Statistics and data science education as a vehicle for empowering citizens – short summary of a survey

Rolf Biehler, Takashi Kawakami, Erna Lampen, Travis Weiland and Lucía Zapata-Cardona

This article is a short summary of the report of survey team 3, presented to the 15th International Congress on Mathematical Education (ICME-15) in Sydney in July 2024.

1 Introduction: Data in society, data science education and citizen empowerment

We face a new challenge in determining which knowledge, skills, and dispositions about data and statistics should and could be taught in secondary education. Data science, as a new field embracing statistics, touches several disciplines, and its scholarly knowledge is interdisciplinary, dynamic, and unstable. This is already a challenge for the process of "didactic transposition" [10] of scholarly knowledge to knowledge to be taught. Moreover, the transposition has to take into account the role of data in society, which affects all communities and individuals. Therefore, the use of and discourse about data in society are more and more pertinent for those who reflect on the relevance and uses of knowledge to be taught in secondary and primary schools. School statistics has not kept pace with how citizens engage with increasingly pervasive data, such as navigating X feeds, using artificial intelligence (AI) to identify photos, and streaming GPS data as a live feed into Google Maps to estimate travel times. Data science and data-driven AI have led to breakthroughs in science and society. Data science can be used for social good, including work to protect the environment and address climate change. But it can also promote the economic and political interests of a few while failing to serve the interests of most citizens. Its use has raised massive concerns about privacy, misuse of data, ethics, and surveillance of citizens, to name a few. Awareness of the non-objective nature of data, such as the underlying gender/racial bias in how and whose data are used to train algorithms, is also at the forefront of these discourses – see, for example, the journals *Big Data & Society* and *AI & Society*.

These developments create the need to redefine what it can and should mean to empower citizens through education. The survey identified various conceptions of citizen education, such as aiming at personally responsible, participatory, justice-oriented citizens, or global citizenship and promoting the UN's sustainable development goals [7, 29, 33, 34]. In the case of expanding data science to empower citizens, these aims compete with forces on the educational system, such as economic, military and political forces as part of international competitions that require workforce education for different purposes.

The survey identified literacy conceptions for citizen education from different disciplines and perspectives, which enhance and transform more traditional conceptions of statistical literacy [15, 17]. These include (critical) statistical or data literacy [24], civic statistical literacy [29], and data-driven mathematical, computational, and algorithmic modeling [20]. Data literacy is also conceived as part of more general literacies, such as digital humanities literacy, media, news, information literacy, digital literacy, Artificial Intelligence (AI), and machine learning literacy [9]. The datafication of disciplines has led to subject-specific data practices in the social and natural sciences. New interdisciplinary approaches have emerged, particularly in addressing socio-scientific problems within the framework of citizen science. Moreover, conceptions such as critical datafication literacy [31], personal data literacy [27], data awareness [18], data acumen, conscience, ethics, activism, and feminism [11] are put forward in the discourse.

These complex and demanding developments impact all school subjects and collide with fully packed curricula in all subjects and with an abundance of problems in many educational systems, for instance, a very high number of low achievers.

The survey identified recent books and special issues of journals that help to orient the discourse in the field. Special issues on data science education have been published since 2020: in the *Journal of the Learning Sciences* (2020); *Teaching Statistics* (2021); *Statistics Education Research Journal* (2022); *Educational Technology and Society* (2022); *Information and Learning Sciences* (2024); *Computers and Education Open* (2024). This shows that the survey went far beyond the usual sources for mathematics and statistics education.

Given this complexity, the survey team focused on four topics.

