The Nature of Data Literacy and Data Science-Related Activities in K-12 Science Classrooms

Hadia Aslam, Simon Fraser University
Dr. Engida Gebre, Simon Fraser University

Research on data literacy and data science education has been increasing in recent years. Much has been written about the importance of teaching and learning with and about data, focusing on open data, big data, data literacy, and data science across disciplines (Wilkerson & Polman, 2020). Data science is viewed as a field that aims to solve real-world problems using large amounts of data, integrating skills from disciplines such as mathematics, computer science, statistics, social science, and art (Frischemeier et al., 2021). A consensus definition of data literacy does not exist; however, the most common way of defining data literacy in the literature focuses on developing the competencies of learners so that they can deal with, or work with, data (Bowler et al., 2017; Gebre, 2022). It is important to teach data-related competencies as part of the curriculum in schools to ensure that our students achieve the required level of data literacy to address personal and community challenges, and recognize the relevance of data for informed decision-making, civic engagement, and data citizenship (Robertson & Tisdall, 2020; Wolff et al., 2016).

Despite the growing interest in better preparing K–12 students to work with data (Lee et al., 2021), there have been few substantial changes in curriculum and instructional practices aimed at fostering data literacy in K–12 education (Gebre, 2018). Finzer (2013) pointed out that K–12 education does not provide meaningful learning experiences designed to develop students' understanding of data science concepts or fluency with data science skills. Philip et al. (2016) reasoned that the cross-disciplinary knowledge and skills needed to develop data related competencies in students are not fluidly integrated into traditional K–12 subject areas. Furthermore, secondary school students' understanding of data is primarily limited to their experiences in school contexts, such as working with surveys, experimental data, or interpreting graphs and charts (Bowler et al., 2017; Gebre, 2018). A close examination of the nature of designed interventions and related publications highlights a way of emphasizing quantitative ways of data handling skills and the use of related technological tools. This narrow approach and uncritical framing of the concept, limit literacy development in the age of data, and constrain what students can possibly achieve in data-rich environments (Gould, 2017). In response to the nature of current approaches to teaching data in schools, there is an increasing call for broader framing and context-oriented perspectives on data literacy and data science education (Gebre, 2018; Lupton, 2017; Markham, 2019, 2020).

This study responds to recent calls for gaining a deeper understanding of data literacy and data science in order to design K–12 learning environments that prepare learners for navigating an increasingly datafied world (D'Ignazio & Bhargava, 2020; Garcia et al., 2021). This paper examines the similarities and differences in how data science and data literacy are conceptualized and implemented in K–12 science classrooms. More specifically, this study answers the question: "In what ways are the conceptualizations and implementations of data science and data literacy different or similar in the context of science education?" Answering this question will provide insights about how both fields intersect in the context of science education. The findings of this study could be significant for teachers and researchers seeking to enhance their understanding of existing data practices in science classrooms. Addressing this question

will also help raise awareness about the evolving role of data in everyday life, and aid in developing context-oriented curricular interventions that promote learning with and about data for young adults.

Method

This study reviewed the existing literature to examine the range of conceptualizations about data literacy and data science in order to develop a comprehensive understanding of these concepts within the context of K–12 science education. In January 2022, we conducted an abstract search in EBSCOhost, Education Source, ERIC, APA, and PsycINFO using the terms "data literacy" OR "data science" AND ("science education" OR "science teaching" OR "science learning" OR "science instruction") AND (K–12 OR "elementary education" OR "secondary education"). This search identified 39 results for data literacy and 165 results for data science from a combination of academic journals, reports, and dissertations. We then used three criteria for inclusion/exclusion: the study should be (a) empirical, where researchers designed intervention; (b) in the area of science education; and (c) in the K–12 context. This parameter limited the search results to 31 papers: 19 papers on data literacy, and 12 papers on data science. We then examined, evaluated, and coded the papers individually.

