Toward a Theoretical Framework for Data Fluency Teaching and Learning in Middle School STEM

Nicole Wong, Rasha Elsayed, Leticia Perez, Kirsten Daehler, Pai-rou Chen

WestEd

Contact: nicolewong@wested.org

This paper presents findings from a qualitative study of eleven experienced STEM educators who worked alongside developers to design and implement data-rich lessons in their grades 6–9 mathematics and science classrooms. In the context of a project that seeks to develop professional learning for data fluency, researchers documented the co-development process to articulate a model of what teachers need to know and be able to do in order to support their students' data fluency. The project team distilled key findings into two framing documents: 1) a description of *high-leverage areas of focus for PL* which highlight challenges faced by teachers, which are common, important for data fluency, and represent opportunities for supporting teacher and student growth; and 2) a *logic model* that describes how the PL course under development is expected to influence teacher, classroom, and student outcomes. This paper contributes to the larger education community by defining the professional learning needs of educators who wish to integrate data into their STEM classrooms. These frameworks provide *designers* and *researchers* with touchpoints to structure and study PL experiences, lesson materials, and other classroom resources for both new and veteran educators. These tools can provide *STEM teachers* with guidance for reflecting on their current knowledge, skills, beliefs, and teaching practices that help their students become more data fluent.

Suggested citation: Wong N., Elsayed, R., Perez, L. R., Daehler, K. R., & Chen, P. (2024, March 17–21). *Toward a theoretical framework for data fluency teaching and learning in middle school STEM* [Poster Session]. Ninety-seventh annual international conference of the National Association for Research in Science Teaching, Denver, CO, United States.

Subject or Problem

Data play a prominent role in everyday life and is vital for modern careers.
The need to support students' data fluency – the ability and confidence to actively make sense of and use data – has been encoded through data-related standards in the *Next Generation Science Standards* (NGSS)(NGSS Lead States, 2013) and The Common Core State Standards for Mathematics (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Despite the need for data fluency, many teachers are woefully underprepared to implement these standards (Banilower et al., 2018; Lovett & Lee, 2018) and they need classroom-ready tools and resources and teacher preparation (National Academies of Sciences, Engineering, and Medicine, 2023). Without professional learning (PL), teachers tend to use their existing personal instructional approaches that may or may not align with research (Pratt, 2002) and rarely are effective in preparing 21st century citizens (Gulamhussein, 2013). If *all* students are to become fluent with data, teachers must be prepared to implement pedagogical approaches and cutting-edge materials that support student learning, including the ability to facilitate high-quality conversations about data (Vahey et al., 2017).

Much is still unknown about how students and teachers progress to data fluency and how teachers can support students on their path to data fluency. To develop PL for teachers who wish to support their students' data fluency, we must first learn more about how teachers and students progress toward data fluency, and we need to articulate a model of the *content and pedagogical content knowledge* required for supporting students' data fluency. While some of this prerequisite knowledge is known (Arnold et al., 2018; Ben-Zvi et al., 2018; Lee et al., 2020), the field is still nascent (da Ponte & Noll, 2018; Petocz et al., 2018; Zieffler et al., 2018). Existing research offers some insight into successful pedagogical practices that support student reasoning about data, such as providing coherent opportunities for students to participate in the construction, representation, analysis, and use of data as evidence, rather than as separate experiences (Lee & Wilkerson, 2018; Hardy et al., 2020). In the introduction to a special issue on situated data in *Journal of the Learning Sciences*, Rubin (2020) summarizes five aspects of working with data that need to be better understood by educators and better integrated into data science teaching and learning: data are contextualized; data exhibit variability; students struggle with aggregated data; visualization is key to reasoning with data; and drawing inferences from data is fraught. Another challenge associated with developing data fluency is that both students and adults exhibit difficulty reasoning about multivariate phenomena, yet most K–12 standards and curricula focus on basic data visualization and univariate and bivariate distributions (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010; Sorto, 2011). We aim to identify topics and practices that are particularly challenging and important to data fluency, and integrate those ideas into a framework for supporting the development of PL and resources to help middle school STEM teachers engage their students in data-rich learning.

Design or Procedure

This paper presents the results of a qualitative study of eleven experienced STEM educators who worked alongside project-based PL developers to design and implement data-rich lessons in their grades 6–9 mathematics and science classrooms.

