Imagining Future Designs of Tools for Youth Data Literacies

Workshop at Connected Learning Summit July 2021

Regina Cheng, Stefania Druga, Emilia Gan, Catherine D'Ignazio, Rahul Bhargava, Victor Lee, Camillia Matuk, Tamara Clegg, Andee Rubin, Yasmin Kafai

1. Introduction

In today's data-driven world, there is little doubt that the ability to understand and use data has become crucial not only for professional data scientists but also for everyday people [5, 12]. Argued by many as a new form of literacy, *data literacies*, or data science related skills, have been brought up as essential competences that young people should acquire [10, 14]. Central to the idea of data literacies are the technical abilities to read data (e.g., understanding what a dataset represents), create data (e.g. generate a dataset), work with data (e.g., cleaning a dataset), analyze data (e.g., filtering data), and argue with data (e.g., visualizing data to support a claim) [6, 17], as well as the critical skills to access and interpret data through the lens of community and social impact [9, 11, 15].

Despite increasing interest and the broad applicability of data literacies, there are few widely used tools specifically designed to assist young learners in developing data literacies. Popular data analytic tools such as Tableau and Excel are designed for the very purpose of "using" rather than "learning" [4]. These tools often focus on generating outputs (e.g., visualizations), rather than scaffolding learners' data analysis and form meaningful arguments from data.

To address this need for learning-focused tools, we propose this workshop to *imagine and propose future tools that could help youth develop data literacies*. We will first present workshop participants with a list of existing tools. Some of these tools, such as CODAP[1] and DataBasic.io [4], were specifically designed to help young people work with and understand data. Others are tools that, while not specifically designed with learning in mind, may still contribute useful functionality. We will invite participants to explore and evaluate these tools. Building on their inquiries with the existing tools, participants will then brainstorm and design new tools and features. By the end of the workshop, we hope to gather a set of novel ideas for features that could inform the design of future learning technologies for data science. Due to the scope of our workshop activities, we propose to split the schedule into two 1-hour slots on consecutive days, 1 hour for tool exploration and 1 hour for brainstorming future designs (we have permission from the conference committee for this extension).

Our workshop design is driven by existing research projects on youth data literacies. Decades of research on the design of constructionist tools for promoting data and broader computational literacies summarize the principles of "low floors" for novice entry, "high ceilings" for skill advancement, "wide walls" for exploration, and "open windows" for social connection [7, 13]. Similarly, other studies (e.g. D'Ignazio and Bhargava [4]) suggest that learner-centered data analytic systems should offer focused, guided, inviting, and expandable tasks and scaffolds. In addition to developing technical skills, learners also need support to understand limitations in data collection, to recognize bias and complexity in data, and to produce insights and artifacts with local communities for social impact [1, 3, 8, 9, 11, 15]. Prior work recognizes that it is also important to allow learners to work with data that is personally meaningful and embedded in daily, familiar activities [2, 14, 18, 16], and to be exposed to data variability, inference, aggregate analysis and various methods of visualization [17, 19]. These previous studies will help us synthesize design recommendations and user scenarios that will guide the activities in our workshop.

2. Workshop Procedures

2.1. Target Participants and Recruitment

We hope to recruit 25 - 30 participants of varying backgrounds and experiences with data literacies for young learners. This includes academic researchers interested in data literacies and digital youth, teachers

[1] https://codap.concord.org/

who have experience or are planning to teach classes about data, after school program educators who organize data learning activities for youth, and any other practitioners who are interested in youth data literacies. We will recruit participants through social media and the co-organizers' networks.

2.2. Workshop Design

We will split the workshop into two 1-hour slots on two consecutive days. All interactive workshop activities will be hosted on Miro[2], a visual collaborative platform, and Zoom will be used for video and voice communication. Given the time constraints, participants will be invited to get familiar with the interface of Miro and create a visually represented personal introduction on Miro prior to the workshop. At the beginning of *Day 1* of the workshop, participants and co-organizers will work together to determine how they define data literacies in order to ground future discussion.

The main activity during *Day 1* will be the exploration and evaluation of existing tools that could be used by youth to learn with and about data. We will present participants with a list of existing online tools (accessible by URLs). We will split the tools into groups based on the aspect of data literacies each tool can scaffold. For example, we can have groups of tools that help kids understand, work with, analyze, and argue using data, respectively. We will survey participants' experience and interest prior to the workshop and assign them to the group with tools of their interest. For each group, we will provide participants with a prompt that involves a potential persona and a user scenario. For example, in the section for analyzing data, the prompt would be: *"Sara, a sixth-grader who just learned basic concepts about central tendency, is trying to study her own preference on lyrics of popular songs. How could tools such as DataBasic.io or Excel help her?"*. Participants will be asked to explore the features, function, and interface design of the group of tools and critically comment on tool features that are helpful, unhelpful, or missing. To facilitate this, we will provide participants with scaffolding such as design recommendations and checklists drawn from literature. After the activity, we will ask each group to present and reflect on its observations and comments. The goal for *Day 1* is to get participants familiar with existing designs, allowing them to think through what features and concepts could be modified and/or added in future designs. At the end of *Day 1*, we will provide participants with a written overview of the design activity in *Day 2* and encourage them to brainstorm and sketch in advance.

