ORIGINAL PAPER

Emerging trends in statistics education

Gail Burrill[1] · Maxine Pfannkuch[2]

Abstract
The rapidly increasing capacity of technology to collect, organize, and manage data has spurred changes in the practice of statistics: new methods of collecting data, large data sets, new forms of data, different ways to visualize and represent data, and recognition of the importance of being able to understand and to communicate data-based arguments and findings from the perspective of data consumers and data producers. Using a narrative review based on Delphi methods, wee asked leading members of the statistics education community to describe trends they have observed in the field and to identify interesting and relevant papers related to those trends. We received 24 responses and over 200 suggestions for papers. Our analysis included papers published in journals, book chapters, conference proceedings, handbooks, and curricular documents. We focused on future directions for statistics education research, and thus included articles based on opinion or principles if the arguments made a strong case supported by evidence as to why the idea was needed. From our analysis of 50 papers in this review, we suggest four emerging themes in statistics education research, challenging what should be taught and suggesting new ways of thinking about the teaching and learning of statistics: Data Science, Visibilizing Statistical Concepts, Social Statistics, and New Contexts for Learning. The review focuses on articles from 2017-mid 2022 and highlights the relevance and importance of each theme. Our choice of a particularly important paper for each theme is annotated in the references.

1 Introduction

Research on the teaching and learning of statistics emerged in the late twentieth century (Zieffler, et al., 2018). Much of this early work came from the college level and was spurred by the changing nature of technology and its impact on the discipline and consequently on what was taught and how it was taught. Of note was the work of Tukey (1977), who introduced "exploratory data analysis", an approach involving exploratory techniques, many of them visual, to uncover patterns, underlying structure, and exceptions in a dataset. Key papers by Cobb and Moore (1997), Moore (1997), Wild and Pfannkuch (1999), Gal (2002), and Cobb (2007) laid the groundwork for further thinking about statistics education

and helped the field come to some consensus about important topics related to the development of statistical understanding. The groundwork included the acknowledgement that statistics was not a branch of mathematics but an independent discipline with its own distinctive modes of reasoning and that technology had shifted statistics "from overtly mathematical approaches to data analysis toward computationally intensive approaches" (Wild, et al., 2018). Statistical investigations (Wild & Pfannkuch, 1999) and literacy (Gal, 2002) were considered important foundations for learning statistics, including developing dispositions such as a critical stance. Furthermore, in a data-rich evidence-based world, there was recognition that all educated people needed to understand statistical ideas whether they were a professional user of statistics, a consumer of statistics, or a decision maker. To enculturate students into the thinking and practice of statisticians, Lehrer and English's (2018) proposal that students should engage in authentic activities that are approximations to professional practice and co-construct the practice of investigative inquiry alongside conceptual development has gained traction in statistics education research and many curricula (e.g., Ministry of Education, 2007). At

✉ Gail Burrill
burrill@msu.edu

Maxine Pfannkuch
m.pfannkuch@auckland.ac.nz

1   Program in Mathematics Education, Michigan State University, East Lansing, MI, USA

2   The University of Auckland, Auckland, New Zealand

the senior level of high school, computer simulation-based approaches such as bootstrapping and randomization tests (Cobb, 2007) instead of mathematical inferential methods began to be adopted in statistics education. For younger students, exploratory data analysis and statistical modelling tools (Biehler, et al., 2013) became more prevalent as statistics education and research sought to provide opportunities for students to investigate multivariate datasets and model situations and to develop statistical reasoning and conceptual understanding. However, statistical literacy in Gal's (2002) definition, whereby people, in response to diverse media reports, activate a set of critical questions to evaluate and communicate reasoned opinions on claims made, has gained limited attention in research and curricula.

Pfannkuch (2022) contended that statistics curricula are shaped and evolved in response to changes in technology, changes in the discipline, the needs and demands of society, research into students' learning and reasoning, and statisticians' calls for reforms. Building on previous research, Pfannkuch (2022), in reimagining curriculum approaches and anticipating possible changes, emphasized the need for fostering statistical reasoning and argumentation including understanding the data context, interrogating data, and using data as evidence for making and supporting claims for both inquiry and advocacy. On the assumption that technology was an integral part of statistics curricula, she argued curricular approaches should include (1) immersing students in data-rich environments, statistical investigations, and modeling, (2) critically evaluating data-based arguments in diverse media, including risk, and (3) facilitating accessibility of statistical concepts through interactive visualizations, learning how to scaffold students' reasoning, and providing coherent conceptual pathways. Technology has continued to advance since Pfannkuch (2022) reimagined curriculum approaches, and hence the landscape of statistics education has undergone considerable change as new directions and learning contexts emerge. Building on the work of Pfannkuch (2022) and the groundwork of the five key papers we identified above, this paper focuses on four emerging trends from 2017 to mid-2022.

2 Method

Using a narrative review based on Delphi methods (Puig & Adams, 2018) we selected an expert group composed of 26 members of the statistics education community from 15 countries based on their publication history, involvement in projects related to the teaching and learning of statistics, contributions to International Association for Statistical Education conferences, and mindful of geographic diversity. These individuals were asked to describe current trends they have observed in the field and to identify interesting

and relevant papers related to those trends. The experts we consulted identified over 200 resources as relevant including articles published in statistics or mathematics education journals, book chapters, conference proceedings, and curricular documents as well as publications in press (e.g., Ridgway, 2022) and projects currently being conducted (e.g., ProDaBi project, https://www.prodabi.de/en/). In addition, we looked at special issues such as those from *Statistics Education Research Journal* (2017, 2021, 2022), *ZDM* (2018), *Teaching Statistics* (2021), *Journal of Statistics and Data Science Education* (2021), and books such as Ben-Zvi et al. (2018) and Leavy et al. (2018). We focused on future directions for statistics education research and thus included articles based on opinion or principles if the arguments made a strong case supported by evidence as to why the idea was important. To help narrow the candidates for inclusion, we established the following criteria for selecting papers (Wilson, et al., 2001). The papers had to be:

- Published within the period from 2017 to mid-2022.
- Significant, challenging current thinking.
- Relevant to the identified trends noted by the expert group.
- Evidence based.
- Scholarly.