2 Topic 1: Civic statistics and humanistic perspectives on data literacy education in the U.S. and Europe

Humanistic perspectives and citizenship development have been prevalent themes in statistics and data science educational research in the U.S. and Europe for several years (e.g., [23, 29]). Scholars in these areas also work transdisciplinary, infusing data literacy across the curriculum. Much of this work has been done in small-scale qualitative projects as these fields try to make sense of the rapidly evolving nature of data literacy in our information-centric societies. Several themes have emerged, including (1) reading the world with data, making sense of the data-based communication of others, including data viz and data journalism [19, 30]; (2) writing the world with data, using data practices to investigate the world around us with students authentically engaging in activities of the discipline through data investigations or creating data stories (e.g., [22, 35]; (3) data structures and handling, focusing on data moves and clean data to make raw data accessible to analysis particularly in the form of "tidy data" (e.g., [13]); and (4) technology, including the development, interaction with, and learning from technology.

3 Topic 2: Critical perspectives on data literacy emerging from Latin America

Data literacy in Latin American countries has taken a critical perspective in the form of critical data literacy, which is the skill set that enables people to use and produce data critically concerning the reality behind the data [4]. It requires a combination of technical skills and the ability to reason critically about data and context. Critical data literacy in Latin America is strongly influenced by the critical pedagogy and popular education of Paulo Freire [32], which seeks to help people to develop their ability to read and write their world. The work of Giroux and Skovsmose has also been influential in this perspective. Critical data literacy in a region with high levels of social, cultural, and economic inequality is needed to help people to (1) make sense of the important data that affects their lives, (2) make informed decisions, (3) participate in public life, (4) recognize the harm that powerful interests can inflict with data, (5) recognize that data is not neutral, (6) recognize that biased algorithms can exacerbate social and economic inequalities, (7) expose systematic social injustices, (8) understand the inequalities within the work system and how disparities between nations exacerbate the oppression of deprived populations [28].

4 Topic 3: Joint discourse between mathematical modeling and statistics/data science communities

More researchers have recently worked "at the boundary" of mathematical modeling (MM) and statistics/data science communities (e.g., [3, 21]). This section aimed to report new trends in the joint discourse between the two scientific communities, focusing on three relevant discourses on data-rich MM since 2020. The first discourse proposes a data-rich MM process with statistics and mathematics at its core to develop statistical and/or mathematical literacy and/or disciplinary learning (e.g., [12]). The second discourse discusses interdisciplinary data-rich MM, involving not only statistics and mathematics but also other disciplines/subjects, to promote STEM literacy as essential for citizens (e.g., [2, 25]). The third discourse focuses on societal data-rich MM, which uses global, social, political, ethical, and everyday contexts to promote critical thinking and citizenship (e.g., [16, 36]). All three discourses emphasize the process of modeling as a cycle and seek to understand the relationship between discipline-specific approaches to modeling and the role of data herein.

5 Topic 4: What can mathematics/statistics education contribute to Artificial Intelligence/Machine learning literacy

The discussion on AI literacy for secondary students has rapidly expanded in a few years, with several review articles emerging in this evolving field, primarily from computer science education (e.g., [1]). Machine learning (ML), which intersects with statistics and mathematics, is recognized as a key area of focus. It involves predicting outcomes through mathematical and statistical modeling and a broader form of inference than sample-to-population inference. A major concern is the opacity of ML models [8]. A consensus is forming that at least one type of machine learning should be taught in schools using a "white box" approach, i.e., a machine learning algorithm not as a black box, but making it partially transparent to learners ("gray box" or "white box") [26]. Researchers propose decision trees [14] and k-nearest neighbors [26] as promising candidates to reduce the opacity of machine learning algorithms. The reviewed discourse on fundamental concepts of ML includes distinguishing between regression and classification, understanding misclassification types, and differentiating training and test data to address overfitting, bias, and fairness (see [6] for an overview with examples for teachers). Research projects were identified that make these concepts accessible to secondary students through tools like data cards [14], CODAP, and Jupyter notebooks (e.g., [5]).

6 Conclusion

This survey clearly indicates that statistics has evolved into data science with new tools, discourses, and various source domains. The data demands that pervade most societies require renewed attention to statistics education at school, data science education across

disciplines, and research-informed decisions about curriculum and pedagogy.

An extended version of this report is available. To order, please email biehler@math.upb.de.

