Designing online professional learning to support advances in teaching strategies in statistics and data science

Hollylynne S. Lee, Gemma F. Mojica, and Emily Thrasher
North Carolina State University, U.S.A.
hollylynne@ncsu.edu

This paper describes the design of an innovative online platform that has over 50 hours of learning experiences to support educators in further advancing their understandings and pedagogical skills in teaching statistics and data science to learners age 11-18+. Two frameworks are described that support effective classroom practices: a Data Investigation Process and Seven Dimensions of Teaching Statistics and Data Science. Examples of features and interface designs are included to illustrate how the frameworks work together to support teachers' learning in a cohesive way, even as they personalize their experience and engage in different learning pathways.

INTRODUCTION

Around the world, countries have been emphasizing more statistics within mathematics curriculum, and more recently are creating standards and curricula to integrate data science into K-12 education (e.g., Sokul, 2024). Many teachers lack access to in-person professional development specific to these topics, prompting educators to increasingly turn to online solutions for broader reach and flexibility. From 2015-2019, Lee and Mojica designed and implemented two 20-hour Massive Open Online Courses for Educators (MOOC-Ed) aimed at developing teachers' knowledge for teaching statistics with a focus on data investigations and inferential reasoning. The courses attracted 3,454 educators from about 100 countries (Hollebrands & Lee, 2020) and had a positive impact on educators' confidence to teach statistics and classroom practices in using real and bigger data (Lee et al., 2020).

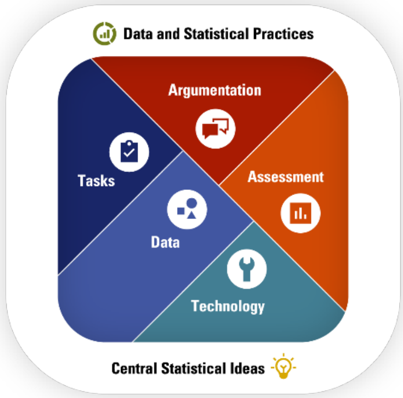
Building from the successes and research with the two MOOC-Eds, we innovated the way educators could engage in professional learning that could be more personalized to their learning goals and occur in smaller modular units. Through funding by the National Science Foundation in the U.S., and a partnership with online learning design experts at RTI International, we designed and implemented the InSTEP platform [instepwithdata.org]. Since 2022, over 1,500 educators globally have enrolled to build their expertise in teaching statistics and data science across STEAM disciplines (i.e., statistics, mathematics, data science, business, engineering, science, social studies). The platform supports asynchronous learning, allowing teachers to engage at their own pace while offering opportunities for community building and knowledge sharing. This paper discusses two frameworks that structure the online learning pathways within InSTEP.

DESIGNING FOR TEACHERS' LEARNING

The InSTEP online professional learning platform is envisioned and designed to be a place where teachers can enroll and learn a little, or a lot, on their own time, and at their own pace. InSTEP is not a static course to complete but is designed to allow educators to drop in. Teachers can complete part or all of a modularized learning experience and come back later to complete others, especially since new learning experiences are added to the platform a few times a year. Thus, we needed an organizational structure that would allow new content to be easily added, and that would provide clear, but flexible pathways teachers could choose to follow based on their own learning goals. To support teachers' ability to push the boundaries of traditional instruction, we developed and used two frameworks to inform the design of teachers' learning experiences in InSTEP (Figure 1). Below, we describe each framework and then illustrate how they are used to design features of the platform and frame teachers' learning through examples from within InSTEP. It is the combination of these two frameworks that are used to help educators learn about teaching statistics and data science and to see how content across the platform is interconnected to support their learning.

In: Birk, L., Loth, G., Jotzo, L., Binder, K., & Frischemeier, D. (Eds.) (2026). *Statistics and Data Science Education in STEAM. Proceedings of the Satellite Conference of the International Association for Statistical Education (IASE), September/October 2025, Münster, Germany*. © ISI/IASE.

a. Framework for engaging in a data investigation process



b. Framework for the seven dimensions of teaching statistics and data science.