Studies were discarded for four reasons: (1) not directly related to one of the identified emerging trends; (2) focused on ideas that had been well covered in papers prior to 2017; (3) did not challenge current thinking; or (4) were based on a single program or country problem that was not applicable across the field of statistics education. In the end, we included 50 papers in this review that highlight the relevance and importance of each theme, challenge what should be taught, or suggest new ways of thinking about the teaching and learning of statistics. A particularly important paper for each theme is annotated in the references.

We used a reductive approach to analyzing the 24 responses received by way of inductive category formation (Mayring, 2014). A first pass through the suggestions, consistent with an open coding technique drawn from grounded theory (Strauss & Corbin, 1998) and referring to the five key papers and Pfannkuch (2022) as described above, enabled us to identify broad categories for emerging themes. Through many discussions and more detailed text analysis, we identified four categories, with the first three aligned to Pfannkuch's reimagined curricular approaches: Data Science, "Visibilizing" Statistical Concepts, Social Statistics, and New Learning Contexts. The impact of technology drives these trends as well as the dual idea of data consumer and data producer, recognizing that to participate as an informed citizen in today's world students must experience both. By data consumer we mean those who interact with

data-based information produced by someone else's efforts, and by data producer we mean those that engage in empirical investigations, interpret their own data, and report their conclusions (Gal, 2002). Figure 1 illustrates these emerging trends with associated subthemes.

As indicated in Fig. 1, the data science trend is characterized by continued advances in software,the connection to computational thinking, and a renewed emphasis on modeling, in particular predictive modeling. The visibilizing statistics concepts trend revealed the strong research focus on young learners, the power of interactive visualizations and the continued attention to inference. The social statistics trend highlighted the importance of attending to risk, critical literacy, and communication, whether in the media or for awareness and advocacy. Trends in the contexts for statistical learning include new methods of data collection and analysis, different forms of data, new representations, and curricula changes. The discussion below describes the four emerging trends in more detail and highlights the relevant papers.

3 Emerging trends

3.1 Data science

The increasing capacity of technology to collect massive amounts of data and to organize and manage these data has led to an increasing demand for what is called data science, which was identified as an emerging trend by over half of those we consulted. While most agree that what is meant by data science is a blend of statistics, mathematics, coding, and context, data science is not well defined (e.g., Gould, 2021). Gould argues that data science encompasses all of statistical reasoning and synthesizes major aspects of computational

thinking as, "an approach to problem solving that attempts to make problems computable" (p. 516) by breaking a problem into smaller steps that can be executed by a machine. Gould (2021) described a data science course offered to secondary students that engaged them with real data and introduced them to coding, arguing that coding can compactly and efficiently communicate statistical models but acknowledged the tension for students between computational thinking and statistical thinking. Supporting the importance of computing, Horton and Hardin (2021) strongly urge the integration of statistics and computing at all levels in their introduction to a special issue of the *Journal of Statistics and Data Science Education* on Computing in the Statistics and Data Science Curriculum. Further clarifying the relationshipbetween statistics and data science, Erickson et al. (2019) and Gould (2021) described elements of data-specific thinking that characterize what data science adds to statistics. For example, data collection has expanded from random sampling and random assignment to accessing secondary data, data that were primarily collected for another purpose, often unknown.

While researchers continue to use the data investigation process PPDAC (problem, plan, data, analysis, conclusions; Wild and Pfannkuch, 1999), several other organizing structures, closely related to PPDAC, have been suggested for implementing the data science process. Building on the work of practicing data scientists, Lee et al. (2022) identified key practices, processes, and dispositions integral to data investigations resulting in a six-phase framework for implementing data science: frame problem, consider and gather data, process data, explore and visualize data, consider models, and communicate and propose action. Attending more to the pedagogy involved in teaching data science, Burrill and Dick (2022) used a set of design and implementation principles for scaffolding understanding when collecting data by

Fig. 1 Emerging trends identified by expert group of statistics education researchers
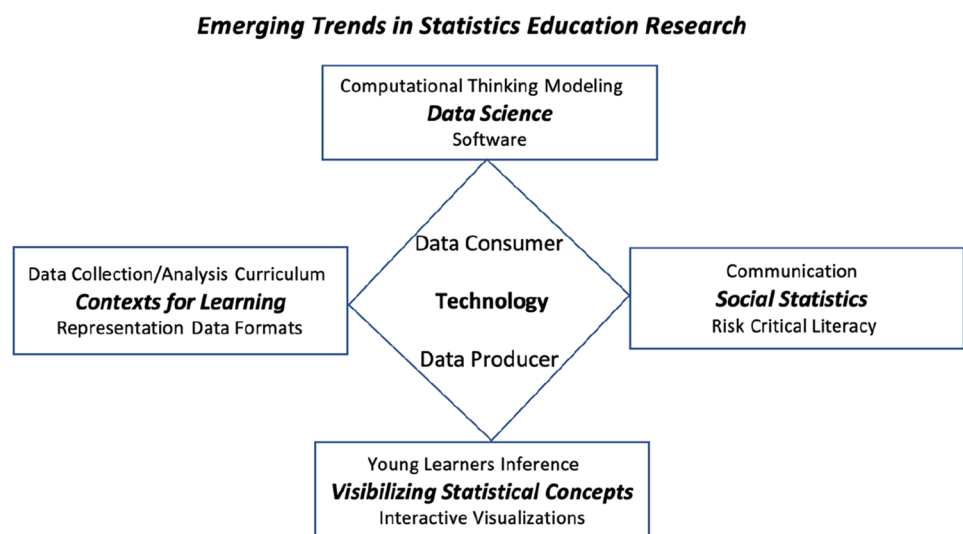
simulations to investigate herd immunity. From the perspective of the increasing collection of data about every aspect of people's lives, Utts (2021) addresses general principles related to ethics and what educators can do to instill sound ethical behavior in students to avoid adverse impact on people and society. Raising some of the same concerns as Utts, Weiland (2018) argues from the perspective of both the data consumer and producer that students also need to learn about and control their personal data trail.

