Proposing a Role-Based Framework for Data Literacy

Zak Risha, University of Pittsburgh, zak.risha@pitt.edu
Aditi Mallavarapu, University of Pittsburgh, aditimal@pitt.edu
Rosta Farzan, University of Pittsburgh, rfarzan@pitt.edu
Jaime Booth, University of Pittsburgh, jmbooth@pitt.edu
Beth Sondel, Independent Scholar, bethsondel@gmail.com
Erin Walker, University of Pittsburgh, eawalker@pitt.edu

Abstract: Efforts to cultivate critical data literacies in youth calls for further research that emphasizes supporting collaboration in team environments. We propose a role-based framework of four roles (Journalist, Detective, Scientist, and Artist) that seeks to go beyond just facilitating data analysis by representing the multitude of ways data literacy plays a role in people's lives. We pilot our role-based curriculum within an advocacy-focused summer program to study how our approach facilitates collaboration and critical data literacies.

Introduction

The advent of Big Data has spurred new theories and practices to prepare a new generation of learners for a future increasingly reliant on the epistemic powers of data. Given data's growing importance for communicating issues and establishing credibility, learners' ability to both comprehend and critique data could be fundamental to bringing about change in the world. Scholars have drawn comparisons between data and the emancipatory capabilities of literacy, extending aspects of *critical literacy* to form the theory of *critical data literacy*. While this theory establishes objectives of using data to empower and transform, the methods for cultivating critical perspectives in learners are not yet well established. Similarly, data literacy and science are often collaborative in practice, calling for a greater understanding of how to support data literacy in team contexts.

Data science teams can be composed of members with a variety of abilities, such as technical, narrative, theoretical, and visual, which can complicate how data literacy is framed and taught. Such diverse competencies call for frameworks that do not simplify data literacy or prioritize one dimension, but rather highlight the multifaceted nature of the domain. Critical data literacy itself has been decomposed into the technical abilities for using data and the critical capacity to understand the context surrounding data and the reality it portrays (Tygel & Kirsch, 2016). Forming critical insights requires going beyond technical or statistical knowledge sometimes favored in curriculum.

Accordingly, our main contribution was to develop and propose a role-based framework that can better represent the various elements of data literacy, while also supporting collaboration and fostering critical data literacy. These disciplinary roles aim to highlight pathways for contribution that can appeal to a wide range of learners and help them engage with the context surrounding data. Curriculum that breaks down data science into more recognizable disciplinary roles can better align with learners' reality, promoting the critical, and social aspects of data that provide various ways to co-construct knowledge. We report on the deployment of our role-based curriculum in an advocacy-focused summer program. Our primary research goal was to investigate how breaking data science into disciplinary roles facilitates collaboration and critical data literacies.

Critical Data Literacy

There has been growing interest in exploring how to support learners' critical engagement with data in the learning sciences community. Such an interest was well represented in a recent special issue of the Journal of Learning Sciences (Wise, 2020) as well as a symposium at ICLS '20 (Irgens et al., 2020) which specifically called for more research supporting critical sociocultural perspectives. To better engage learners and demonstrate a direct relationship with data, some scholars have developed curriculum around personal data. Such examples include data storytelling using personal data collected from daily lives (Stornaiuolo, 2020) or exploring family geobiographies and personal family migration stories (Kahn 2020). However, these studies noted the need for more explicit support or additional interactions to tackle issues of injustice or inequality (Stornaiuolo 2020; Kahn 2020). In contrast to personal data, which has a more individualistic focus, other scholars have utilized local community data which provides more opportunities for collaboration. Van Wart et al. (2020), described two participatory mapping activities where youth worked together to tackle local community challenges such as air quality and park planning. Similarly, Rubel et al. (2017) describes a classroom module where students investigate their city's distribution of various financial institutions (e.g., pawn shops and banks). We build on this work to explore how a collaborative framework related to data literacy roles can guide interactions in this context.