This research took place in the context of a larger research and PL development project known as *Data Fluency*[1], in which math and science teachers were recruited to help project-based developers to generate a set of classroom narratives that describe ways in which teachers integrate data into their instruction. These narratives are then used in a PL course for *other* educators. Researchers documented the work with the co-development teachers in order to inform the development of the PL course and to develop a model of teacher knowledge needed to support students' data fluency.

Figure 1 provides an overview of the theoretical framework development activities within the broader *Data Fluency* project. The white boxes in Figure 1 show the areas of focus for this paper. First, the team conducted a review to help articulate the knowledge, skills, practices, dispositions, challenges, and existing professional learning supports and tools related to teachers' and students' data fluency, including terms such as data literacy, data science, statistical reasoning, and computational thinking. At the same time, we recruited, engaged, and studied the work of a cohort of co-development teachers who helped our team design professional learning. Through a review of the literature and our analysis of the research with our co-developers, we generated two documents for guiding the development of our professional learning course that are the primary focus of this paper:

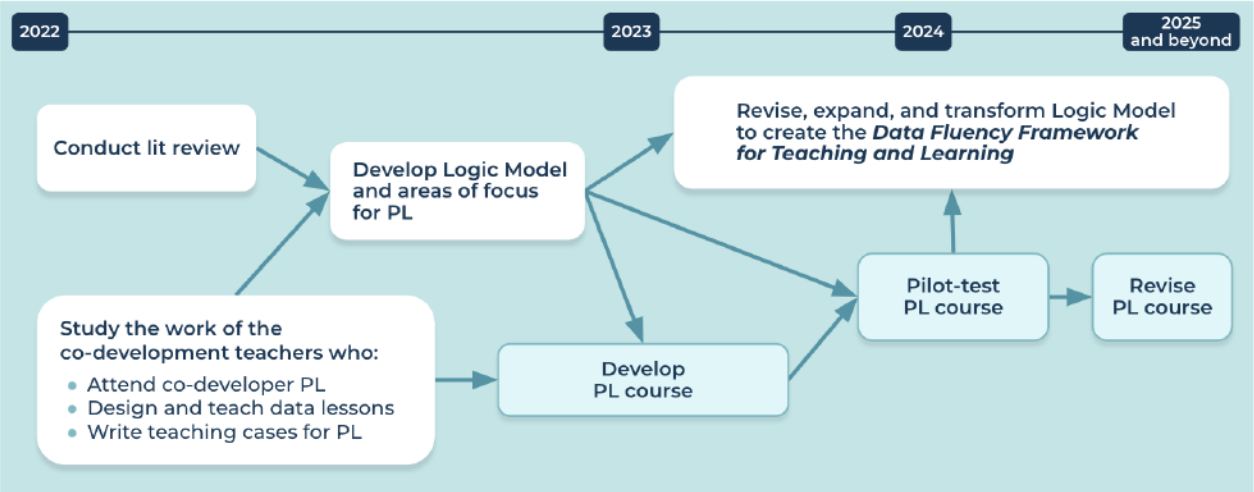
1. **A logic model** that describes how the PL course under development could potentially influence teacher, classroom, and student outcomes, and
2. **A description of high-leverage areas of focus for data fluency PL** that highlights challenges faced by teachers, which are common, important for data fluency, and represent opportunities for supporting teacher and student growth.

This work is still ongoing, and we are currently expanding the logic model beyond the scope of this project's PL course into a broader *Data Fluency Framework for Teaching and Learning*, which

[1] The full project name is *Boosting Data Science Teaching and Learning in STEM*. This material is based upon work supported by the National Science Foundation under Grant No. 2101049. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

describes the knowledge and skills teachers need to support their students' data fluency. This framework will provide designers and researchers with touchpoints to structure and study PL experiences, lesson materials, and other classroom resources. These tools can provide STEM teachers with guidance for reflecting on their own knowledge and practices.

Figure 1. *Overview of Theoretical Framework Development Activities in the Data Fluency Project*



The literature review and the study of the co-development teachers were guided by the following questions: (a) What do teachers need to know and be able to do to support students in becoming data fluent? (b) What are common student misconceptions and roadblocks in students' progress to data fluency? (c) What are the core components of PL that boost teachers' data fluency and their ability to support students becoming data fluent?