In addition to the introduction of computational thinking with the emergence of data science, technology is essential for the management of large data sets. Wild (2018) discussed how to combine software such as TinkerPlots™ and CODAP (common online data analysis platform, a free online tool for data exploration and visualization) with point and click interfaces specifically designed for learning statistics and software that relies on the user communicating with the device through a language (coding). Given the expanding needs for teaching data science at the secondary level, researchers have adapted packages such as R or Python. For example, students in the Mobilize *Introduction to Data Science* project (Gould, 2021) analyzed data using Rstudio and Shiny apps. From a slightly different perspective but still making use of a simplified approach to coding, Fergusson and Wild (2021) discussed APIs (application programming interfaces), pieces of code from different sources that can be put together like building blocks to enhance applications, as a method for making data from an online database accessible to a broad range of secondary students.

The importance of modeling is supported by Lee et al.'s (2022) framework for the data science process, which identifies modeling as a key component of the process. From the statistical perspective, Pfannkuch et al. (2022) argued that statistical modeling provides a way to bridge data, chance, and context as a holistic approach for developing key statistical ideas and described the research knowledge base before 2018 with respect to statistical modeling, giving examples of research related to data modeling and modeling data observations with random generating devices to produce simulated distributions that mimic real data distributions. With the advent of data science, however, algorithmic predictive modeling, a process such as regression or decision trees that transforms the data, where the focus is on whether the model is effective in predicting outcomes (Fleischer et al., 2022), has become prominent. Fleischer et al. (2022) and Podworny et al. (2022) have embarked on the multigrade curriculum ProDaBi project in which secondary students worked with samples from large data sets to design a "training" model, beginning with hands-on experiences using data cards to build decision trees (algorithms for classifying data according to specified conditions), tested the model with another sample, then used classification rate data to evaluate the model. Students transitioned to CODAP and

eventually used Jupyter Notebooks (a programming environment related to Python that enabled them to write and execute code and that appear menu driven to the user) to apply an algorithm to generate decision trees and to evaluate their results (see https://youtu.be/9ol9HuTlXLw). Their instructional sequence is similar to the principles advocated by Burrill and Dick (2022). From the same perspective, Horton et al. (2022) described a form of supervised learning or predictive modeling by providing an overview of four approaches, based on the use of different technologies, to engage students in thinking about classifying email messages as spam or non-spam.

Several researchers have engaged teachers with algorithmic models. Zieffler et al. (2021) involved secondary teachers in algorithmic modeling activities designed to introduce classification trees (decision trees where the variables are discrete) and highlighted aspects of the statistical modeling process that appeared to be understood by the teachers and aspects that were challenging, in particular evaluating a model. They argued that algorithmic models offer opportunities to expand multivariate thinking, something typically not done in the past when activities involved small carefully designed data sets. Also working with secondary teachers, Fergusson and Pfannkuch (2018) illustrate how they introduced the teachers to predictive modeling involving simple linear regression models. The teachers used data from an API to develop and train their prediction models, which they then tested on a different set of data.

Because data science is such a new field, research has tended to focus on understanding what data science is and on related curricular materials. Investigating how students develop an understanding of algorithmic predictive modeling, effective ways to integrate computational thinking into the curriculum, and identifying ways to support teachers as they engage their students in data science activities are questions that need answers as we move forward.

3.2 Visiblizing statistical concepts

Researchers continue to show the difficulties students have in understanding statistical ideas and, in particular, they are identifying statistical concepts and argumentation that have been absent in traditional curricula and instruction (Biehler et al., 2018). That is, researchers are making visible previously unknown concepts such as posing a statistical investigative question (e.g., Arnold & Franklin, 2021) or creating dynamic interactive visualizations of concepts to make them visible to learners (e.g., Burrill 2018), which we call "visiblizing" statistical concepts. For example, statistical inference procedures and the concept of sampling distribution were traditionally introduced in the final year of high school, but technology has made it possible for students to be introduced much earlier to

inferential ideas (e.g., van Dijke-Droogers et al., 2020) and to a richer and bigger array of fundamental concepts such as sampling variability, sample distribution, and uncertainty. As Makar and Rubin (2018) demonstrated, statistical inference ideas and concepts can be nurtured and grown across the curriculum toward non-mathematical simulation-based statistical inference methods. Capitalizing on advances in technology, researchers have attended to inference by developing student reasoning and concepts through "visualizations, simulation, and powerful problem contexts" (Makar & Rubin, 2018).

The adoption of interactive visual approaches for learning statistics has led researchers to understand more about students' reasoning and conceptual processes. For instance, because simulation-based inference made 17–18-year-old students' conceptions visible, Case and Jacobbe (2018) were able to identify the difficulties students had in coordinating "the population …, the distribution of single sample, and the distribution of statistics collected from multiple samples". Similarly, Noll et al. (2018), using narrative theory, found that when students constructed models through interactive visual means, they were influenced by the story of the situation, guessing music notes correctly, and hence created and preferred models that were narrative in nature. The use of interactive visual data analysis tools to reveal students' intuitive reasoning is captured in Dvir and Ben-Zvi's (2018) analysis of two 12-year-olds reasoning from a scatterplot. Their analysis revealed that the students were holding conjectures about what the data would show, which they compared and assessed against the data they observed. Dvir and Ben-Zvi believe that focusing on students' acts of comparison will be helpful in enhancing their learning.

Visibilizing statistical concepts and actions in very young learners is a burgeoning area of research (e.g., Leavy, et al., 2018), an area that has had little attention since the early 2000s (e.g., Watson, 2006). The use of visual means to assist very young students to engage, reason, and interact with data is exemplified in the work of Leavy and Hourigan (2018), who explored 5–6-year-old students' inscriptions when collecting data by tracking the appearance of zoo animals in a video. They contended that the inscriptions served as a record of the event and represented the beginnings of abstract thoughts. Within the predictive modeling arena, Oslington et al. (2018) illustrate how high-ability 6–7-year-olds used self-portraits drawn by children in Kindergarten and Year 3 to develop a rule-based classification model. Students tested their model on larger sets of self-portraits and developed their own illustrations to support the rule-based model. The students used graphical representations, reflective statements, and tabular displays to inform their judgements regarding the strengths and weaknesses of their models.

