Learning to reason from samples

Dani Ben-Zvi · Arthur Bakker · Katie Makar

Abstract The goal of this article is to introduce the topic of *learning to reason from samples*, which is the focus of this special issue of Educational Studies in Mathematics on *statistical reasoning*. Samples are data sets, taken from some wider universe (e.g., a population or a process) using a particular procedure (e.g., random sampling) in order to be able to make generalizations about this wider universe with a particular level of confidence. Sampling is hence a key factor in making reliable statistical inferences. We first introduce the theme and the key questions this special issue addresses. Then, we provide a brief literature review on reasoning about samples and sampling. This review sets the grounds for the introduction of the five articles and the concluding reflective discussion. We close by commenting on the ways to support the development of students' statistical reasoning on samples and sampling.

Keywords Informal statistical inference · Sample · Sampling · Statistics education · Statistical reasoning

1 The importance of understanding samples and sampling

The pervasiveness of data in everyday life is a global trend. People are encircled by statistics in their everyday life and must become savvy consumers of data (Watson, 2002). In the workplace, data are vital for quality control, monitoring and improving productivity, and anticipating problems (Bakker, Kent, Noss, & Hoyles, 2009). With data increasingly being used to add or verify credibility, there are new pressures for schools to prepare both citizens and professionals to be able to create and critically evaluate data-based claims. Progress in the understandings of teaching and learning of statistical reasoning and the availability of high-quality technological tools for learning statistics have enabled the relatively young field of statistics education to integrate and readily capitalize on these advances (Ben-Zvi, 2000).

D. Ben-Zvi
Faculty of Education, University of Haifa, Haifa, Israel
e-mail: dbenzvi@univ.haifa.ac.il

A. Bakker (✉)
Freudenthal Institute, University of Utrecht, Utrecht, The Netherlands
e-mail: a.bakker4@uu.nl

K. Makar
School of Education, University of Queensland, Brisbane, Australia

Taking representative samples of data and using samples to make inferences about unknown populations are at the core of statistics. An understanding of how samples vary (sampling variability) is crucial in order to make reasoned data-based estimates and decisions. Learning to reason about samples and sampling includes making distinctions between samples and populations, developing notions of sampling variability by examining similarities and differences between multiple samples drawn from the same population, and examining distributions of sample statistic (Saldanha & Thompson, 2002; Wild & Pfannkuch, 1999). Looking at distributions of sample means for many samples drawn from a single population allows us to see how one sample compares to the rest of the samples, leading us to determine if a sample is surprising (unlikely) or not surprising. This is an informal precursor to the more formal notion of p value that comes with studying statistical inference (Garfield & Ben-Zvi, 2008).

Comparing means of samples drawn from the same population also helps build the idea of sampling variability, which leads to the notion of sampling error, a fundamental component of statistical inference whether constructing confidence intervals or testing hypotheses. Sampling error indicates how much a sample statistic may be expected to differ from the population parameter it is estimating. The sampling error is used in computing margins of error (for confidence intervals) and is used in computing test statistics (e.g., the t statistic when testing hypotheses).

The two central ideas of sampling—sampling representativeness and sampling variability—have to be understood and carefully balanced in order to understand statistical inference. Rubin, Bruce, and Tenney (1990) cautioned that over reliance on sampling representativeness leads students to think that a sample tells us everything about a population, while over reliance on sampling variability leads students to think that a sample tells us nothing useful about a population. Students should be given the opportunity to engage in the "middle ground between representativeness and variability" (Shaughnessy, 2007, p. 977) with increasing levels of certainty and confidence. In other words, we should help students develop the seeds of probabilistic language to articulate "not nothing, not everything, but something" (Rubin et al., 1990, p. 314) while making informal statistical inferences.

In fact, the ideas of sample and sampling distribution build on many core concepts of statistics, and if these concepts are not well understood, students may never fully understand the important ideas of sampling. Based on Wild and Pfannkuch's (1999) framework for statistical thinking, Pfannkuch (2008, p. 4) noted that students' statistical reasoning about sampling involves many underlying statistical concepts. For example, the fundamental ideas of distribution and variability underlie an understanding of sampling variability (how individual samples vary) and sampling distribution (the theoretical distribution of a sample statistic computed from all possible samples of the same size drawn from a population). The idea of center (average) is also involved (understanding the mean of the sampling distribution) as is the idea of model (the Normal Distribution as a model that fits sampling distributions under certain conditions). We also interpret empirical sampling distributions of simulated or collected data in similar ways, sometimes referring to these as, for example, a distribution of many sample means, rather than referring to it as a (theoretical) sampling distribution. Samples and sampling variability also build on basic ideas of randomness and chance, or the study of probability. Finally, sample size is related to the Law of Large Numbers, the fact that larger samples better represent the population from which they were sampled, and their sample statistics are closer to the population parameters. It is therefore important that students understand samples and sampling, and their interrelationships with other key concepts in order to make statistical inferences (Bakker, 2004).