The co-developers were guided and supported to design and implement data lessons by a series of workshops led by the project team to deepen their own understanding of data, their knowledge of how to use a data visualization tool called Common Online Data Analysis Platform (CODAP) (2023), and pedagogical practices related to data. The learning topics spanned a range of data-related ideas and practices including communicating with data, generating and obtaining data (including support for navigating the collection of first-hand data and finding and curating datasets for classroom use), exploring data, and humanizing data. These workshop topics were selected and crafted to meet the emerging needs of the educators. The teachers also participated in small working groups and "office hour" sessions with PL specialists that included tailored support to help plan, design, implement, and reflect on the data lessons. Each teacher selected the topic for their own data lesson, and the lessons varied in length from a few days to several weeks. During these planning sessions, PL specialists supported teachers' efforts to increase accessibility and inclusion by encouraging reflection about how their tasks and facilitation strategies could potentially meet the needs of student groups with differing levels of data and content understanding.

The research team observed and video-recorded the workshops, office hours, working group meetings, and classroom enactments. During the classroom observations, researchers took notes, collected samples of student work, video recorded students and teachers working with each other, video recorded pairs of students working with data using technological tools, and conducted a post-lesson debriefing interview with the teacher. Notes and interview transcripts were coded, summarized, and compiled into memos. Pre-structured codes focused on teacher knowledge; teacher skills and practices; requests for assistance; PL supports; technological supports; and student knowledge, skills, and misconceptions related to data fluency. Open coding captured other salient aspects of the work.

Analysis and Findings

High-Leverage Areas of Focus for Data Fluency PL

Below, we describe 7 high-leverage areas of focus for data fluency PL. This list captures our current thinking about challenges faced by educators seeking to incorporate data into their classrooms. We selected areas of difficulty which are common, important for data fluency, and represent opportunities for supporting teacher and student growth.

#1 Examine the nature of data. Typical applications of data in science and mathematics classrooms involve numeric data, and their existence is often treated as authoritative and unproblematic by both teachers and students. Professional learning has the opportunity to *broaden teachers'* notions about data types and data experiences by engaging educators with both traditional (e.g., data tables, recorded values, plots) and non-traditional (e.g., photos, satellite images, community science observation records) data types and representations.

#2 Amplify sociocultural and humanistic aspects of working with data. Related to the point above, PL providers have the opportunity to introduce a *humanistic approach* to data (Lee et al., 2021), which emphasizes the roles humans play in all activities connected to data. Data are not presented as objective but as products which are influenced by human values, beliefs and decisions. PL that amplifies sociocultural aspects of data may support teachers to engage their students in discussions about the origins of data and ethical practices around the use of data. For example, students and teachers might think together about the following questions: For what purposes and by whom were these data collected? What beliefs and values are encoded within these data and their original uses? In what ways do we, as users of data, have the power to reinforce or resist these values and beliefs?

#3 Consider tradeoffs of primary v. secondary data. Many teachers value providing their students with the experience of collecting their own data. Students' involvement in generating data has the potential to support their understanding of the statistical, computational, or evidentiary features of the data and can offer a sense of agency and personal connection with

the data. However, classroom datasets are often too limited to support deeper inquiry (e.g., they may not include enough relevant variables or they might not have enough data points to answer the questions at hand). In contrast to more typical bivariate datasets, multivariate datasets provide more opportunities for students to pose interesting statistical investigative questions. PL can help educators reason about the choice to have students collect data or to use data provided to them. For example, PL might invite teachers to consider different methods for obtaining data and weigh the benefits, limitations, and tradeoffs of each choice in relation to specific learning goals.

#4 Engage students in multiple entryways to the data investigation process. Educators struggle to find data activities that are both *meaningful and feasible*. Lesson plans sometimes limit work with data to only one bite-sized aspect of working with data (e.g., "asking questions"), or they ask teachers to engage their students in a full data investigation, which can take weeks or months and is often depicted as following a rigid, fixed-sequence "data cycle." By using a model such as our project's *Entryways to Data* (Figure 2), PL can help educators begin to see that the phases of the data investigation process 1) are inextricably linked (not phases done in isolation), 2) do not take place in a fixed sequence, and 3) that not all phases of an investigation process need to take place in order for a data lesson to be successful. The *Entryways to Data* can help provide educators with a holistic sense of how scientists and students could use the process to make sense of data, while also offering multiple entry points to learning. Teachers can use this model to identify learning goals and create lessons that focus on just a few interrelated entryways.