Figure 1. Diagrams for two frameworks used to support teachers' learning in InSTEP

Framing Teachers' Learning to Work with Data

The first framework (Figure 1a) depicts a Data Investigation Process (Lee et al., 2022), which guides teachers through interrelated phases to investigate real world phenomena by: framing problems, gathering and processing data, exploring and visualizing data, considering models, and communicating results. A Data Investigation Process (Lee et al., 2022) is built on theory from statistics education and data science practices and processes. Within statistics education, several of these frameworks describe four phases: pose a question, collect or consider data, analyze data, and interpret results (e.g., Bargagliotti et al., 2020; Friel et al., 2006; Graham, 1987), while others describe a five-phase model, emphasizing planning and data exploration, as well as the role of visualization (Watson et al., 2018; Wild & Pfannkuch, 1999). Scholars in data science education have begun to articulate frameworks describing key practices and processes, typically involving six or seven phases. These often start with defining a problem and context, including key goals and attributes (EDC, 2014; Goldstein, 2017) and are followed by gathering, cleaning, transforming and managing data—often described as wrangling (Education Development Center [EDC], 2014) or processing raw data (Goldstein, 2017). Subsequent phases focus on exploring and analyzing data and conclude with communicating results, often through data visualization (EDC, 2014; Goldstein, 2017).

With an aim to support teaching and learning of statistics and data science, we draw on a 6-phase data investigation framework that synthesizes key elements from these prior frameworks in statistics education and the practices of data scientists (Lee et al., 2022). As shown in Figure 1a, the phases of a Data Investigation Process fit together like pieces of a puzzle for a holistic approach, requiring ongoing refinement and movement between the phases. Though most often dynamic, the process can also be linear or cyclic. Regardless of the entry point, the core focus remains on using data to solve a problem or answer a question within a context. In Lee et al.'s (2022) framework, ***Framing the problem*** begins by developing an understanding of the context and broader issues surrounding a phenomenon. One or more investigative questions are posed that anticipate variability in data and can be addressed using statistical methods. Understanding the broader context may be often revisited and questions may be refined. ***Considering and gathering data*** involves identifying and collecting data needed to answer a question, which involves attending to issues of data collection, measurement, study design, potential bias, and ethical considerations. ***Processing data*** requires organizing and preparing data for analysis. It includes addressing missing or erroneous values and applying techniques such as merging, sorting, grouping, transforming, or filtering data to support your investigation. ***Exploring and visualizing data*** uses visualizations (e.g., graphs, diagrams) and statistical summaries to explore data in context and involves looking for patterns, trends, and relationships among variables that relate to an investigative question. ***Considering models*** entails selecting and applying appropriate models (e.g., statistical measures, visualizations, predictive models, or distribution models) to answer questions, requiring one to evaluate variability, uncertainty, and underlying assumptions to ensure model viability.

Communicating and proposing action requires interpretation of results in the context of a problem and communicating evidence-based conclusions. It can also include making recommendations or outlining next steps.