Roles to Play in Data Literacy

Recent research has also explored the different facets of data science and literacy in curriculum, helping describe the various roles that can be played in collaborative settings. Such approaches highlight the interdisciplinary nature of data by integrating other creative fields, such as art and journalism. An arts-centric approach called *Data Murals* (Bhargava et al, 2016) engaged the community to reflect on their data and create a visual design that tells a data-driven story. The project, informed by critical pedagogy, aimed to utilize art's history of helping people engage and critique power structures in their lives. D'Ignazio (2017) further described such interdisciplinary uses of data as *creative data literacy*, which highlights non-technical aspects of data. Another related study integrated aspects of journalism by having students generate infographics in a science classroom (Gebre & Polman, 2020). Students were able to creatively communicate with statistics, labeled figures, and data visualizations about a topic of interest. These examples helps explore and inform the various roles youth can play when working with data.

While new in the context of data literacy, role theory has long been used as a framework for explaining characteristic behavior patterns (Biddle, 1986). The theory grew in popularity in collaborative learning for studying both naturally emerging roles as well as scripted roles to facilitate collaborative processes (Strijbos & Weinberger, 2010). In an effort to engage learners in interdisciplinary learning and integrate science and literacy, Jiang et al. (2019) implemented "discipline-specific roles," a type of functional role related to academic disciplines. In the study, students could assume the role of writer, scientist, artist, or engineer to produce multimodal science fiction stories in small groups. Though not yet applied to the domain of data literacy, using disciplinary roles could help capture the many dimensions of using data.

Curriculum and Framework Development

We developed our role framework by generating 8 aspects that are encompassed in using community data science for advocacy: narrative, artistic, investigative, activist, technical, making, scientific, and exploratory. Our next step was to consolidate aspects into roles by minimizing redundancy and preserving diverse ways to participate and engage with data. The roles of *artist* and *journalist* were conceived to cover the narrative, artistic, and making aspects of data. For the remaining aspects of investigative, technical, scientific, and exploratory we created the *detective* and *scientist* roles. Last, we reasoned that activism was an aspect all roles could adopt. We piloted initial role concepts in a youth advisory board, consisting of 3 high school youth from underrepresented groups in data science. Sessions were focused on understanding their attitudes and impressions on proposed disciplinary roles and identifying technologies to aid in data analysis. Using local datasets and data analysis software, youth assumed roles to work collaboratively with their own individual role-related objectives (such as forming a story). Feedback on the role and experience was collected through discussions at the end of each activity. The advisory board helped refine role concepts and identify Tableau (https://www.tableau.com), an interactive visualization and analytics software, as a suitable technology due to the affordances for tinkering and user-friendly interface.

After roles were finalized, we began planning to implement them in a broader curriculum for an advocacy-focused summer program. The research team mapped learning objectives of the curriculum to each role and developed four sessions, each focusing on a single role. Our goal for the sessions was to introduce each respective role and develop the proficiencies needed to enact them in a collaborative context. Each session began with instructional lessons among all youth covering learning objectives and role competencies which was followed by a collaborative activity in smaller groups with facilitators. Our four role-based sessions were:

- Journalist: Exploring personally relevant data (e.g. popularity of one's name over time), sharing results in groups, and writing headlines to communicate the most interesting insights.
- Detective: Learning how to explore datasets with Tableau through an interactive tutorial. Working in groups to find interesting insights in data using summary statistics (counts and averages) and charts.
- Scientist: Watching a video that introduced statistical hypothesis testing and confidence intervals. Applying confidence intervals in Tableau to draw insights from their own data.
- Artist: Introductory activity looking at examples of data art and engaging in sketching to represent data

The sessions were designed to empower the participating teens to use data to inform the advocacy campaigns they developed to redress systemic injustice in their communities, wherein all activities were designed to build towards their final projects (Figure 1). An additional fifth session was planned where students assumed a role according to their preference and collaborated to incorporate prior work with data into their own advocacy campaign.