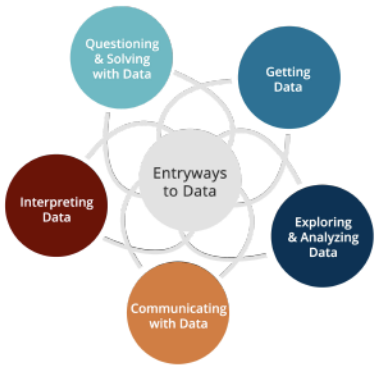


Figure 2. *Entryways to Data*

#5 Use data to support conceptual understanding within math and science content areas. When planning lessons, teachers often had difficulty seeing the relationship between data-related goals and their math and science goals. This can lead to the creation or adoption of lessons that engage students in data purely for data's sake. PL can help teachers consider the question, "How can the patterns within data introduce or reinforce mathematical or science content concepts?" It may also be helpful to acknowledge that not all lessons have to do all things: PL can help teachers prioritize and balance data fluency goals with STEM conceptual goals. Toward this end, it should also be noted that our co-development teachers needed a great deal of support to identify and curate datasets that are aligned to their mathematics or science curriculum and standards. Creating and cataloging these classroom-facing resources is a non-trivial challenge for PL providers and curriculum developers that should be addressed on a larger scale.

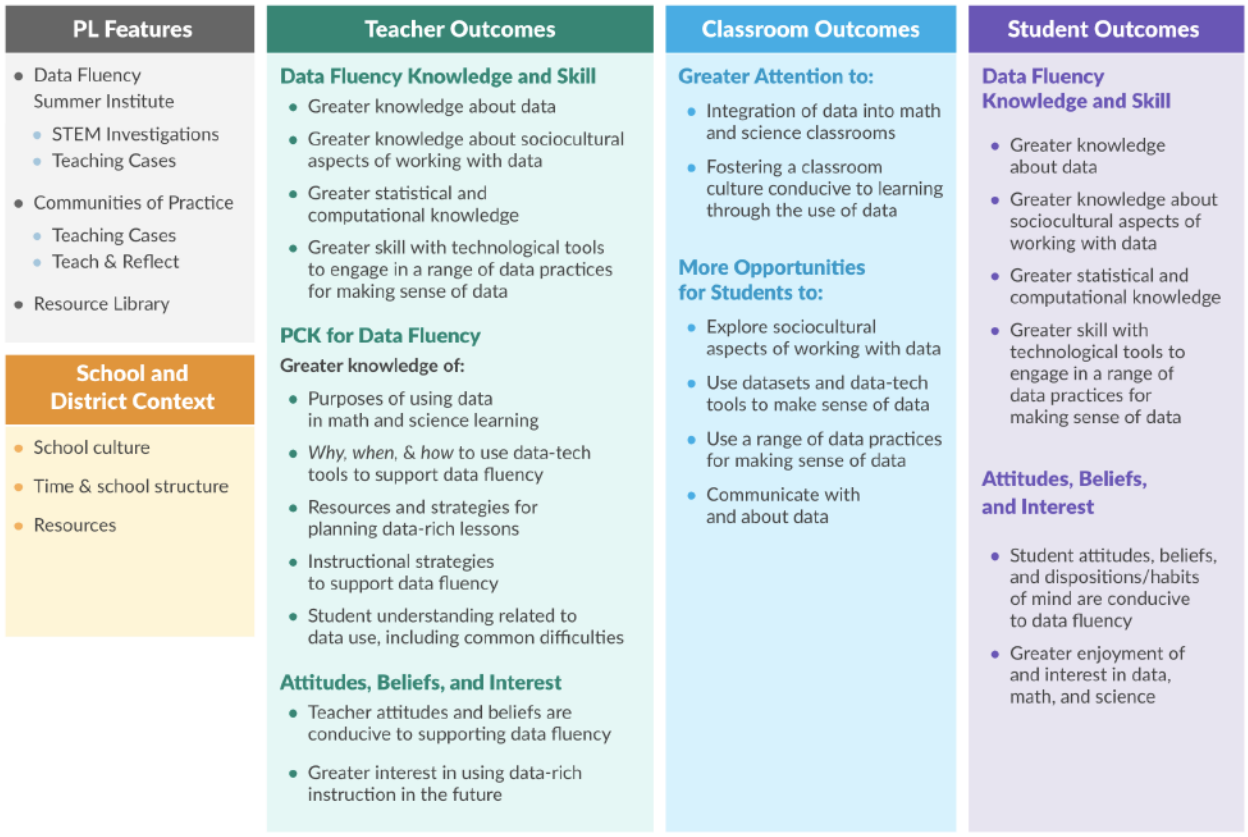
#6 Examine the role, benefits, and tradeoffs of technology. When considering how to use technology in their classrooms, teachers took great care to consider how and when to introduce technology. PL can help educators learn how to use data-tech tools *and also* consider tradeoffs of particular tools (e.g., paper and pencil vs. spreadsheet vs. data visualization software) for particular instructional purposes and learning goals. PL might illustrate the power/affordances that technological tools offer when differentiating experiences to support all learners develop their data fluency, including adaptive technologies. For example, through working with and reflecting on their experience of using the CODAP data visualization tool, our co-developers discovered that the technology helped to offload the burden of calculation by automatically creating means for students. Offloading this cognitive demand can support student engagement in other conceptual activity, such as increasing opportunities for conversation about variability rather than procedural calculation. Our co-development teachers recognized that this would allow their students who were less confident in math to participate more equitably.