There is a growing network of statistics education researchers who are interested in studying the development of students' statistical reasoning. The International Collaboration for Research on Statistical Reasoning, Thinking, and Literacy (SRTL, http://srtl.info) is a community of researchers and statistics educators studying the nature and development of students' statistical literacy, reasoning, and thinking, and exploring the challenges posed to educators at all levels in supporting students to achieve these goals (Ben-Zvi & Garfield, 2004a). The topics of these research studies conducted by members of the SRTL community reflect the shift in emphasis in statistics instruction from developing procedural understanding, that is, statistical techniques, formulas, computations, and procedures, to developing conceptual understanding and statistical literacy, reasoning, and thinking (Garfield & Ben-Zvi, 2004c). Over the last 15 years, through conference participation and collaborative projects, the SRTL members have become familiar with the research conducted by the group, advancing significantly the research on statistical reasoning, studying fallow fields in statistics education. The group has been fertile in producing publications, including several books (Ben-Zvi & Garfield, 2004b; Garfield & Ben-Zvi, 2008; Zieffler, Ben-Zvi, Chance, Garfield, & Gould, 2015); special issues of statistics and mathematics education journals: reasoning about variability in Statistical Education Research Journal (SERJ) (Ben-Zvi & Garfield, 2004a, b, c; Garfield & Ben-Zvi, 2005); reasoning about distribution in SERJ (Pfannkuch & Reading, 2006); informal inferential reasoning in SERJ (Pratt & Ainley, 2008); the role of context in informal statistical inference in Mathematical Thinking and Learning (Makar & Ben-Zvi, 2011) ; as well as many journal and conference proceedings papers.

During the past decade, the SRTL community has thus followed and shaped the research on *informal statistical inference* (ISI) (Pfannkuch, 2006). ISI has been characterized as a generalized conclusion expressed with uncertainty and evidenced by, yet extending beyond, available data (Makar & Rubin, 2009). Over the past years, a group of researchers have come to focus on *sample* and *sampling* as being at the heart of statistical inference and relevant at all levels of schooling, even in the early years.

This special issue of Educational Studies in Mathematics aims at presenting state-of-the-art studies on students' understanding of samples and sampling when making informal statistical inferences. The contributions were invited among members of SRTL that devoted a conference to this theme (SRTL7). This conference was hosted by Arthur Bakker in the Netherlands and attended by 22 participants from 7 countries. This special issue addresses the following questions:

- How does students' reasoning about data, samples, and sampling develop in the context of learning to make informal statistical inferences?
- What are innovative and effective approaches, tasks, tools, or sequences of instructional activities that may be used to promote students' understanding of reasoning about samples and sampling in making statistical inferences?

2 Reasoning about samples and sampling

The idea of a sample is likely to be familiar even to young students. They have all taken samples (e.g., tasting a food sample) and have an idea of a sample as something that is drawn from or represents something bigger. Students may have a fairly good sense that each sample may differ from the other samples drawn from the same larger entity. However, they have difficulty making the transition to the formal statistical meaning of sample (Watson & Moritz,

2000), understanding the behavior of samples when they study statistics, how they relate to a population, and what happens when many samples are drawn and their statistics accumulated in a sampling distribution (Garfield & Ben-Zvi, 2008). Even though sampling and sampling distributions are so important in statistics and challenging to students, there has been somewhat less research attention on this topic compared to other statistical concepts, such as distribution, variability, correlation, and regression (Garfield & Ben-Zvi, 2007).

In the following section, we review several studies that examined how students understand and misunderstand ideas of samples and sampling.