Summer Program Curriculum

All sessions were in the context of a partnership with a non-profit organization, The Women and Girls Foundation (WGF). For the past fifteen years, WGF has run GirlGov, which is a civic engagement and social justice advocacy

program for high school-aged women and femmes. The primary goal of GirlGov is to train teens to translate their passion for social justice and equity into tangible systemic change. The GirlGov curriculum is rooted in critical theory and community organizing and aims to develop teens' civic knowledge, critical consciousness, connection to community, and confidence as they develop an advocacy campaign around an issue they are passionate about. First, the teens break into groups based on their passions and identify an issue they want to address. After they identify an issue, they participate in a series of lessons, activities, and sessions to research root causes, learn from experts and other advocates, identify a SMART goal or intended objective of their campaign, participate in power-mapping to identify who holds the levers of power to create change, and identify the multiple strategies they will use to convince those in power to change policies. There were also additional opportunities for teens to use the arts and creative expression towards activism.

Program Deployment

The summer program occurred over three weeks and lasted 3.5 hours per day (4 days a week). In coordination with program leaders from GirlGov, we designed five sessions, each lasting approximately 75 minutes, occurring twice a week for the first two weeks and once for the final week. Due to the ongoing effects of COVID-19, all but the first of our sessions happened remotely utilizing video conferencing software and breakout rooms for small group work. As part of the program registration application, the youth provided advocacy-based topics of interest; for example, "education and desegregating school", "racial equality", or "school-to-prison pipeline." Based on these topics, the researchers identified relevant datasets which they introduced in an initial session with the youth. The GirlGov team then held a session for teens to discuss their personal connections to each of the data sets and share examples of the issues in their own lived experience. Based on these discussions, youth formed their own groups around two prominent issues: *the school to prison pipeline* and *teacher and student racial disparities in high schools*. The first dataset contained records of juvenile referrals in the local county. The second described the school demographics of public schools in the county. A total of 24 students enrolled in the program with session attendance varying from 17 to 23 students (ages 14-17). For smaller breakout sessions, each advocacy group was divided further within the two topics by program leaders to create 4 subgroups (4-6 youth) for collaborating and 1-2 facilitators (including research team and GirlGov staff), wherein the youth worked on their selected dataset. Sessions and breakout rooms were recorded, transcribed, and de-identified for analysis

Figure 1
Collage of youth's work developed for social media posts or final presentations that pertain to advocacy campaigns focusing on either the school to prison pipeline (a) or teacher and student racial disparities (b).
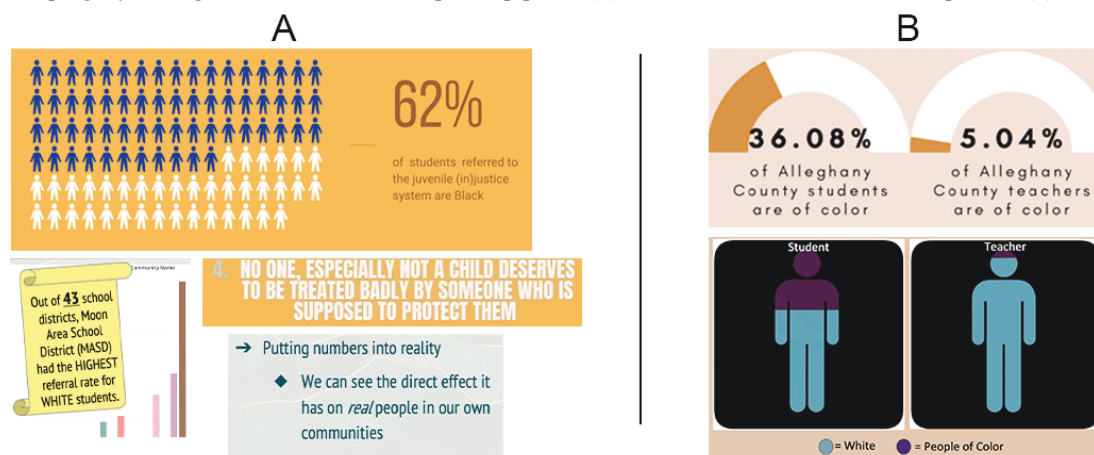
Results and Discussion

From a curriculum perspective, there was promising evidence that framing data science around roles facilitated critical engagement with data by helping youth form insights about reality, consider the context surrounding data, and use data to enact change. Specifically, roles were unique in which critical aspects they most fostered. For example, the Artists and Journalist roles promoted creative ways to connect with the human element of data while the Detective and Scientist roles supported critical reflection on how the data was generated. Such diversity suggests a role-based approach could provide more affordances for students to critically engage with data.