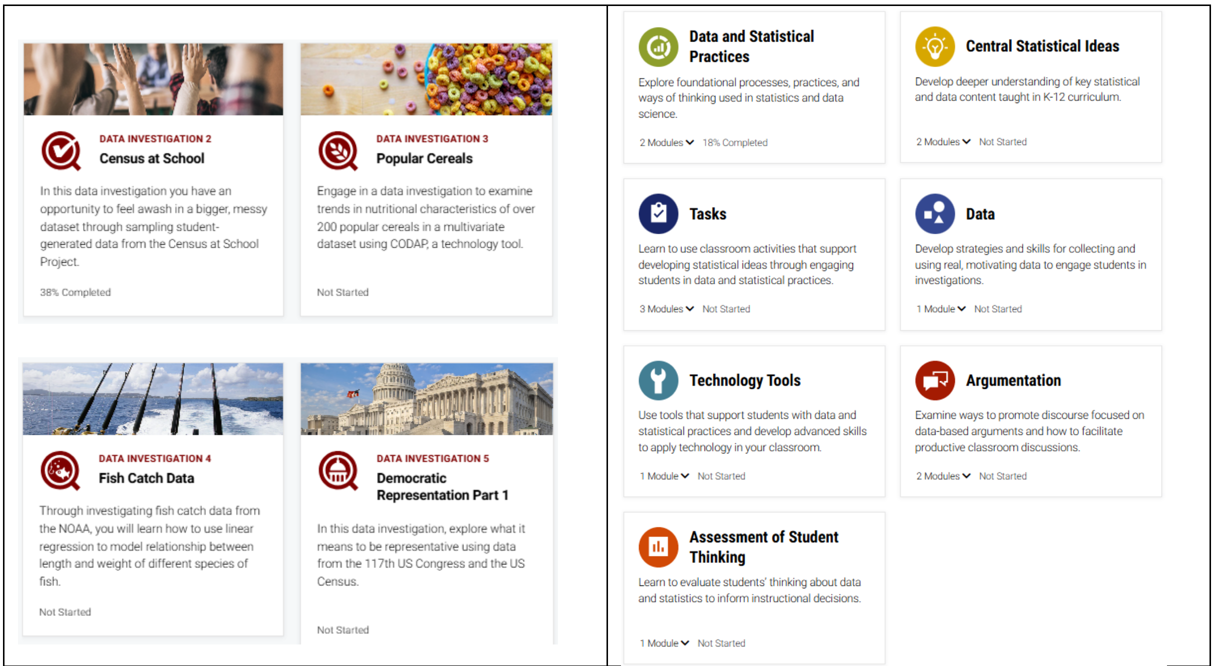
Framing Teachers' Learning to Create a Data Intensive Learning Environment for Students

The second framework, shown in Figure 1b, describes elements of an effective learning environment for statistics and data science based on decades of research. Traditional teaching of statistics often has less emphasis on the context of data, the process of ensuring good data is collected, and making claims about data that are uncertain in nature (Pierce & Chick, 2011). Eichler (2011) posited that the focus of teachers' intended curriculum in statistics can be considered on a continuum from traditionalists focused on procedures, to those wanting students to be prepared to use statistics in everyday life and focused on an investigative process tightly connected to contexts of real data. A goal in teachers' professional learning is to move teachers along this continuum, which requires impacting teachers' beliefs about the nature of statistics and learning goals for students. To support students in developing productive statistical thinking, teachers need to carefully reconsider the use of procedurally-oriented, teacher-centered learning environments. "Learning statistics is not about passively acquiring a set of facts and procedures but rather about actively constructing meaning and understandings of big ideas, ways of reasoning, and articulating arguments" (Ben-Zvi et al., 2018, p. 475). Ben-Zvi et al. (2018) suggest that focusing on how to bring together different dimensions of instructional design are needed to impact how statistics is taught. Building from others (e.g., Ben-Zvi et al., 2018; Cobb & McCain, 2004; Garfield & Ben-Zvi, 2009), we argue that effective instructional design should focus on seven interrelated dimensions and how each operates in relation to others: 1) the nature of data and statistics practices, 2) focus on central ideas; 3) use of well-designed tasks; 4) use of real data; 5) support for discourse and argumentation with data; 6) integration of technological tools; and, 7) understanding students' reasoning about data and statistics. We designed the diagram and set of icons shown in Figure 1b to illustrate our framework.