2.1 Studies involving school students

A study of students' conceptions of sampling in upper elementary school by Jacobs (1999) suggested that students understood the idea that a sample is a small part of a whole and that even a small part can give you an idea of the whole. Watson and Moritz (2000) also studied children's intuitive ideas of samples, and identified six categories of children's thinking about this concept. They point out that while students have a fairly good "out of school" understanding of the concept of sample, they have difficulty making the transition to the formal, statistical meaning of this term and the related connotations and conceptualizations. For example, students can make appropriate generalizations from a small sample of food to the larger entity from which it was drawn, but these intuitive ideas do not generalize to the notion of sampling variation and the need for large, random, and representative samples in making statistical estimates and inferences. Watson and Moritz (2000) suggest making explicit these differences, for example between taking a small food sample which represents a homogeneous entity, with taking a sample from the population of students to estimate a characteristic such as height, which is a population that has much variability.

Watson (2004), in a review of research on reasoning about sampling, describes how students often concentrate on "fairness," and prefer biased sampling methods such as voluntary samples because they do not trust random sampling as a process producing fair samples. Saldanha and Thompson (2002) found both of these types of conceptions on sampling in high school students. In a classroom teaching experiment, they discovered that presenting and using the concept of sampling as part of a repeated process, with variability from sample to sample, supported the understanding of distribution, which is required to understand sampling distributions.

In a teaching experiment with eighth grade students, Bakker (2004) was able to help students understand that larger samples are more stable (less variable) and better represent the population, using a sequence of "growing samples" activities. Growing samples is an instructional heuristic mentioned by Konold and Pollatsek (2002), worked out by Bakker (2004, 2007) and Bakker and Frederickson (2005), and elaborated by Ben-Zvi (2006) and Ben-Zvi, Aridor, Makar and Bakker (2012). In this approach, students are gradually introduced to increasing sample sizes that are taken from the same population. For each sample, they are asked to make sense of it and make an informal inference. They then predict what would remain the same and what would change in the following larger sample (for a visual description of this process, see Fig. 1). Thus, students are required to search for and reason with stable features of distributions or variable processes, and compare their hypotheses regarding larger samples with their observations in the data. They are also encouraged to think about how certain they are about their inferences.

What is the pedagogical rationale for this task design? When asked how to investigate a particular problem, students often suggest asking a few people, often mentioning low numbers. Allowed to begin with small samples (e.g., $n=8$), students are expected to experience the
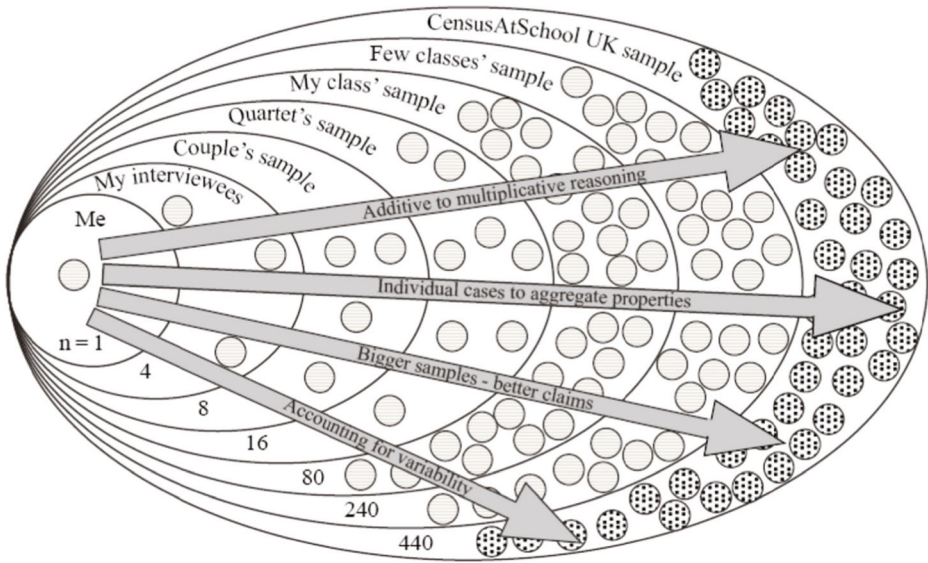
Fig. 1 A growing samples sequence and the change in students' reasoning about informal inference (Ben-Zvi, 2006, p. 4)