However, our analysis struggled to draw conclusions on how assuming a specific role facilitated collaboration. While roles seemed to demonstrate ways in which students could contribute, they did not always lead to clear division of labor or responsibilities. Groups often focused on a single role at one time, leading youth to sometimes shift to that role. This shifting, similarly observed by Jiang et al. (2019), could be embraced by a role framework where roles responsibilities are modeled as leading and coordinating the group in a task. For example, Detectives could be prompted to lead the group through data analysis, rather than conduct it alone.

Out of all roles, the Scientist was the least represented in the work produced by students. It proved the most challenging role to implement, as students struggled to latch on to ideas of statistical hypothesis testing. This was not surprising, as the role required more specialized knowledge making it the least intuitive to youth. Similarly, the term Scientist might draw connotations from natural sciences or chemistry, which differ in practices from data scientists. Furthermore, there may have been less clarity into how this role was to contribute to a final deliverable. Although fundamental to data literacy and data science, a role centered around Scientist may need to be reframed around a less statistical curriculum to be more intuitive and easier to engage with for youth.

Future Work

Our future work will further analyze the results of this deployment to support the refinement of our role-based data science curriculum. Our analysis will also attempt to describe collaborative aspects of the program, such as inter-role collaboration, and potential methods for future support. Insights will inform our plans to design and develop a collaborative learning system specifically for facilitating role-based critical data literacies in youth.

References

Bhargava, R., Kadouaki, R., Bhargava, E., Castro, G., & D'Ignazio, C. (2016). Data murals: Using the arts to build data literacy. *The Journal of Community Informatics*, *12*(3), 197-216.

Biddle, B. J. (1986). Recent developments in role theory. *Annual Review of Sociology, 12*(1), 67–92.

D'Ignazio, C. (2017). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, *23*(1), 6-18.

Gebre, E. H., & Polman, J. L. (2020). From "context" to "active contextualization": Fostering learner agency in contextualizing learning through science news reporting. *Learning, Culture and Social Interaction*, *24*, 100374.

Jiang, S., Shen, J., & Smith, B. E. (2019). Designing discipline-specific roles for interdisciplinary learning: Two comparative cases in an afterschool STEM+L programme. *International Journal of Science Education, 41*(6), 803–826.

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.

Irgens, G. A., Simon, K., Wise, A., Philip, T., Olivares, M. C., Van Wart, S., Vakil, S., Marshall, J., Parikh, T. S., Lopez, M. L., Wilkerson, M., Gutiérrez, K., Jiang, S., & Kahn, J. (2020). Data literacies and social justice: Exploring critical data literacies through sociocultural perspectives. In Gresalfi, M. and Horn, I. S. (Eds.), *14th International Conference of the Learning Sciences (ICLS) 2020 on the Interdisciplinarity of the Learning Sciences* (Vol. 1; pp. 406–413). International Society of the Learning Sciences.

Rubel, L. H., Hall-Wieckert, M., & Lim, V. Y. (2017). Making space for place: Mapping tools and practices to teach for spatial justice. *Journal of the Learning Sciences, 26*(4), 643–687.

Stornaiuolo, A. (2020). Authoring data stories in a media makerspace: Adolescents developing critical data literacies. *Journal of the Learning Sciences, 29*(1), 81–103.

Strijbos, J. W., & Weinberger, A. (2010). Emerging and scripted roles in computer-supported collaborative learning. *Computers in Human Behavior, 26*(4), 491–494.

Tygel, A. F., & Kirsch, R. (2016). Contributions of Paulo Freire for a critical data literacy: A popular education approach. *The Journal of Community Informatics*, *12*(3), 108—121.

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

Wise, A. F. (2020). Educating data scientists and data literate citizens for a new generation of data. *Journal of the Learning Sciences*, *29*(1), 165–181.

Acknowledgements