In *Day 2*, the main activity will be to brainstorm and design new features to support young people's learning about data. Before the activity, we will prepare a set of visual patterns in Miro such as basic UI elements, shapes, and images which can be used directly in the design activity. We will start with a short warm-up activity to introduce participants to the design affordances and to get everyone into a creative mind frame. After the warm-up, we will once again divide participants into groups where they will brainstorm and design features to help kids understand, work with, analyze, and argue with data, respectively. For each group, we will provide participants with a design prompt featuring a persona and scenario that will build on the prompts from *Day 1*. For example, "*How can we design a tool to help Sara learn from her analysis?*." To scaffold their designs, we will present participants with the notes and design recommendations from *Day 1*. We will encourage participants to bring their experience of research, teaching, and working with youth into their designs. During the brainstorm and design activity, participants will be able to create graphical design solutions in Miro by using the prepared visual elements or by importing pictures of their own sketches. They will also be able to note down explanations and general design ideas. Towards the end of the workshop, we will invite participants to reflect on their designs and experience, and to synthesize a list of design guidelines for future data literacy tools. After the workshop, we will organize the designs and takeaways into a report, sharing it with all the participants and potentially writing it up as a white paper.

2.3. Timeline

Preparation: Organizers surveying participants' interests; Participants reading description of the workshop and instructions for using Miro; Organizers and participants making personal introductions on Miro.

Day 1: Introduction & overview (15 mins); Tool exploration and evaluation activity in groups (30 mins); Discussion (15 mins). Day 2: Overview and warm-up (5 mins); Brainstorming and design creation (40 mins); Discussion and wrap-up (15 mins)

[2] https://miro.com

REFERENCES

[1] Rahul Bhargava et al. "Data murals: using the arts to build data literacy." In: *The Journal of Community Informatics* 12.3 (2016).

[2] Tamara Clegg et al. "Data Everyday: Data Literacy Practices in a Division I College Sports Context." In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–13.

[3] Catherine D'Ignazio. "Creative data literacy: Bridging the gap between the data-haves and data-have nots." In: *Information Design Journal* 23.1 (2017), pp. 6–18.

[4] Catherine D'Ignazio and Rahul Bhargava. "DataBasic: Design principles, tools and activities for data literacy learners." In: *The Journal of Community Informatics* 12.3 (2016).

[5] Catherine D'Ignazio and Lauren F Klein. *Data feminism*. Mit Press, 2020.

[6] Catherine D'Ignazio and Rahul Bhargava. "Approaches to Building Big Data Literacy." In: *Proceedings of the Bloomberg Data for Good Exchange Conference 2015*. Bloomberg Data for Good Exchange Conference. New York, N.Y., Sept. 28, 2015.

[7] Sayamindu Dasgupta and Benjamin Mako Hill. "Scratch Community Blocks: Supporting Children as Data Scientists." In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. New York, New York: ACM Press, 2017, pp. 3620–3631. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025847.

[8] Erica Deahl. "Better the data you know: Developing youth data literacy in schools and informal learning environments." In: *Available at SSRN 2445621* (2014).

[9] Juliana Elisa Raffaghelli. "Is Data Literacy a Catalyst of Social Justice? A Response from Nine Data Literacy Initiatives in Higher Education." In: *Education Sciences* 10.9 (2020), p. 233.

[10] Andreas Grillenberger and Ralf Romeike. "Developing a Theoretically Founded Data Literacy Competency Model." In: *Proceedings of the 13th Workshop in Primary and Secondary Computing Education*. WiPSCE '18. Potsdam, Germany: Association for Computing Machinery, 2018. ISBN: 9781450365888. DOI: 10.1145/3265757.3265766.

[11] Samantha Hautea, Sayamindu Dasgupta, and Benjamin Mako Hill. "Youth Perspectives on Critical Data Literacies." In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. (forthcoming). New York, New York: ACM Press, 2017, pp. 919–930. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025823.

[12] Benjamin Mako Hill et al. "Democratizing Data Science: The Community Data Science Workshops and Classes." In: *Big Data Factories: Scientific Collaborative Approaches for Virtual Community Data Collection, Repurposing, Recombining, and Dissemination*. Ed. by Nicolas Jullien, Sorin A. Matei, and Sean P. Goggins. New York, New York: Springer Nature, 2017, pp. 115–135.

[13] Yasmin Kafai. "Connected Gaming: An Inclusive Perspective for Serious Gaming." In: *International Journal of Serious Games* 4.3 (Sept. 2017). DOI: 10.17083/ijsg.v4i3.174.

[14] Victor R Lee and Joel Drake. "Quantified recess: Design of an activity for elementary students involving analyses of their own movement data." In: *Proceedings of the 12th international conference on interaction design and children*. 2013, pp. 273–276.

[15] Camillia Matuk et al. "Data Literacy for Social Justice." In: *International Society of the Learning Sciences (ISLS)* (2020).

[16] Yim Register and Amy J Ko. "Learning Machine Learning with Personal Data Helps Stakeholders Ground Advocacy Arguments in Model Mechanics." In: *Proceedings of the 2020 ACM Conference on International Computing Education Research*. 2020, pp. 67–78.

[17] Andee Rubin. "Learning to Reason with Data: How Did We Get Here and What Do We Know?" In: *Journal of the Learning Sciences* 29.1 (2020), pp. 154–164.

[18] Andee Rubin, James Hammerman, and Cliff Konold. "Exploring informal inference with interactive visualization software." In: *Proceedings of the Seventh International Conference on Teaching Statistics*. International Statistical Institute Voorburg. 2006.

[19] Andee Rubin and James K Hammerman. "Understanding data through new software representations." In: *Thinking and reasoning with data and chance* (2006), pp. 241–256.