As teachers advance their understandings and skills in each dimension, they are supporting their overall development in teaching statistics and data science. There are two dimensions in the outer rim (see Figure 1b). ***Data and Statistical Practices*** is comprised of foundational practices, processes, and ways of thinking with data. The data investigation process is a critical component of this dimension. ***Central Statistical Ideas*** include core conceptual ideas necessary to support students' learning in statistics and data science classrooms (e.g., reasoning about variability, comparing distributions, and inferential reasoning). The inner five dimensions further support essential pedagogical strategies. Within the ***Tasks*** dimension, teachers learn to use classroom activities that support developing statistical ideas through engaging students in data and statistical practices. Teachers develop strategies and skills in the ***Data*** dimension for collecting and using real, motivating data to engage students in investigations. With such a wide array of tools available, teachers need to learn about different ***Technology*** and develop advanced skills to support students with data and statistical practices in their classrooms. A critical aspect of teaching statistics and data science includes ***Argumentation*** and learning to support students in developing data-based claims and how to facilitate productive classroom discussions. And finally, within ***Assessment***, teachers learn to evaluate students' thinking about data and statistics to inform their instruction.

Bringing These Frameworks Together to Design Professional Learning Experiences

These two frameworks work together to empower teachers through flexible, structured learning experiences. The integration of these two frameworks within the InSTEP platform provides a coherent and flexible structure that supports participants in building expertise in the teaching and learning of data science and statistics. This integrated design not only scaffolds meaningful engagement with core content but also empowers users with the freedom to tailor their learning journeys to their individual interests and needs. To help achieve this vision, we organized learning experiences by two types: 1) activities to develop teachers' content knowledge about working with data themselves to investigate phenomena that may be of interest to both them and their students, and 2) activities to develop teachers' pedagogical abilities to create data intensive learning environments to support students in learning statistics and data science. On the Learning Hub page within the platform, participants are presented with a series of tiles labeled as either Data Investigations (see sample in Figure 2a), or Self-Paced modules organized by the seven dimensions (Figure 2b). Data investigations provide a pathway for educators to experience investigations as a learner and doer, and the self-paced modules allow them to dig deeper into specific dimensions to advance their knowledge and skills in targeted ways.



a. Sample of 4 tiles describing data investigation learning experiences

b. Seven tiles that organize self-paced modules into corresponding dimensions.

Figure 2. Tiles on InSTEP's Learning Hub for Navigating to Two Types of Learning Experiences

There are currently six investigations for teachers to learn to work with data and key statistical modeling through exploring topics of: *science* (roller coaster characteristics such as speed, height, time, and track length), *science and recreational hobbies* (fishing restrictions related to length and weight of different species of fish), *health and societal impacts* (nutrition of cereal and ways they are marketed), *social practices of teenagers* (self-reported data about practices such as social media and texting), and *demographic representation of the U.S. Congress* (sex and race/ethnicity of members of congress and comparison to U.S. population). Within each investigation, teachers are guided through how to engage in the investigation themselves, with the language and puzzle pieces from the Data Investigation process used throughout to make explicit connections to the framework (Figure 3). In addition, within the investigations, if there is a strong connection to learning materials within the modules in a dimension, educators are encouraged to engage in that resource through a popup window (rather than leaving the investigation page). See the tile labeled "Thinking Through a Data Investigation" in Figure 3.

The Dimensions of Teaching Statistics and Data Science framework is used to design several key features of the platform. As shown by the tiles in Figure 2b, there are brief modules (~1.5-2 hours), currently 13, with more under development, organized within each of the seven dimensions. We use the color and icons from the framework diagram (Figure 1b) to orient users throughout the platform to their learning within these dimensions. For example, on the page describing two modules in the Data and Statistical Practices dimension (Figure 4), the color green is used along with an image of the framework diagram with just that dimension highlighted to indicate the focus of these modules.