Results

The results present our analysis of how data literacy and data science have been conceptualized in the context of science education in K–12 classrooms. We first tried to capture how the literature defined both concepts in the specific context of science education. This was not very successful because most researchers did not define the terms data literacy and data science. When definitions were provided, they were often borrowed from other disciplines outside of science education. Robertson and Tisdall (2020), adapting from Data-Pop Alliance (2015), defined data literacy as "the desire and ability to constructively engage in society through and about data" (p. 8). Similarly, Kahn (2020) conceptualized data science education as education aiming to generate innovative and ethical solutions to societal problems with data technologies. So, both domains share the same focus (i.e., society), but they have different levels of granularity about the nature of data sets. Highlighting the conceptual similarities between data literacy and data science, Kjelvik and Schultheis (2015) argued that data literacy is at the intersection of quantitative reasoning and data science, and both are rooted in real-world contexts. Instruction in either area will likely support the other (Prado & Marzal, 2013). Further similarities between the nature of data science and data literacy-related activities are discussed below.

Inquiry-based Approach

The process of inquiry is at the core of science education. Seven publications adopted an inquiry-based approach to learning in science classrooms (e.g., Forster et al., 2018; Gebre, 2018; Harris et al., 2012; Hug & McNeill, 2008; Maltese, 2015; Ucar & Trundle, 2011; Wolff et al., 2016). Inquiry learning is a process in which learners diagnose problems, formulate hypotheses, conduct investigations, and communicate coherent arguments (Pedaste et al., 2015). For example, in a data literacy intervention, Wolff et al., (2016) used inquiry-based instruction to engage students in learning about energy consumption by interpreting the visualization of smart

meter data. Following a similar inquiry-based approach, in a data science education intervention, eighth-grade students wore physical activity monitor devices for approximately four months to collect and analyze personal data as a way of making sense of the world (Ching & Hagood, 2019).

This highlights that for K–12 science students, phenomena and problems should be compelling to figure out (Lee & Campbell, 2020). Students should be able to see clear connections between scientific data and the real world they inhabit. So, the nature of activities in both data literacy and data science focuses on real-world data, and students figure out complex real-world phenomena and problems as the problem space of their pursuits, such as pollution in local streams and issues of water availability, quality, and use.

Personal Relevance of Data

We noticed a focus on personal contexts in six publications related to data literacy and data science education (Ching & Hagood, 2019; Gebre & Polman, 2020; Kahn, 2020; Lee et al., 2021; Lupton, 2017, 2018; Stornaiuolo, 2020). High school biology students worked on a project of their own interest and produced data-informed posters in an intervention designed to promote data literacy (Gebre & Polman, 2020). Similarly, Kahn (2020) introduced family geobiography as a new personal context for data science education, where high school students used large-scale data sets and dynamic visualization tools to explain why their families moved.

Using personal data in these interventions reframes students' understanding of themselves as data agents and expands what they understand as data and its purposes (Stornaiuolo, 2020). Secondly, students who are authors and architects of data are better equipped to recognize its potential and limitations than those who only engage with descriptive and inferential analysis using structured data (Gebre & Polman, 2020).

Authentic Contexts

Our analysis revealed that data science education interventions aimed to engage students in authentic science tasks where they had the opportunity to work with the same tools in the same way as scientists do and reason using the data they collected. For example, middle school students used the NOAA Weather and Climate Toolkit, the same computer application scientists employ to track changes in cloud heights (Goldberg et al., 2015). Similarly, authentic contexts were also provided for data literacy interventions as students collected data by visiting local streams (Harris et al., 2012), zoos (Mokros & Wright, 2009), and deserts (Jones, 2016).

It was also quite evident from our literature analysis that both data literacy and data science aim to promote interdisciplinary collaboration. For example, Vahey et al. (2012) engaged seventh-grade students in a Thinking with Data project that used a cross-disciplinary approach and integrated modules for social studies, mathematics, science, and English language arts classes. Students were asked to investigate and present coherent arguments using data related to equitable water distribution in the Tigris/Euphrates watershed, thus fostering a deep understanding of data across the curriculum. Similarly, the Data Science Outreach Educational Program (aimed to promote data science education amongst STEM students) also emphasized the interdisciplinary collaboration of the dynamic fields of data science and other STEM (Science, Technology, Engineering, and Mathematics) concentrations to pique students' interest (Sami et al., 2020).