Introducing data modeling or statistical investigations to younger students is emphasized in the work of Lehrer and English (2018), who synthesized

diverse research studies investigating the potential of inducting elementary grade children into the statistical practice of modeling variability in the light of uncertainty. Data modeling allows young students to co-construct concepts such as distribution, center, and variation and to reason visually using a combination of pencil-and-paper, and data analysis tools. Fielding-Wells (2018) explored this direction as a precursor to more formal work with inference, providing an example of how this worked with 10–11-year-old students engaged in constructing paper airplanes. Software platforms such as TinkerPlots, have also allowed young students to create visual models rather than equations, using random generating devices or factors that account for variation in order to produce simulations that mimic real data distributions. Kazak et al. (2018) showed the beginnings of this statistical modeling process in their study involving 11–12-year-old students, who collected and modeled the distances jumped by different sized origami frogs.

Because technology can remove learning barriers (e.g., constructing graphs by hand and mathematical procedures), *all* students can engage in activities that are authentic to the discipline, access statistical ideas from a very young age, and experience reasoning from and about data and modeling (Lehrer & English, 2018). With students using digital tools or other forms of visualizations and meaningful contexts, researchers have identified and explicated new statistical conceptual underpinnings and students' intuitive reasoning and conceptions or productive resources (Findley & Lyford, 2019) on which to build better conceptual infrastructure across the curriculum. Technology-rich visual learning environments are opening new avenues for researchers to explore and gain better understanding of students' reasoning, actions, and conceptual development and to identify new concepts, which often require the invention of new language for articulating reasoning and argumentation, that need to be part of the teaching repertoire.

3.3 Social statistics

Statistical literacy, which was elucidated by Gal (2002) from the perspective of societal needs, was recognized but gained only limited action in researching and implementing curricula changes (e.g., Callingham & Watson, 2017). The ProCivicStat Partners (2018) project, however, recognized the urgency of introducing Gal's ideas into the curriculum to strengthen democracies and that the nature of statistical literacy was changing. They argued the future of the world is at a critical juncture and that statistics educators must take up the challenge of improving student understanding of statistics about important societal concerns such as inequality and

injustice. Building on Gal's work they claimed that students need the knowledge, skills, and disposition to engage with civic statistics from a wide variety of information sources, where data are often "multivariate, aggregate, dynamic, and communicated through rich text and data visualizations and embedded in a social context" (p. 5). They suggest students need to develop a habit of mind to interrogate data-based information, to know what questions to ask, to communicate a reasoned opinion about whether the evidence provided is convincing, to discuss the implications of the findings, to make decisions in the presence of uncertainty, and to advocate for social action. Their research resulted in the provision of instructional resources and conceptual frameworks for educators to design tasks based on the needed skills for today's world (Ridgway, 2022).

Aligned with ProCivicStat Partners' (2018) concerns about improving student understanding of statistics in the media is the lack of attention to risk in the curriculum. Engel and Wilhem (2021) highlighted this need saying, "The Covid-19 crisis has impressively raised the general awareness that our social coexistence and political decisions are essentially based on data, the weighing of risks and thus on probability estimates". Brown et al. (2021) concurred, arguing that students should develop risk-know-how to protect themselves "against misinformation and misperception and encapsulate the concepts that equip people to access and deal [with such] information". There is a noticeable gap in research in this area at the school level (Bargagliotti et al., 2020). Martignon et al. (2022) delved into the area of risk literacy, focusing mainly on conditional probability, and illustrated how the use of specific iconic and interactive dynamic representations can make the ideas accessible. The understanding of risk, however, is a bigger field than conditional probability and relative risk, and work needs to be done on mapping out important ideas and a developmental pathway for enhancing students' notions of risk. A good place to start could be the risk-know-how framework of Brown et al. (2021) that identifies eight focal points for making sense of risk and includes the importance of knowing what is being discussed with respect to concepts such as numerator, denominator, or population and ideas such as recurrence intervals.

Given that citizens form opinions based on information from diverse communication channels, Gal and Geiger (2022) updated Gal's (2002) work and researched current statistical demands in the media. A main finding was that the primary means of communicating statistical information was text-based, that is written and spoken text, not graphs, and that students need to identify and comprehend statistical information embedded in text, which is often conveyed implicitly and in everyday language. Regarding everyday language, ProCivicStat Partners (2018) stated learners should be able "to deconstruct the diverse rhetorical and argumentative styles they will encounter when reading and interpreting statistical messages". Urgent attention, however, needs to be given to the increasingly worrying societal issue of disinformation. Souza and Araújo (2022) suggest educators need to broaden their perspective on how to interrogate data in an era of fake news and consider how to alert students to methods of communication that appear evidence based. They deconstructed the speeches of a journalist and the techniques he used to continually spread statistics-based disinformation about the pandemic and to shape public opinion including the specific language and tone of voice used to emotionally engage and persuade people to adopt his beliefs about the pandemic.

The necessity for statistics education to address social justice issues is raised in the work of several researchers. Weiland (2017) conceived the data consumer and data producer as "reading the world" and "writing the world", which he classified together as statistical literacy. He argued that critical literacy such as "interrogating social structures in the world" and "actively influencing and shaping" (p. 41) those structures should be combined with statistical literacy to form a new conception of statistics education as critical statistical literacy. Other researchers suggest that the purpose of students learning about statistical inquiry should be broadened to include the social context and how to advocate and argue from findings. From a socio-cultural perspective Zapata-Cardona (2018) argues students should conceive of themselves as being in an inextricable relationship with the world outside and their culture. In this view students are immersed in tasks where the data context allows them to "develop awareness of their surroundings and participation in the world" to become a critical citizen. To illustrate, Zapata-Cardona described a lesson with 12–13-year-old students who investigated a question about the nutritious value of the food they brought to school for lunch. In the analysis she contended there were nascent signs that students were not only developing statistical knowledge but also were becoming aware of their world by personalizing their findings with comments such as "we are overconsuming calories", that is, students were learning to become critical citizens.

Becoming aware of the world in which they live and adopting an evidence-based critical stance is promulgated further in Souza et al.'s (2020) research. Based on the theory of creative insubordination, they urge educators to realize the potential of statistics education to provide students with the tools to become political activists. Souza et al. demonstrated how two students investigated an issue from their environment, namely school canteen food. The teacher supported the students in repositioning their concerns about not liking the food into a social concern about food wastage. The students used the findings from their investigation to advocate for change. Souza et al. contend that statistics educators

need to assist students to develop the skills "to challenge the status quo, address inequity, and promote change". Rubel et al.'s (2017) research extends the idea of students becoming aware of their surroundings by considering the geo-mapping of data, visual displays that have traditionally not been part of statistics curricula, but which can graphically show spatial patterns on maps. Through the students geo-plotting their photographs of shops in their community, they realized that their community, compared to other neighborhoods, had different patterns, for example, more pawn shops, fewer bank facilities. Considering spatial data provided opportunities to open students' eyes to inequities in society and to groups who have power.