Additionally, when educators register and create an account, they are asked to rank order their professional learning goals from a list (randomized when presented to users), which align with learning goals of the seven dimensions as well as data investigations. Sample learning goals include,

- Strengthen my understanding of how to engage students in the practices related to statistics and data science. (Data and Statistics Practices)

- Strengthen my understanding of key statistics and data concepts and skills. (Central Statistical Ideas)
- Improve my ability to lead productive discussions about important ideas related to data and statistics. (Argumentation)
- Improve my ability to collect and use real-world data to support student's learning in statistics and data science. (Data)

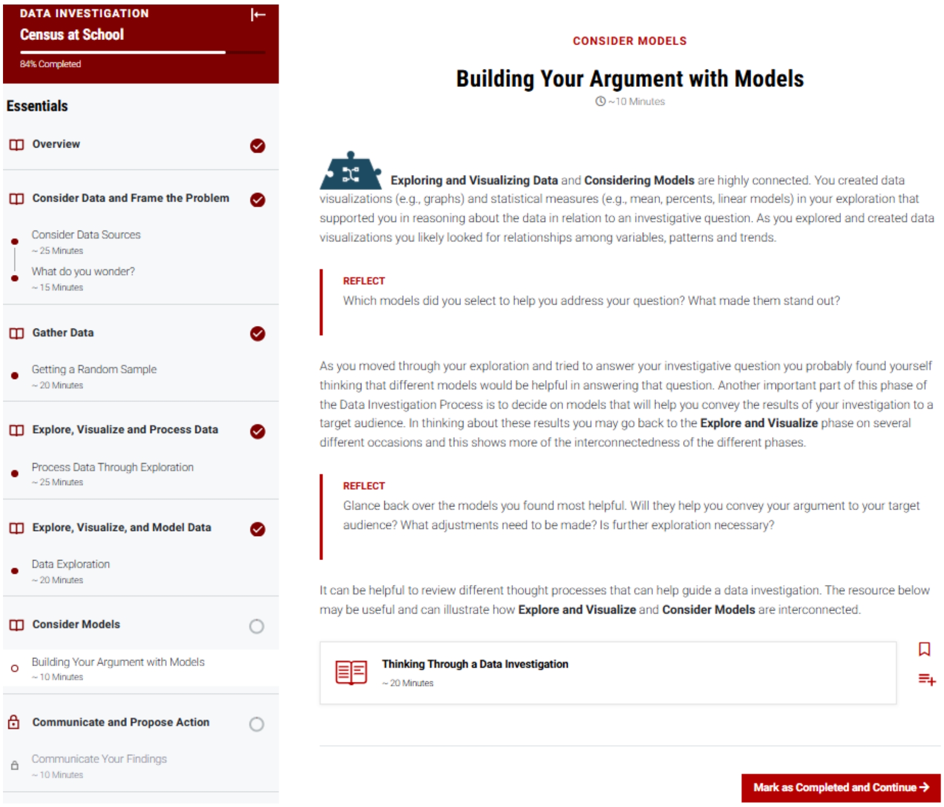


Figure 3. A Consider Models Page within the Census at School Data Investigation

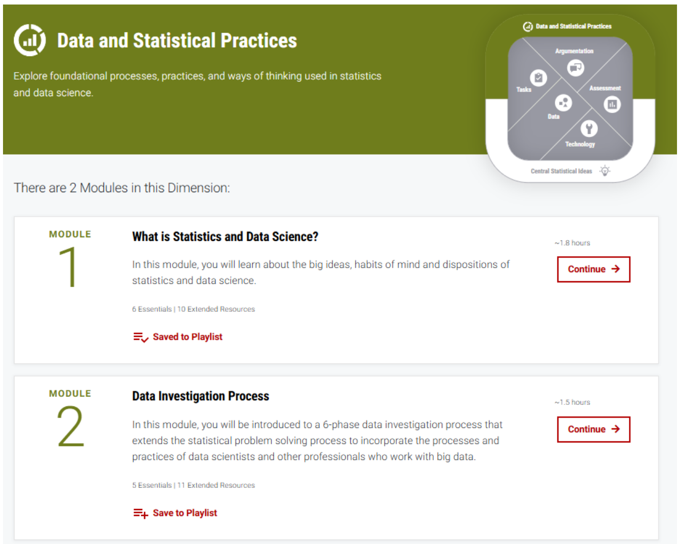


Figure 4. Two Modules within the Data and Statistical Practices Dimension

Along with other user demographics (e.g., grade level focus, subjects taught), an educator's ranked goals are used to generate a list of recommended learning experiences that are available on their Dashboard, thus linking their goals onto the framework and helping support teachers' engagement with the platform. As users engage and complete learning experiences, their log data informs subsequent recommendation (e.g., if they complete a module within a dimension, a second module in that dimension will be recommended).