Quantitative Approach Towards Data

A review of existing literature shows that data literacy and data science activities in science classrooms have focused on infusing quantitative thinking through specific data-related competencies such as data collection, analysis, and communication. Details are provided in Table 1. This limited and purely quantitative approach toward data science and data literacy can pose a societal challenge because it fails to raise questions about the value-laden nature of data collection and analysis. For example, focusing solely on quantitative data does not sufficiently empower students to recognize and question instances where data is unjustly used against communities or individuals, disregarding their experiences and positionalities (Van Wart et al., 2020).

Table 1

Competencies Developed by Data Science and Data Literacy in Science Classrooms

Competency	Data Literacy	Data Science Education
Collect or acquire data	Harris et al. (2012) Mokros & Wright (2009) Forster et al. (2018)	Kahn (2020)
Analyse and create explanations from data or data based argumentation	Vahey et al. (2012) Jones (2012) Mokros & Wright (2009)	Hug & McNeill (2008) Ching & Hagood (2019) Sami et al. (2020) Frischemeier et al. (2021)
Evaluate the validity of explanations based on data and formulate new questions	Vahey et al. (2012)	Goldberg et al. (2015)
Create representation or visualization of data	Gebre & Polman (2020) Robertson & Tisdall (2020) Maltese et al. (2015)	Kahn (2020)

Differences in Conceptualizations

In light of the discussion above, we can infer that data literacy is closely related to data science. However, there are some differences, such as the difference in the level of competence. While data science education prepares learners for careers as trained data specialists, data literacy education aims to prepare learners for all roles of an increasingly datafied society (Pangrazio & Sefton-Green, 2022).

The nature of data sets differs in data science and data literacy activities. In some publications about data science education, modeling with open large-scale data sets (LSDS) was emphasized (e.g., Frishemeier et al., 2021; Kahn, 2020; Maltese, 2015; Philip et al., 2016; Ucar

& Trundle, 2011). In contrast, interventions used to teach data literacy in schools typically used small data sets (e.g., Vahey et al., 2012; Gebre & Polman, 2020).

Moreover, a core aspect of data science also involves working with data sets in which the student is not involved in the process of data collection (Wise, 2020). Web-based, real-time data could, therefore, offer a promising tool for conducting data science scientific inquiries within classroom environments (e.g., Ucar & Trundle, 2011). In contrast, data literacy activities used both first and second-hand data. For example, students measured the leg length of preserved grasshoppers to develop an understanding of variation in a population. Then, they compared this with the second-hand data available on the website. Thus, our findings align with Gould (2017), who viewed data literacy as a modern extension of statistical literacy characterized by smaller datasets.

Use of Simple and Advanced Software

Learning to reason with visual data is a central part of learning science. Different software for data visualization have been used for data literacy and data science education interventions. Data literacy-focussed interventions used software such as Google Sheets and Paint (Rosenberg et al., 2022), while advanced educational software such as TinkerPlots, Common Online Data Analysis Platform (CODAP), and NOAA Weather and Climate Toolkits were used for data science interventions (Frischemeier et al., 2021; Goldberg et al., 2015). Data science educational software such as InspireData, was used in an intervention that helped chemistry students analyze data related to temperature and humidity measurements within their school classrooms (Gunter, 2007). With integrated table and plot views created by software, students moved instantly from data collection to data analysis, learned to dig deep into the information, saw the different meanings of data and strengthened their analytical skills.

Discussion

Although data literacy and data science are two different domains, they both share considerable similarities in their aims, choice of contexts, cross-curricular approaches, and nature of data sets in science classrooms. Both adopt a situated learning approach as data-related activities are grounded in authentic contexts. Our analysis also shows that both data literacy and data science place a strong instructional emphasis on developing specific competencies in quantitative reasoning. This is evident in the tools used to measure data literacy skills that gauged students' ability to interpret complex data tables (Vahey et al., 2012), read graphs, and analyze box and whisker plots (Jones, 2016). Therefore, our findings support the idea that data literacy lies at the intersection of quantitative reasoning and data science (Kjelvik & Schultheis, 2015).