limitations of what they can infer about this current sample. Ben-Zvi et al. (2012) view it as a useful pedagogical tool to sensitize and slowly introduce students to the decreasing variability of apparent signals in samples of increasing sizes. In each stage, students are given additional data that creates an updated sample and asked to draw conclusions, speculating what can be inferred about the next and larger sample. Bakker (2004) found such an approach to be helpful in supporting coherent reasoning with key statistical concepts such as data, distribution, variability, tendency, and sampling. He suggests that asking students to make conjectures about possible samples of data pushes them to use conceptual tools to predict the distributions, which helps them develop reasoning about samples. In these processes, "what-if" questions prove to be particularly stimulating. Ben-Zvi (2006) found that the growing samples task design combined with "what-if" questions not only helped students make sense of the data at hand, but also supported their informal inferential reasoning by observing aggregate features of distributions, identifying signals out of noise, accounting for the constraints of their inferences, and providing persuasive data-based arguments (see also Gil & Ben-Zvi, 2011). The growing awareness of students to uncertainty and variation in data enabled students to gain a sense of the middle ground of "knowing something" about the population with some level of uncertainty and helped them develop a language to talk about the grey areas of this middle ground (Ben-Zvi et al., 2012).

Biased sampling techniques used by sixth-grade students were observed in a study by Schwartz, Goldman, Vye, and Barron (1998). These researchers examined and supported fifth- and sixth-grade children's evolving notions of sampling and statistical inference. Only 40 % of the students proposed sampling methods that avoided obvious bias. A follow-up study indicated that fifth and sixth graders overwhelmingly preferred a stratified or stratified-random sample to a biased sample. However, the students remained somewhat skeptical about using truly random sampling methods. For example, to select a sample of school children, roughly 60 % of the students indicated a preference for selecting the first 60 children in a line

over drawing 60 student names from a hat. When they can, students purposefully select individuals to represent the critical population characteristics. Watson and Moritz (2000) reported similar findings with third- and sixth-grade students.

The primary finding of Schwartz et al. (1998) has been that the context of a statistical problem exerts a profound influence on children's assumptions about the purpose and validity of a sample. A random sample in the context of drawing marbles, for example, is considered acceptable, whereas a random sample in the context of an opinion survey is not. It is interesting to note that students appear to accept randomness in games of chance and see merit in stratification techniques when sampling opinions. What they often struggle with is how these two ideas connect. As Schwartz et al. (1998) observe:

> Even though the children could grasp stratifying in a survey setup and randomness in a chance setup, they did not seem to have a grasp of the rationale for taking a random sample in a survey setup. They had not realized that one takes a random sample precisely because one cannot identify and stratify all the population traits that might covary with different opinions. (p. 257)

2.2 Studies involving tertiary students

Confusion about samples and sampling has also been found in tertiary students and even professionals. Tversky and Kahneman (1971) suggested in their seminal article "Belief in the Law of Small Numbers" that:

> People have strong intuitions about random sampling; that these intuitions are wrong in fundamental aspects; that these intuitions are shared by naïve subjects and by trained scientists, and that they are applied with unfortunate consequences in the course of scientific inquiry. (p. 24)

They also claim that:

> People view a sample randomly drawn from a population as highly representative… consequently, they expect any two samples drawn from a particular population to be more similar to one another and to the population than sampling theory predicts, at least for small samples. (p. 24)

Since the publication of this article, many researchers have examined and described the difficulties students have understanding samples, sampling variability, and inevitably, sampling distributions and the Central Limit Theorem (CLT). Well, Pollatsek, and Boyce (1990) noted that people sometimes reason correctly about sample size (e.g., that larger samples better represent populations) and sometimes do not (e.g., thinking that both large and small samples equally represent a population). To reveal the reasons for this discrepancy, they conducted a series of experiments that gave college students questions involving reasoning about samples and sampling variability. The researchers found that students used sample size more wisely when asked certain types of questions (e.g., which sample size is more accurate) than on questions that asked them to pick which sample would produce a value in the tail of the population distribution, indicating that they do not understand the variability of sample means. They also noted that students confused distributions for large and small samples with distributions of averages based on large and small samples. They concluded that students' statistical intuitions are not always incorrect, but may be crude and can be developed into correct conceptions through carefully designed instruction.

Summarizing the research in this area as well as their own experience as statistics teachers and classroom researchers, delMas, Garfield, and Chance (2004) list the following common students' beliefs in reasoning about sampling distributions:

- The sampling distribution should look like the population (for $n>1$).
- Sampling distributions for small and large sample sizes have the same variability.
- Sampling distributions for large samples have more variability.
- A sampling distribution is not a distribution of sample statistics.
- One sample (of real data) is confused with all possible samples (in distribution) or potential samples.
- The Law of Large Numbers (larger samples better represent a population) is confused with the CLT (distributions of means from large samples tend to form a Normal Distribution).
- The mean of a positive skewed distribution will be greater than the mean of the sampling distribution for large samples taken from this population.