References

[1] O. Almatrafi, A. Johri and H. Lee, A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Comput. Educ. Open* **6**, article no. 100173 (2024)

[2] K. Aridor, M. Dvir, D. Tsybulsky and D. Ben-Zvi, Living the DReaM: The interrelations between statistical, scientific and nature of science uncertainty articulations through citizen science. *Instr. Sci.* **51**, 729–762 (2023)

[3] J. B. Ärlebäck and T. Kawakami, The relationships between statistics, statistical modelling and mathematical modelling. In *Advancing and consolidating mathematical modelling: Research from ICME-14*, pp. 293–309, Springer, Cham (2023)

[4] S. Baack, Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data Soc.* **2(2)** (2015)

[5] R. Biehler and Y. Fleischer, Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teach. Stat.* **43**, S133–S142 (2021)

[6] R. Biehler and S. Schönbrodt (eds.), KI verstehen: Wie Maschinen lernen. *mathematik lehren* **244** (2024)

[7] Brookings Institution, UNESCO and GEFI-YAG (eds.), *Measuring Global Citizenship Education - A Collection of Practices and Tools*. The Brookings Institution, Washington D.C. (2017)

[8] J. Burrell, How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data Soc.* **3(1)** (2016)

[9] L. Casal-Otero, A. Catala, C. Fernández-Morante, M. Taboada, B. Cebreiro and S. Barro, AI literacy in K-12: a systematic literature review. *Int. J. STEM Educ.* **10**, article no. 29 (2023)

[10] V. Chevallard, *La transposition didactique: du savoir savant au savoir enseigné*. La Penseé Sauvage, Grenoble, France (1985)

[11] C. D'Ignazio and L. F. Klein, *Data Feminism*. The MIT Press, Cambridge, MA (2020)

[12] L. D. English, Mathematical and interdisciplinary modeling in optimizing young children's learning. In *Exploring mathematical modeling with young learners*, pp. 3–24, Springer, Cham (2021)

[13] T. Erickson, M. Wilkerson, W. Finzer and F. Reichsman, Data moves. *Technology Innovations in Statistics Education* **12(1)** (2019)

[14] Y. Fleischer, S. Podworny and R. Biehler, Teaching and learning to construct data-based decision trees using data cards as the first introduction to machine learning in middle school. *Statistics Education Research Journal* **23(1)**, article no. 3 (2024)

[15] I. Gal, Adults' statistical literacy: meanings, components, responsibilities. *Int. Stat. Rev.* **70**, 1–25 (2002)

[16] I. Gal and V. Geiger, Welcome to the era of vague news: a study of the demands of statistical and mathematical products in the COVID-19 pandemic media. *Educ. Stud. Math.* **111**, 5–28 (2022)

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

[18] L. Höper and C. Schulte, The data awareness framework as part of data literacies in K-12 education. *Inf. Learn. Sci.* **125**, 491–512 (2024)

[19] J. B. Kahn, L. M. Peralta, L. H. Rubel, V. Y. Lim, S. Jiang and B. Herbel-Eisenmann, Notice, wonder, feel, act, and reimagine as a path toward social justice in data science education. *Educ. Technol. Soc.* **25**, 80–92 (2022)

[20] T. Kawakami and A. Saeki, Roles of mathematical and statistical models in data-driven modelling: A prescriptive modelling perspective. In *Researching Mathematical Modelling Education in Disruptive Times*, pp. 595–605, Springer, Cham (2024)

[21] S. Kazak, T. Fujita and M. P. Turmo, Students' informal statistical inferences through data modeling with a large multivariate dataset. *Math. Think. Learn.* **25**, 23–43 (2023)

[22] H. Lee, G. Mojica, E. Thrasher and P. Baumgartner, Investigating data like a data scientist: Key practices and processes. *Statistics Education Research Journal* **21(2)**, article no. 3 (2022)