To further illustrate the interplay of these two frameworks, consider the example of Comparing Distributions: A Learning Activity. This activity is situated in the ***Central Statistical Ideas*** dimension and nested within the *Comparing Distributions Module* as seen in Figure 5 through the left-hand navigation panel. This placement reflects the pedagogical emphasis of the activity (teaching comparing distributions) and guides users toward related content within the same instructional dimension. Once in the learning activity, the Data Investigation Process framework is used as subheadings with associated puzzle piece icons to help situate the learner to the data science activity. This reinforces the data science activities that permeate throughout the different Dimensions of Teaching Statistics and Data Science. The use of these two frameworks throughout the platform helps reinforce important activities of data science and statistics while also supporting navigation and scaffolded teachers' learning while allowing free choice to navigate the platform in the order the participant chooses.

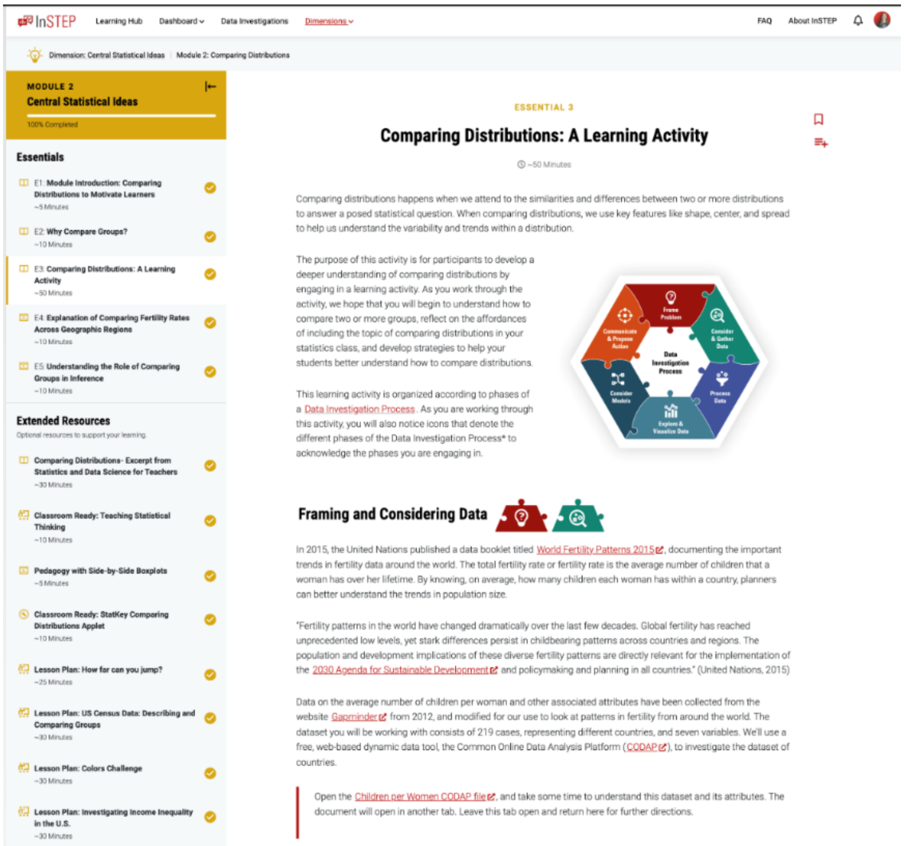


Figure 5. Page Within Comparing Distributions Module in Central Statistical Ideas dimension