#7 Foreground support for fostering data-friendly habits of mind. Individuals and classrooms demonstrating greater data fluency typically utilize habits of mind that shape how they approach data. These habits of mind/dispositions include skepticism, imagination, curiosity and awareness, openness, a propensity to seek deeper meaning, being logical, engagement, perseverance (Wild & Pfannkuch, 1999). Learners who are accustomed to more traditional classroom structures may feel discomfort in taking a skeptical stance in the face of what they may think of as authoritative data, or they may feel that imagination is not welcome in a science or math classroom. PL that calls attention to these habits can help educators foster classroom cultures where these habits are valued and affirms the presence and value each student brings to a data task. In PL, educators could, themselves, engage in practices that reinforce data-friendly habits of mind, such as using a reflective journal to document when they used certain habits throughout an investigation.

Logic Model

The logic model in Figure 3 describes how the project's Data Fluency Professional Learning is expected to influence teacher, classroom, and student outcomes. It also acknowledges the important role of the school and district context. The components of the logic model were derived from key themes identified during our work with the co-development teachers. The descriptors of teacher and student knowledge represent an explicit integration of multiple disciplines involved in data education in STEM, including domain knowledge, data knowledge, technological knowledge, statistical knowledge, and sociocultural and affective aspects of working with data. Furthermore, the "classroom outcomes" reflect our belief that data fluency exists as a quality or set of qualities that reside in individuals *and* communities: classroom communities can vary in the degree to which they exhibit features that welcome and support sensemaking with data.

Figure 3. *Data Fluency PL Logic Model*



Directions for Future Research

We are currently pilot testing PL that is designed around the features described in this model. We are expanding our project logic model to a more broadly applicable framework that further specifies the content knowledge, pedagogical content knowledge, and technological pedagogical content knowledge needed to teach lessons that support data fluency. We hope that this tool will be useful for other researchers and developers who are interested in supporting teachers' and students' data fluency.

Contributions & General Interest

This paper contributes to the members of Strand 4, the members of NARST, and the education community at large because it helps to define the professional learning needs of educators who wish to integrate data into their STEM classrooms. The Data Fluency PL areas of focus and the logic model described in this paper are intended as tools to support developers, researchers, and educators. These frameworks provide *designers* with touchpoints to structure PL experiences, lesson materials, and other classroom resources for both new and veteran

educators. These tools can provide *STEM teachers* with guidance for reflecting on and self-assessing their current knowledge, skills, beliefs, and teaching practices that help their students become more data fluent and identify PL experiences that may support their growth in these areas. *Researchers* can use these tools to guide the development measures of data fluency related to teachers, classrooms, and students. The design and study process for this framework heavily centers around our collaboration with educators who work with a wide range of students across many teaching contexts. We aim to center their voices and choices (The Centering Voices Workgroup, 2018), their success and challenges, and suggest future areas of research for designing inclusive instructional experiences.