In addition, students have been found to believe that a sample is only "good" (e.g., representative) if the sample size represents a large percentage when compared to the population (e.g., Smith, 2004). To confront the common misconceptions that develop and to build sound reasoning about samples and sampling distributions, statistics educators and researchers have turned to visual and interactive technological tools (Chance, Ben-Zvi, Garfield & Medina, 2007) to illustrate the abstract processes involved in repeated sampling from theoretical populations and help students develop statistical reasoning (we address technology separately in the next section).

In a series of studies, Sedlmeier and Gigerenzer (1997) revealed that when students seem to have a good understanding of the effect of sample size, they are thinking of one frequency distribution (of a sample). When they show confusion about sample size, they are struggling with the more difficult notion of sampling distribution (statistics from many samples). Sedlmeier (1999) continued this research, and found that if he converted items that required subjects to consider sampling distributions, to ones that instead required frequency distributions, a higher percentage of correct solutions was obtained.

Building on this work, Saldanha and Thompson (2002) studied high school students' reasoning about samples and sampling distributions. They identified a multiplicative concept of samples that relates the sample to the population as well as to a sampling distribution in a visual way. This interrelated set of images is believed to build a good foundation for statistical inference, which suggests that instructors clearly help students distinguish between three levels of data: the population distribution, the sample distribution, and the sampling distribution. Lane-Getaz (2006) provides such a visual model in her Simulation Process Model, which Garfield and Ben-Zvi (2008, p. 241) adapted and called the *Simulation of Samples* model (SOS). This model distinguishes between the first level of data (population), many random samples from the population (level 2) along with sample statistics for each sample, and the distribution of sample statistics (level 3). In the level 3, a sample outcome can be compared to the distribution of sample statistics to determine if it is a surprising outcome, an informal approach to statistical inference.

2.3 Using technology to help develop reasoning about sampling

Several articles discuss the potential advantage of simulations in providing examples of the process of taking repeated random samples and allowing students to experiment with variables that affect the outcomes (sample size, population parameters, etc.; see Mills, 2002 for a review

of these articles). In particular, technology allows students to be directly involved with the "building up" of the sampling distribution, focusing on the process involved, instead of being presented only the end result.

Numerous instructional computer programs have been developed that focus on the use of simulations and dynamic visualizations to help students develop their understanding of sampling distributions and other statistical concepts (e.g., Aberson, Berger, Healy, Kyle & Romero, 2000). However, research suggests that just showing students demonstrations of simulations using these tools will not necessarily lead to improved understanding or reasoning. Chance, delMas, and Garfield (2004) report the results of a series of studies over a 10-year period that examined various ways of having students interact with the *Sampling SIM* software (delMas, 2001). *Sampling SIM* software allows students to specify different population parameters and generate random samples of simulated data along with many options for displaying and analyzing these samples. They found that it worked better to have students first make a prediction about a sampling distribution from a particular population (e.g., its shape, center, and spread), than to generate the distribution using software, and then to examine the difference between their prediction and the actual data. They then tried different ways to embed this process, either having students work through a detailed activity, or be guided by an instructor. Despite students appearing to be engaged in the activity and realizing the predictable pattern of a normal looking distribution for large samples from a variety of populations, they nonetheless had difficulty applying this knowledge to questions asking them to use the CLT to solve problems. A favorable approach to using the software combines a concrete activity with the use of some Web applets, before moving to the more abstract *Sampling SIM* Software.

Recently, several instructional sequences have been developed that focus on the use of modern simulations and dynamic visualizations (e.g., *TinkerPlots*, Konold & Miller, 2014) to help students develop their understanding of sampling, modeling, and other statistical concepts (e.g., Konold & Kazak, 2008). Despite the development of flexible and visual tools, research suggests that showing students demonstrations of simulations using these tools will not necessarily lead to improved understanding and reasoning (Chance et al., 2007). The current generation of simulation and modeling tools has a potential of promising results in helping students develop their understanding of randomness, sampling, and other key statistical ideas.