Such research reinforces the urgent need for educators to consider their role in assisting students to become knowledgeable citizens and advocates by learning how to investigate data-based situations as a producer and consumer of data. Multivariate thinking, nurturing the disposition to question and interrogate data-based information, and recognizing the need for data to judge a situation are some aspects where a research focus is needed. Moreover, there is an urgent need for research into how people interpret and understand statistical information embedded in text, both written and spoken. For students to navigate the world of data in which they live, curriculum developers and researchers need to pay attention to critical statistical literacy from a social statistics perspective.

3.4 New contexts for learning

The continuing evolution of technology offers new contexts for studying how students learn, for developing understanding, and for engaging students in statistical investigations. As noted by many of those we consulted, data today are multivariate and may consist of a variety of formats: images, text, sounds, dates, coded symbols, and locations. For example, Lee and Wilkerson (2018) discuss new forms of data accessible through emerging technologies. They specified four forms of data: data that are (1) collected through automated means, (2) algorithmically-generated, (3) non-quantitative, and (4) curated and publicly-available. These include data coming from the use of wearable sensor technologies, computational log data, such as clicks on websites or keystrokes on a personal computer, and remote laboratories. For each form, they discussed the implications for classroom practice, teacher preparation, and educational research. As another example, the Oslington et al. (2018) study described above used student art as data.

Several researchers recognize the potential of technology to create new forms of visualization. Andre and Lavicza (2019) identified opportunities to leverage technological innovations, offering examples of how to integrate data visualization and Open Data from primary through secondary school. Engel et al. (2020) illustrated how technology today provides tools for data visualization that allow the user to explore data without requiring deep mathematical knowledge and how interactive data visualizations can be used to promote conceptual understanding far beyond the graphical representations typically part of the school curriculum. They described data visualization as a rapidly evolving mix of science and art that opens up new avenues for communicating both in video and print, particularly with respect to data about the social and economic well-being of humans and the realization of civil rights. From a very different perspective, Rubel et al. (2021) looked at data representations from an equity viewpoint. They used examples related to the pandemic, defined dimensions of critically reading data and described how "data visualization's interrelated formatting, framing, and narrating might resonate with privileged perspectives, uphold certain hierarchies of power, perpetuate particular values over others, and guide towards certain decisions" (p. 265).

To improve and understand the conceptual infrastructure needed across curricula, researchers themselves are beginning to adopt technology-based analytical techniques for qualitative data to probe not only students' reasoning but also their actions when using data analysis tools. Frischemeier and Biehler (2018) used computer supported data analysis and stringent coding to determine pre-service teachers' reasoning and actions when they were comparing groups. Gould et al. (2017) described the coding of videos of teachers' reasoning and actions as they progressed through a statistical investigation centered around participatory sensing data in an interactive dashboard data analysis environment. Using discrete Markov chains, the researchers graphically depicted the strategic decisions the teachers made, which revealed the "importance of frequent questioning and crafting productive statistical questions" (p. 305). By delving into research data in a unique way, Gould et al. reinforced the notion that developing students' disposition to ask questions during statistical investigations was an important aspect, which Arnold and Franklin (2021) illustrate in their work. Other promising methods of uncovering previously inaccessible students' thinking and reasoning are the use of eye-tracking technology to collect data and machine learning algorithms for data analysis. Through eye-tracking 16-year-old students' strategies when interpreting histograms, Boels et al. (2019) found that students used a case-value plot interpretation, which is well-known in research, and a computational strategy, which is rarely reported.

Pfannkuch (2018) argued for reimagining the statistics curriculum, and other researchers agree. The importance of preparing students to deal with real data to address real questions has led educators such as Gould (2021), to claim there is a need to rethink secondary-level statistics education. Erickson et al. (2019) argued that specific data moves

such as filtering, merging two data sets, or making hierarchy should be introduced into the K-12 curriculum to prepare students for working with large data sets. At the middle school level, a study by Wilkerson and Laina (2018) raised issues of what should be in the curriculum when students repurposed a large publicly available data set to investigate questions about their local communities, reasoning about data and context through storytelling and explicit construction of hand-drawn or digital inscriptions. The repurposing raised issues related to data collection, selection, and manipulation such as recalculation, summarization, and data merging and purging but provided opportunities to connect to students' lives and to bring a new perspective to the role of statistics in classrooms. Burrill (2020) makes suggestions related to both mathematical and statistical content for integrating data/statistical literacy into the school curriculum.

The emphasis since 2018 on multivariate data and on computational thinking has opened new horizons for the teaching and learning of statistics, which has consequences for what should be taught. It also raises questions about how students learn in these new contexts and how researchers can utilize new contexts and tools to better understand the ways in which students learn.

4 Conclusion

Statistics education research began to flourish in the 1990s alongside the recognition that statistics was an independent intellectual discipline with its own unique ways of thinking and arguing. The changes wrought by technology in the last few years, however, have been so profound that school curricula and thus research need a major paradigm shift to educate future generations (Andre & Lavicza, 2019). Furthermore, statistics education needs to recognize its unique position in school curricula for developing students' awareness of exploitation, injustices, and inequities in society. Technological advances have thrust data science to the forefront, a major trend that will irrevocably change the statistics education landscape. At the same time the pandemic has provided the impetus to bring social statistics into prominence as a major trend to address. New learning contexts and visibilizing statistical concepts are two other major trends underpinned by the increasing prevalence of a wide array of tools to integrate into the learning environment, which open up many avenues to explore. Across the trends we emphasized that curricula need to enculturate students into multivariate and computational thinking, to deal with new forms of data, and to develop their disposition to ask questions.

We acknowledge that the trends are based on the perspectives of the experts that we approached, which are reflected in our final choice of articles. We recognize that every country is at a different stage in the development of their school

statistics curricula, in their access to technology, and the research that is possible to conduct. However, as many of the articles show, it is still possible to introduce data science ideas using unplugged activities with data cards (Podworny et al., 2022) and social justice issues in a pencil-and-paper environment (Zapata-Cardona, 2018).