The study of data literacy and data science practices in science classrooms also indicates that the marked boundaries between the two domains appear to diminish when the learning objectives involve teaching with or about data. As the curricular expectations for K–12 students are already densely packed, developing educational materials that focus separately on data literacy and data science can be challenging. To address this, we need to develop curricula and practices (e.g., thematic teaching) that eliminate rigid boundaries between subjects; emphasize the connections among science, mathematics, and technology; and present scientific endeavors as a social enterprise. Additionally, we propose broadening the scope of data-related concepts in the science curriculum to include both quantitative and qualitative aspects of data in order to

emphasize the everyday relevance of data for science students. This is particularly important in today's "post-truth world" (Chinn et al., 2021), where learners need to engage with broader issues about how data is used in society.

References

Bhargava, R., Deahl, E., Letouzé, E., Noonan, A., Sangokoya, D., & Shoup, N. (2015). *Beyond data literacy: Reinventing community engagement and empowerment in the age of data.* [White paper]. Data-Pop Alliance.

Bowler, L., Acker, A., Jeng, W., & Chi, Y. (2017). *"It lives all around us": Aspects of data literacy in teen's lives. Proceedings of the Association for Information Science and Technology, 54*, 27–35. https://doi.org/10.1002/ pra2.2017.14505401004

Campbell, T., Wang, S. K., Hsu, H. -Y., Duffy, A. M., & Wolf, P. G. (2010). Learning with web tools, simulations, and other technologies in science classrooms. *Journal of Science Education and Technology, 19*(5), 505–511.

Ching, C. C., & Hagood, D. (2019). Activity monitor gaming and the next generation science standards: Students engaging with data, measurement limitations, and personal relevance. *Journal of Science Education and Technology, 28*, 589–601. https://doi.org/10.1007/s10956-01909789-5

Chinn, C. A., Barzilai, S., & Duncan, R. G. (2021). Education for a "post-truth" world: New directions for research and practice. *Educational Researcher, 50*(1), 51– 60.

D'Ignazio, C., & Bhargava, R. (2020). Data visualization literacy: A feminist starting point. In M. Engebretsen & H. Kennedy (Eds.), *Data visualization in society* (pp. 207–222). Amsterdam University Press.

Data-Pop Alliance. (2015). *Beyond data literacy.* http://datapopalliance.org/wp content/uploads/2015/11/Beyond-Data-Literacy 2015.pdf

Doering, A., & Veletsianos, G. (2007). An investigation of the use of real-time, authentic geospatial data in the K–12 classroom. *Journal of Geography, 106*, 217–225.

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V. R., Tsou, A., Weingart, S., & Sugimoto, C. R. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology, 66*(8), 1523–1545. https://doi.org/10.1002/asi.23294

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

Forster, M., Bestelmeyer, S., Baezrodriguez, N., Berkowitz, A., Caplan, B., Esposito, R., Grace, E., & Mcgee, S. (2018). *Data jams. The Science Teacher, 86*(2), 48–53.

Frischemeier, D., Biehler, R., Podworny, S., & Budde, L. (2021). A first introduction to data science education in secondary schools: Teaching and learning about data exploration with CODAP using survey data. *Teaching Statistics, 43*(1), 182–189. https://doi.org/10.1111/test.12283

Garcia, A. G., McGrew, S., Mirra, N., Tynes, B., & Kahne, J. (2021). Rethinking digital citizenship: Learning about media, literacy, and race in turbulent times. In C. D. Lee, G. White, & D. Dong (Eds.), *Educating for civic reasoning and discourse* (pp. 319–352). National Academy of Education.

Gebre, E. (2018). Young adults' understanding and use of data: Insights for fostering secondary school students' data literacy. *Canadian Journal of Science, Math and Technology Education, 18*, 330–341.

Gebre, E. (2022). Conceptions and perspectives of data literacy in secondary education. *British Journal of Educational Technology, 53,* 1080–1095. https://doi.org/10.1111/bjet.1324

Gebre, E. & 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,* doi.org/10.1016/j.lcsi.2019.100374

Goldberg, R., Magee, N., & Madden, L. (2015). Using authentic data to enhance middle school atmospheric science instruction. *Science Scope, 38*(8), 25–32.

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal, 16*(1), 22–25.

Gunter, G. A. (2007). Building student data literacy: An essential critical-thinking skill for the 21st century. *MultiMedia & Internet@Schools, 14*(3), 24–28.