References

Arnold, P., Confrey, J., Jones, R. S., Lee, H. S., & Pfannkuch, M. (2018). Statistics learning trajectories. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 295–326). Springer, Cham. https://doi.org/10.1007/978-3-319-66195-7_9

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 national survey of science and mathematics education NSSME+*. Chapel Hill, NC: Horizon Research, Inc.

Ben-Zvi, D., Gravemeijer, K., & Ainley, J. (2018). Design of statistics learning environments. In D. Ben- Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 473–502). Springer, Cham. https://doi.org/10.1007/978-3-319-66195-7_16

The Centering Voices Workgroup (2018). Retrieved July 29, 2023, from https://uwphi.pophealth.wisc.edu/wp-content/uploads/sites/316/2019/04/Centering-Voices-Principles_MATCH_Sept-2018.pdf

CODAP - Common Online Data Analysis Platform. (2023). Retrieved from https://codap.concord.org/

da Ponte, J. P., & Noll, J. (2018). Building capacity in statistics teacher education. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 433–455). Springer, Cham. https://doi.org/10.1007/978-3-319-66195-7_14

Gulamhussein, A. (2013). Teaching the teachers: Effective professional development in an era of high stakes accountability. *Center for Public Education*, *1*, 1–47.

Hardy, L., Dixon, C., & Hsi, S. (2020). From data collectors to data producers: Shifting students' relationship to data. *Journal of the Learning Sciences, 29*(1), 104–126. https://doi.org/10.1080/10508406.2019.1678164

Lee, H. S., Mojica, G. F., & Lovett, J. N. (2020). Examining how online professional development impacts teachers' beliefs about teaching statistics. *Online Learning, 24*(1), 5–27.

Lee, V. R., & Wilkerson, M. (2018). *Data use by middle and secondary students in the digital age: A status report and future prospects.* Commissioned Paper for the National Academies of Sciences, Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6-12. Washington, D.C.

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. https://doi.org/10.3102/0013189X211048810

Lovett, J. N., & Lee, H. S. (2018). Preservice secondary mathematics teachers' statistical knowledge: A snapshot of strengths and weaknesses. *Journal of Statistics Education, 26*(3), 214–222. https://doi.org/10.1080/10691898.2018.1496806

National Academies of Sciences, Engineering, and Medicine. (2023). *Foundations of Data Science for Students in Grades K-12: Proceedings of a Workshop.* Washington, DC: The National Academies Press. https://doi.org/10.17226/26852.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common core state standards for mathematics.* Washington D.C.: Author.

NGSS Lead States (2013). *Next Generation Science Standards: For States, By States.* Washington, DC: The National Academies Press.

Petocz, P., Reid, A. & Gal, I. (2018). Statistics education research. In Ben-Zvi, D., Makar, K. & Garfield, J. (Eds.). *International handbook of research in statistics education* (pp. 71–99). Springer, Cham. https://doi.org/10.1007/978-3-319-66195-7_3

Pratt, D. D. (2002). Good teaching: One size fits all?. *New Directions for Adult and Continuing Education*, *2002*(93), 5–16.

Rubin, A. (2020). Learning to reason with data: How did we get here and what do we know? *Journal of the Learning Sciences, 29*(1), 154–164. https://doi.org/10.1080/10508406.2019.1705665

Sorto, M. A. (2011). Data analysis and statistics in middle grades: An analysis of content standards. *School Science and Mathematics*, *111*(3), 118–125. https://doi.org/10.1111/j.1949-8594.2010.00068.x

Vahey, P., Finzer, W., Yarnall, L., & Schank, P. (2017). CIRCL primer: Data science education. In *CIRCL Primer Series*. http://circlcenter.org/data-science-education

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67(*3), 223–248.

Zieffler, A., Garfield, J., & Fry, E. (2018). What is statistics education? In D. Ben-Zvi, K. Makar, & J. Garfield, (Eds.). *International handbook of research in statistics education* (pp. 37–70). Springer International Publishing. https://doi.org/10.1007/978-3-319-66195-7_2