Research is needed to inform curricula in the next five years based on the four trends we have identified, trends that challenge what should be taught in statistics, how it should be taught, and the role of statistics in school curricula. We hope that the trends we have identified will provide a catalyst for rethinking school statistics and for identifying research opportunities. Some avenues for future research including some questions that could be considered are:

- *More in-depth studies on introducing a coherent pathway of data science ideas across the curriculum.* As data science becomes integrated into curricula, what new tools and techniques will enable learners to use secondary data efficiently and purposefully, and make sense of and reason from such data? Given the increasing amount of content necessary to prepare students for a future in which data and statistical reasoning are essential, what are feasible strategies for incorporating these ideas into the curriculum? Should statistics curricula go beyond the boundary of a single subject and if so, how might this affect learning opportunities? As new methods for data collection emerge, what ethical principles do students need to activate and use?

- *More insight into fostering student statistical argumentation, advocacy, and risk-know-how.* What environments encourage learning how to make claims in a data-rich world, and how to critically evaluate written and spoken data-based evidence in diverse media and risk arguments? What is the impact of engaging students in social statistics? In the long term, how do student views change about the world in which they live? What is the developmental path for understanding risk? How do we include risk in the curriculum, so all students have sufficient knowledge to understand how to make informed decisions? How do we build students' intuitive conceptions into productive conceptions for understanding risk?

- *Visibilizing how students interact with and interpret data and conceptual visualizations.* How do learners interpret and interact with non-traditional interactive data visualizations? In technology-rich visual learning environments, what reasoning, actions and conceptual development are occurring and what new concepts need attention? What new language for articulating reasoning and argumentation might be needed? When young learners are introduced to core statistical concepts, how do these concepts become cultivated across the grades and become connected and deepened in students' reasoning

processes? How do we support the development of and make widely available software such as TinkerPlots and shape research into the learning opportunities provided by such software?

- *Impacts of new learning contexts and the development of learning content and pedagogical pathways.*How do learners interact with new forms of data? As new learning contexts emerge, in what ways do statistical thinking, reasoning and critical statistical literacy develop? What professional development preparation in terms of content, pedagogy, and technology do teachers need to engage with and teach new learning contexts? How can teachers from a wide range of countries and backgrounds be very quickly upskilled?

As Pfannkuch (2018) stated, curricula are shaped and challenged by technology changes, new societal perspectives, research findings on student reasoning and the constant evolution of statistical knowledge and practice. The advances in technology since 2017 are profound and challenging. We anticipate these advances will continue at an increasing rate and will require a concerted and rapid effort by researchers and curriculum developers to stay abreast of current and future developments that will be relevant to students' lives.

References

Andre, M., & Lavicza, Z. (2019). Technology changing statistics education: Defining possibilities, opportunities and obligations. *Electronic Journal of Mathematics and Technology, 13*(3), 253–264.

Arnold, P., & Franklin, C. (2021). What makes a good statistical question? *Journal of Statistics and Data Science Education, 29*(1), 122–130. https://doi.org/10.1080/26939169.2021.1877582

Bargaliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *guidelines for assessment and instruction in statistics education II (GAISE II)*

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clement, A. J. Bishop, C. Keitel, J. Kilpatrick, & A. Y. L. Leung (Eds.), *Third international handbook on mathematics education* (pp. 643–689). https://doi.org/10.1007/978-1-4614-4684-2_21

Ben-Zvi, D., Makar, K., & Garfield, J. (Eds.). (2018). *International handbook of research in statistics education*. Springer. https://doi.org/10.10007/978-3-319-66195-7.

Biehler, R., Frischemeier, D., Reading, C., & Shaughnessy, J. M. (2018). Reasoning about data. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp.139–192). Springer.

Boels, L., Bakker, A., & Drijvers, P. (2019). Eye-tracking secondary school students' strategies when interpreting statistical graphs. In M. Graven, H. Venkat, A. Essien, & P. Vale (Eds.), *Proceedings of the 43rd Conference of the International Group for the Psychology of Mathematics Education, Pretoria, South Africa, July 2019* (Vol.2, pp.113–120). PME.

Burrill, G. (2018). Building concept images of fundamental ideas in statistics: the role of technology. G. Burrill, & D. Ben-Zvi (Eds.), Topics and trends in current statistics education research:

International perspectives. Springer. https://doi.org/10.1007/978-3-030-03472-6_6

Burrill, G. (2020). Statistical literacy and quantitative reasoning: Rethinking the curriculum. In P. Arnold (Ed.), *New skills in the changing world of statistics education. Proceedings of Roundtable conference of the International Association for Statistical Education (IASE) July 2020*. https://doi.org/10.52041/iase.20104

Burrill, G., & Dick, T. (2022). Connecting mathematics to the world: Engaging students with data science. In J. Morska, & A. Rogerson (Eds.). *Building on the past to prepare for the future: Proceedings of the 16th International Conference of The Mathematics Education for the Future Project*, King's College, Cambridge.

Brown, T., Franklin, C., & Sierra, L. (2021). What is risk know-how? *Significance, 18*(6), 30–33.

Callingham, R., & Watson, J. (2017). The development of statistical literacy at school. *Statistics Education Research Journal, 16*(1), 181–201. https://doi.org/10.52041/serj.v16i1.223

Dvir, M., & Ben-Zvi, D. (2018). The role of model comparison in young learners' reasoning with statistical models and modeling. *ZDM-Mathematics Education, 50*(7), 1183–1196. https://doi.org/10.1007/s11858-018-0987-4

Case, C., & Jacobbe, T. (2018). A framework to characterize student difficulties in learning inference from a simulation-based approach. *Statistics Education Research Journal, 17*(2), 9–29. https://doi.org/10.52041/serj.v17i2.156

Cobb, G., & Moore, D. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly*, *104*(9), 801–823. https://doi.org/10.2307/2975286.