Harris, C., Berkowitz, R, A., Alvarado, A. (2012). Salt pollution as a case study for teaching data literacy. *The American Biology Teacher, 74*(7), 479–484.

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

Jones, L. P. (2016). *Teaching data literacy in high school biology* [Unpublished master's thesis]. Montana State University.

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, DOI: 10.1080/10508406.2019.1693377

Kjelvik, M. K., & Schultheis, E. H. (2015). Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. *CBE Life Sciences Education, 18*(2), es2. https://doi.org/10.1187/cbe.18-02-0023

LaMar, T., & Boaler, J. (2021). The importance and emergence of K–12 data science. *Phi Delta Kappan, 103*(1), 49–53. doi:10.1177/00317217211043627

Lee, O., & Campbell, T. (2020). What science and STEM teachers can learn from COVID-19: Harnessing data science and computer science through the convergence of multiple STEM subjects. *Journal of Science Teacher Education, 31*(8), 932–944.

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. https://doi.org/10.1080/10 508406.2019.1666011

Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A call for a humanistic stance toward K–12 data science education. *Educational Researcher, 50*(9), 664–672.

Lupton, D. (2017). Feeling your data: Touch and making sense of personal digital data. *New Media & Society, 19*, 1599–1614.

Lupton, D. (2018). How do data come to matter? Living and becoming with personal data. *Big Data & Society, 5*(2). https://doi.org/10.1177/2053951718786314

Maltese, A. V., Harsh, J. A., & Svetina, D. (2015). Data visualization literacy: Investigating data interpretation along the novice—expert continuum. *Journal of College Science Teaching, 45*(1), 84–90. http://www.jstor.org/stable/43631889

Markham, A. N. (2019). Critical pedagogy as a response for datafication. *Qualitative Inquiry, 25*(8), 754–760.

Markham, A. N. (2020). Taking data literacy to the streets: Critical pedagogy in the public sphere. *Qualitative Inquiry, 26*(2), 227–237.

Mokros, J., & Wright, T. (2009). Zoos, aquariums, and expanding students' data literacy. *Teaching Children Mathematics, 15*(9), 524–530. https://doi.org/10.5951/TCM.15.9.0524

Pangrazio, L. & Sefton- Green, J., (2022). The death of the educative subject? The limits of criticality under datafication. *Educational Philosophy & Theory*. Epub ahead of print. https://doi.org/10.1080/00131857.2021.1978072

Pedaste, M., Maeots, M., Siiman, L. A., Jong, T. D., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions of the inquiry cycle. *Educational Research Review, 14*, 47–61.

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.

Prado, J. C., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri, 63*(2), 123–134.

Robertson, J., & Tisdall, K. E. M. (2020). The importance of consulting children and young people about data literacy. *Journal of Media Literacy Education, 12*(3), 58–74.

Rosenberg, J. M., Schultheis, E. H., Kjelvik, M. K., Reedy, A., & Sultana, O. (2022). Big data, big changes? The technologies and sources of data used in science classrooms. *British Journal of Educational Technology, 53*(5), 1179–1201. https://doi.org/10.1111/bjet.13245

Sami, J., Sinclair, K., & Medsker, L. (2020). Data science outreach educational program for high school students focused in agriculture. *Journal of STEM Education, 21*(1), 18–33.

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

Ucar, S. & Trundle, K.C. (2011). Conducting guided inquiry in science classes using authentic, archived, web-based data. *Computers & Education, 57*, 1571–1582.

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.

Vahey, P., Rafanan, K., Patton, C., Swan, K., Van't Hooft, M., Kratcoski, A., & Stanford, T. (2012). A cross-disciplinary approach to teaching data literacy and proportionality. *Educational Studies in Mathematics, 81*(2), 179–205. http://www.jstor.org/stable/23254237.

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.

Wise, F.A. (2020). Educating data scientists and data literate citizens for a new generation of Data. *Journal of the learning Sciences, 29*(1), 165–181, DOI: 10.1080/10508406.2019.1705678

Wolff, A., Gooch, D., Cavero Montaner, J. J, Rashid, U., & Kortuem, G., (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics, 12*(3), 9–26.