysis of the politics of food, cultural practices, and other necessary elements supportive of nondominant youths' sociocritical literacies (Gutiérrez, 2008).

References

- Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data moves. *Technology Innovations in Statistics Education*, 12(1).
- Finzer, W., & Damelin, D. (2016). Design perspective on the Common Online Data Analysis Platform (CODAP). In *annual meeting of the American Educational Research Association, Washington, DC*.
- Gutiérrez, K. D. (2008). Developing a sociocritical literacy in the third space. *Reading research quarterly*, 43(2), 148-164.
- Irgens, G. A., Dabholkar, S., Bain, C., Woods, P., Hall, K., Swanson, H., ... & Wilensky, U. (2020).
- 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.
- Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A Call for a Humanistic Stance Toward K–12 Data Science Education. *Educational Researcher*, 0013189X211048810. Modeling and measuring high school students' computational thinking practices in science. *Journal of Science Education and Technology*, 29(1), 137-161.
- 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*, 29(1), 127-153.
- Wilkerson, M. H., & Polman, J. L. (2020). Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences*, 29(1), 1-10.