[23] V. R. Lee, M. H. Wilkerson and K. Lanouette, A call for a humanistic stance toward K–12 data science education. *Educ. Researcher* **50**, 664–672 (2021)

[24] J. Louie, Critical data literacy: Creating a more just world with data. Technical report. National Academy of Sciences' Workshop on Foundations of Data Science for Students in Grades K–12, Washington D.C. (2022)

[25] K. Makar, K. Fry and L. English, Primary students' learning about citizenship through data science. *ZDM* **55**, 967–979 (2023)

[26] K. Mike and O. Hazzan, Machine learning for non-majors: A white box approach. *Statistics Education Research Journal* **21(2)**, article no. 10 (2022)

[27] L. Pangrazio and N. Selwyn, 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media Soc.* **21**, 419–437 (2018)

[28] J. E. Raffaghelli, Alfabetización en datos y justicia social ¿Un oxímoron? Respuestas desde la contra-hegemonía. *Revista Izquierdas* **51**, 1–18 (2022)

[29] J. Ridgway (ed.), *Statistics for empowerment and social engagement: Teaching civic statistics to develop informed citizens*. Springer, Cham (2022)

[30] L. H. Rubel, C. Nicol and A. Chronaki, A critical mathematics perspective on reading data visualizations: Reimagining through reformatting, reframing, and renarrating. *Educ. Stud. Math.* **108**, 249–268 (2021)

[31] I. Sander, *Critical datafication literacy: A framework and practical approaches*. transcript, Bielefeld, Germany (2024)

[32] A. F. Tygel and R. Kirsch, Contributions of Paulo Freire to a critical data literacy: A popular education approach. *The Journal of Community Informatics* **12**, 108–121 (2016)

[33] T. Weiland, Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educ. Stud. Math.* **96**, 33–47 (2017)

[34] J. Westheimer and J. Kahne, What kind of citizen? The politics of educating for democracy. *Am. Educ. Res. J.* **41**, 237–269 (2004)

[35] M. H. Wilkerson and V. Laina, Middle school students' reasoning about data and context through storytelling with repurposed local data. *ZDM* **50**, 1223–1235 (2018)

[36] M. H. Wilkerson, K. Lanouette and R. L. Shareff, Exploring variability during data preparation: A way to connect data, chance, and context when working with complex public datasets. *Math. Think. Learn.* **24**, 312–330 (2022)

Rolf Biehler is a full professor emeritus for didactics of mathematics at Paderborn University, Germany. His research encompasses probability, statistics, and data science education, with a focus on integrating digital tools into learning processes. He co-directs the "Data Science and Big Data at School" project,[1] developing curricula and professional development courses for teachers to introduce data exploration and data-based machine learning at various educational levels.

biehler@math.upb.de

Takashi Kawakami is an associate professor of mathematics education at the Cooperative Faculty of Education, Utsunomiya University, Japan. His research interests include mathematical modeling, statistics and data science education, STE(A)M education, and mathematics teachers' professional learning. His current primary focus is exploring the intersection of mathematical modeling and statistics/data science education using conceptual and empirical approaches.

t-kawakami@cc.utsunomiya-u.ac.jp

Erna Lampen is a retired senior lecturer from Stellenbosch University, South Africa. Her research interests include mathematical and statistical reasoning and implications of technological advancement on such reasoning. She focuses on teacher education and materials development to integrate STEAM subjects.

ernalampen@sun.ac.za

Travis Weiland is an assistant professor at the University of North Carolina, Charlotte. His interests are at the intersection of teacher education, statistics education, data science education, and critical education. Currently, much of his work is focused on studying how mathematics teachers develop critical statistical literacies for doing and teaching data investigation concepts and practices.

tweilan1@charlotte.edu

Lucía Zapata-Cardona is a full professor at the Universidad de Antioquia, Colombia. Her interests are in teacher education, statistics education and critical data science. She is currently working on the development of data science teaching materials that help citizens understand and transform critical social issues.

lucia.zapata1@udea.edu.co

---

[1] www.prodabi.de/en