CONCLUSION

To reiterate, the Dimensions in Teaching Statistics and Data Science framework helps organize learning activities according to seven key instructional dimensions, providing a cohesive lens through which educators can explore effective strategies for teaching data science and statistics. The Data Investigation Process framework, on the other hand, brings structure to individual data investigations and lessons within modules by highlighting the iterative and inquiry-based nature of data science. This dual use of frameworks not only strengthens conceptual understanding and engagement with data science practices but also enhances personalized learning support in the InSTEP platform. Participants can navigate the platform in a non-linear fashion, choosing activities that align with their professional goals or areas of interest, while still benefiting from a structured, research-based learning environment. The consistent application of these frameworks ensures that, even as participants exercise choice and agency, they remain anchored in best practices for teaching and learning in data science and statistics.

ACKNOWLEDGEMENT

The work described in this paper is supported by the National Science Foundation under Grant DRL 1908760 awarded to NC State University. Any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank Alex Dreier, as well as the instructional design team at RTI International that helped co-design and enact our vision within the InSTEP platform: Patrick Brown, Andrzej Proczka, and Britta Hansen. All images are from the platform and developed by the research and design team, and not copyrighted. InSTEP can be accessed at https://instepwithdata.org.

REFERENCES

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre-K-12 Guidelines for assessment and instruction in statistics education (GAISE) report II*. American Statistical Association and National Council of Teachers of Mathematics. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf

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. https://doi.org/10.1007/978-3-319-66195-7_16

Cobb, P., & McCain, K. (2004). Principles of instructional design for supporting the development of students' statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of developing statistical literacy, reasoning, and thinking* (pp. 375–395). Springer. https://doi.org/10.1007/1-4020-2278-6_16

Education Development Center. (2014). *Profile of a big-data-enabled specialist*. http://oceansofdata.org/our-work/profile-big-data-enabled-specialist.

Friel, S., O'Connor, W., & Mamer, J. (2006). More than "Meanmedianmode" and a bar graph: What's needed to have a statistical conversation? In G. Burrill and P. Elliott (Eds.), *Thinking and reasoning with data and chance: Sixty-eighth Yearbook* (pp. 117–137). National Council of Teachers of Mathematics.

Garfield, J., & Ben-Zvi, D. (2009). Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment. *Teaching Statistics, 31*(3), 72–77. https://doi.org/10.1111/j.1467-9639.2009.00363.x

Goldstein, A. (2017, January 14). *Deconstructing data science: Breaking the complex craft into it's simplest parts*. Mission.org. Retrieved September 25, 2018 from https://medium.com/the-mission/deconstructing-data-science-breaking-the-complex-craft-into-its-simplest-parts-15b15420df21

Graham, A. T. (1987). *Statistical investigations in the secondary school*. Cambridge University Press.

Hollebrands, K. F., & Lee, H. S. (2020). Effective designs of Massive Open Online Courses for mathematics teachers to support their professional learning. *ZDM Mathematics Education* 52(5), 859-875. https://doi.org/10.1007/s11858-020-01142-0

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. https://doi.org/10.24059/olj.v24i1.1992

Lee, H. S., Mojica, G. F., Thrasher, E. P., & Baumgartner, P. (2022). Investigating data like a data scientist: Key practices and processes. *Statistics Education Research Journal* 21(2), https://doi.org/10.52041/serj.v21i2.41

Sukol, S. (2024, March 18). *Beyond borders 2024: Primary and secondary data science education around the world*. Data Science 4 Everyone. https://www.datascience4everyone.org/post/beyond-borders-2024

Watson, J., Fitzallen, N., Fielding-Wells, J., & Madden, S. (2018). The practice of statistics. In D. BenZvi, K., Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 105–138). Springer. https://doi.org/10.1007/978-3-319-66195-7_4

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–248. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x