Cobb, G. (2007). The introductory statistics course: A ptolemaic curriculum? *Technology Innovations in Statistics Education*. https://doi.org/10.5070/T511000028

Engel, J., & Wilhelm, A. (2021). Data and statistics as basis for political decisions: Lessons to be learnt from the COVID-19 pandemic. In R. Helenius, & E. Falck (Eds.), *Statistics education in the era of data science. Proceedings of the Satellite conference of the International Association for Statistical Education (IASE), Aug–Sept 2021*.

Engel, J., Campos, P., Nicholson, J., Ridgway, J., & Teixeira, S. (2020). Visualizing multivariate data: Graphs that tell stories. In P. Arnold (Ed.), *New skills in the changing world of statistics education. Proceedings of the Roundtable Conference of the International Association for Statistical Education (IASE), July 2020*

Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data moves. *Technology Innovations in Statistics Education*. https://doi.org/10.5070/T5121038001. This paper defines and gives examples of data moves, actions made possible by digital tools designed to facilitate the manipulation and analysis of large, complex datasets. The authors argue data moves should be part of data analysis, give four recommendations to consider in shaping curriculum and instruction, and identify potential research questions that should be addressed in moving forward.

Fergusson, A., & Pfannkuch, M. (2022). Introducing high school statistics teachers to predictive modelling and APIs using code-driven tools. *Statistics Education Research Journal, 21*(2), 8–8. https://doi.org/10.52041/serj.v21i2.49

Fleischer, Y., Biehler, R., & Schulte, C. (2022). Teaching and learning data-driven machine learning with educationally designed Jupyter Notebooks. *Statistics Education Research Journal, 21*(2),7–7.

Fielding-Wells, J. (2018). Dot plots and hat plots: Supporting young students emerging understandings of distribution, center and variability through modeling. *ZDM-Mathematics Education, 50*(7), 1125–1138. https://doi.org/10.1007/s11858-018-0961-1

Findley, K., & Lyford, A. (2019). Investigating students' reasoning about sampling distributions through a resource perspective. *Statistics Education Research Journal, 18*(1), 26–45. https://doi.org/10.52041/serj.v18i1.148

Frischemeier, D., & Biehler, R. (2018). Pre-service teachers comparing groups with TinkerPlots–An exploratory laboratory study. *Statistics Education Research Journal, 17*(1), 35–60. https://doi.org/10.52041/serj.v16i2.192

Fergusson, A., & Wild, C. (2021). On traversing the data landscape: Introducing APIs to data-science students. *Teaching Statistics, 43*(1), 71–83. https://doi.org/10.1111/test.12266

Gal, I. (2002). Statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1–52. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x.

Gal, I., & Geiger, V. (2022). Welcome to the era of vague news: A study of the demands of statistical and mathematical products in the COVID-19 pandemic media. *Educational Studies in Mathematics, 111*(1), 5–28. https://doi.org/10.1007/s10649-022-10151-7. This research describes a rigorous systematic analysis of actual statistical and mathematical demands of diverse media items with which people need to engage. The authors identify nine categories of media demands and note that most of the communication is done in words. The paper lists its substantial contributions to statistics education including the urgent need to revise current curricula, particularly in the area of text-based media items.

Gould, R., Bargagliotti, A., & Johnson, T. (2017). An analysis of secondary teachers' reasoning with participatory sensing data. *Statistics Education Research Journal, 16*(2), 305–334. https://doi.org/10.52041/serj.v16i2.194

Gould, R. (2021). Toward data-scientific thinking. *Teaching Statistics, 43*(1), 11–22. https://doi.org/10.1111/test.12267. Gould argues with explication that data science curricula should center on developing students' "data scientific thinking," which consists of a core of statistical thinking, components of computational thinking, and "a dash of mathematical thinking." He describes an example of such a curriculum, and the final section provides consequences and challenges that arise from this particular view of data science.

Horton, N., Chao, J., Finzer, W., & Palmer, P. (2022). Spam four ways: Making sense of text data. *CHANCE, 35*(2), 32–40. https://doi.org/10.1080/09332480.2022.2066414

Horton, N. J., & Hardin, J. S. (2021). Integrating computing in the statistics and data science curriculum: Creative structures, novel skills and habits, and ways to teach computational thinking. *Journal of Statistics and Data Science Education, 29*, S1–S3. https://doi.org/10.1080/10691898.2020.1870416

Kazak, S., Pratt, D., & Gökce, R. (2018). Sixth grade students' emerging practices of data modelling. *ZDM–Mathematics Education, 50*(7), 1151–1163. https://doi.org/10.1007/s11858-018-0988-3

Lehrer, R., & English, L. (2018). Introducing children to modeling variability. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 229–260). Springer. https://doi.org/10.1007/978-3-319-66195-7 The article focuses on a data modeling framework, used by many researchers in the field, which suggests how approximations to professional practices of modeling and inference can be introduced to elementary students. Based on diverse research, the authors argue for the need to involve students in practices of inquiry, visualizing through data displays and attending to modeling and measuring variability.

Lee, H., Mojica, G., Thrasher, E., & Baumgartner, P. (2022). Investigating data like a data scientist: Key practices and processes. *Statistics Education Research Journal, 21*(2), 3–3. https://doi.org/10.52041/serj.v21i2.4

Leavy, A., & Hourigan, M. (2018). Inscriptional capacities and representations of young children engaged in data collection during a statistical investigation. In A. Leavy, M. Meletiou-Mavrotheris, & E. Paparistodemou (Eds.), *Statistics in early childhood and primary education: Supporting early statistical and probabilistic thinking* (pp.89–107). Springer.

Leavy, A., Meletiou-Mavrotheris, M., & Paparistodemou, E. (Eds.). (2018). *Statistics in early childhood and primary education: Supporting early statistical and probabilistic thinking* Springer. https://doi.org/10.1007/978-981-13-1044-7

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.Although directed at science education, thiscommissioned paper discusses new and emerging forms of data made possible byadvances in technology from an interdisciplinary perspective. The review putsinto context how the nature of data available to students is changing, forexample, sourced data from public datasets or photograph data. The discussionincludes implications for classroom instruction

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173

Martignon, L., Frischemeier, D., McDowell, M., & Christoph Till, T. (2022). Dynamic, interactive trees and icon arrays for visualizing risks in civic statistics. In J. Ridgway (Ed.), *Statistics for empowerment and social engagement: Teaching civic statistics to develop informed citizens* (pp. 477–501). Springer. https://doi.org/10.1007/978-3-031-20748-8_19

Makar, K., & Rubin, A. (2018). Learning about statistical inference. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp 261–294). Springer. https://doi.org/10.1007/978-3-319-66195-7_8

Ministry of Education. (2007). *The New Zealand curriculum*. Wellngton, NZ: Learning Media.

Moore, D. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123–165. https://doi.org/10.1111/j.1751-5823.1997.tb00390.x.

Noll, J., Clement, K., Dolor, J., Kirin, D., & Petersen, M. (2018). Students' use of narrative when constructing statistical models in TinkerPlots. *ZDM–Mathematics Education, 50*(7), 1267–1280. https://doi.org/10.1007/s11858-018-0981-x. Set in narrative discourse literature, the authors show how students privileged a more communicative and narrative form of thinking to construct statistical models in TinkerPlots. Because the students preferred models that were narrative in nature and their use of narrative seemed to support them toward productive statistical modeling, their findings have important pedagogical implications for teaching and curricula.

ProCivicStat, P. (2018). Engaging civic statistics: A call for action and recommendations. A product of the ProCivicStat Project. http://IASE-web.org/ISLP/PCS

Oslington, G., Mulligan, J., & Van Bergen, P. (2018). Young children's reasoning through data exploration. In V. Kinnear, M. Y. Lai, & T. Muir (Eds.), *Forging connections in early mathematics teaching and learning* (pp. 191–212). Springer. https://doi.org/10.1007/978-981-10-7153-9_11

Pfannkuch, M., Ben-Zvi, D., & Budgett, S. (2018). Innovations in statistical modeling to connect data, chance and context. *ZDM–Mathematics Education, 50*(7), 1113–1123. https://doi.org/10.1007/s11858-018-0989-2

Pfannkuch, M. (2018). Reimagining curriculum approaches. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp.387–413). Springer. This overview describes why and how statisticalmodeling is a research environment to learn about students' statisticalreasoning; statistical modeling provides a wayto bridge data, chance, and context as a holistic approach for developing keystatistical ideas. The nature and purposes for statistical modeling aswell as several theoretical frameworks are discussed from this perspective

Podworny, S., Fleischer, Y., & Hüsing, S. (2022). Grade 6 students' perception and use of data-based decision trees. In S. A. Peters, L. Zapata-Cardona, F. Bonafini & A. Fan (Eds.), *Bridging the gap: Empowering and educating today's learners in statistics. Proceedings of the Eleventh International Conference on Teaching Statistics, Rosario, Argentina*. https://doi.org/10.52041/iase.icots11.T2H3

Puig, A., & Adams, C. (2018). Delphi technique. In B. Frey (Ed.), *The sage encyclopedia of educational research, measurement, and evaluation*. Washington, DC: Sage Publications. https://doi.org/10.4135/9781506326139

Ridgway, J. (Ed.). (2022). *Statistics for empowerment and social engagement: Teaching civic statistics to develop informed citizens* Springer. https://doi.org/10.1007/978-3-031-20748-8

Rubel, L., Nicol, C., & Chronaki, A. (2021). A critical mathematics perspective on reading data visualizations. Reimagining through reformatting, reframing, and renarrating. *Educational Studies in Mathematics, 108*(1–2), 249–268. https://doi.org/10.1007/s10649-021-10087-4

Rubel, L., Hall-Wiekert, M., & Lim, V. (2017). Making space for place: Mapping tools and practices to teach for spatial justice. *Journal of the Learning Sciences, 26*(4), 643–687. https://doi.org/10.1080/10508406.2017.1336440

Souza, L., & Araújo, J. (2022). Fake news phenomenon: Formation of beliefs under pragmatic optics and mathematical education. *Acta Scientiae, 24*, 1–29. https://doi.org/10.17648/acta.scientiae.6602

Souza, L., Lopes, C., & Fitzallen, N. (2020). Creative insubordination in statistics teaching: Possibilities to go beyond statistical literacy. *Statistics Education Research Journal, 19*(1), 73–91. https://doi.org/10.52041/serj.v19i1.120. The paper reports on a study showing how two students become agents of change for a problem they identified in their social and cultural context, illustrating how data can be used to empower and transform lives. The authors used an innovative narrative analysis method reporting their results through a google translation of the Portuguese written narrative to maintain cultural connections.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*, 2nd Ed. Washington, DC: Sage Publications.

Tukey, J. W. (1977). Exploratory data analysis addision-wesley. *Reading, Ma, 688*, 581–582.

Utts, J. (2021). Enhancing data science ethics through statistical education and practice. *International Statistics Review, 89*(1), 1–17. https://doi.org/10.1111/insr.12446

van Dijke-Droogers, M., Drijvers, P., & Bakker, A. (2020). Repeated sampling with a black box to make informal statistical inference accessible. *Mathematical Thinking and Learning, 22*(2), 116–138. https://doi.org/10.1080/10986065.2019.1617025

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics, 96*(1), 33–47. https://doi.org/10.1007/s10649-017-9764-5

Wilson, S., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps and recommendations*. Center for the Study of Teaching and Policy, University of Washington.

Wilkerson, M. H., & Laina, V. (2018). Middle school students' reasoning about data and context through storytelling with repurposed local data. *ZDM–Mathematics Education, 50*(7), 1223–1235. https://doi.org/10.1007/s11858-018-0974-9

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

Wild, C., Utts, J., & Horton, N. (2018). What is statistics? In D. Ben-Zvi, D. K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp.5–36). Springer. https://doi.org/10.1007/978-3-319-66195-7_1

Wild, C. (2018). Gaining iNZights from data. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10), July 2018, Kyoto, Japan*.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.

Zieffler, A., Justice, N., delMas, R., & Huberty, M. (2021). The use of algorithmic models to develop secondary teachers' understanding of the statistical modeling process. *Journal of Statistics and Data Science Education, 29*(1), 131–147. https://doi.org/10.1080/26939169.2021.1900759

Zapata-Cardona, L. (2018). Students' construction and use of statistical models: A socio-critical perspective. *ZDM–Mathematics Education, 50*(7), 1213–1222. https://doi.org/10.1007/s11858-018-0967-8

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. https://doi.org/10.1007/978-3-319-66195-7_2