3 The special issue

This special issue includes a collection of five articles dealing with various aspects of helping students to reason about data, samples, and sampling, and a reflective discussion. In the first article, "Data seen through different lenses," Konold, Higgins, Russell, and Khalil discuss a major statistical idea, that of "data as aggregate." This notion of an aggregate seems crucial in the step from seeing data as individual data points to seeing a data set as having properties that it may share with a population. More generally, statistical reasoning focuses on properties that belong to the entire aggregate rather than the individual data values (Bakker & Gravemeijer, 2004; Ben-Zvi, 2002, 2004; Ben-Zvi & Arcavi, 2001). The authors analyze students' statements from three different sources to explore possible building blocks of the idea of data as aggregate and suggest ways in which young students develop these ideas. An important reasoning about data framework is suggested that is based on four general perspectives that students use in working with data: regarding data as pointers, as case values, and as classifiers, and the aggregate perspective. The three studies show that some students seem inclined to view data from only one of these three alternative perspectives to the aggregate view. This

inclination influences the types of questions they ask, the data representations they generate or prefer, the interpretations they give to notions such as the average, and the conclusions they draw from the data.

The second article in this collection, "Developing students' reasoning about samples and sampling variability as a path to expert statistical thinking," by Garfield, Le, Zieffler, and Ben-Zvi, describes the importance of developing students' reasoning about samples and sampling variability as a foundation for statistical thinking as defined by Wild and Pfannkuch (1999). The authors study a theory on expert–novice thinking and claim that statistical thinking can be explained via this lens. A case is made that statistical thinking is a type of expert thinking, and as such, research comparing novice and expert thinking can inform the research on developing statistical thinking in students. It is also speculated that developing students' informal inferential reasoning, akin to novice thinking, can help build the foundations of experts' statistical thinking.

Statistical instruction typically pays little attention to the development of students' sampling variability reasoning in relation to statistical inference. Pfannkuch, Arnold, and Wild, the authors of the third article, "What I see is not quite the way it really is: Students' emergent reasoning about sampling variability," therefore designed sampling variability learning experiences for students aged about 15. In their case study, they examine assessment and interview responses from four students to describe their emergent reasoning about sampling variability. Students' statistical reasoning is analyzed using an adaptation of a statistical inference framework and a mental processes framework. The findings suggest that these students are beginning to develop understanding of sampling variability concepts from probabilistic and generalization perspectives and to articulate the evidence used from the data. These students' understanding of sampling variability is explained to be supported by the development of three mental processes: visualization, analysis, and verbal description.

The fourth article, "Proper and paradigmatic metonymy as a lens for characterizing student conceptions of distributions and sampling," by Noll and Hancock, focuses on what can be learnt from students' utterances about their reasoning. Metonymy is the substitution of the name of an attribute or adjunct for that of the thing or idea meant. For example, "the suits on Wall Street" is a metonymy; suits stand in for the people who work on Wall Street or, more generally, financial corporations. This study investigates what students' use of statistical language can tell us about their conceptions of distribution and sampling in relation to informal inference. Trying to address the reported difficulties students have to understand ideas of distribution and sampling as tools for making informal statistical inferences, the authors choose to focus on the ways in which their language mediates their statistical problem-solving activities within the realm of distribution, sampling, and informal inference. To achieve their goal, Noll and Hancock interviewed undergraduate students asking them to focus on (1) distinctions between distributions of populations, samples, and sample statistics; (2) properties of sampling distributions; and (3) how to use sampling distributions to make informal inferences. The analysis focused on students' use of metonymy. The results show two particular metonymies that were used by the students: (1) a "paradigmatic metonymy" in which students applied the properties of the Normal Distribution to all distributions and (2) a "proper metonymy" in which students talked about sampling distributions as compilations of many samples. The impact of these metonymies on students' ability to solve problems and the implications for teaching are discussed.

The final research article in this collection by Meletiou-Mavrotheris and Paparistodemou—"Developing students' reasoning about samples and sampling in the context of informal inferences"—studies the informal inferential reasoning of younger students in the upper elementary school. In a two-phase study, students' initial understandings of samples and

sampling were first examined through an open-ended written assessment and follow-up interviews. Then, a teaching experiment was implemented to support the emergence of children's reasoning about sampling through the provision of an inquiry-based learning environment designed to offer ample opportunities for informal, data-based inferences. The findings indicate that the teaching experiment supported students in moving towards more nuanced forms of reasoning about sampling.

This special issue is concluded with a reflective discussion by three key researchers in the statistics education community. Ainley, Gould, and Pratt discuss the collection of articles drawing on deliberations arising at the Seventh International Collaboration for Research on Statistical Reasoning, Thinking, and Literacy (SRTL7). They choose the perspectives *task design* and the emergence of *big data*, on which they have been working for some time, to offer a commentary from these two perspectives on what might be learnt from the articles in this special issue. In their reflections and reactions, they raise important questions and issues which may carry forward into the future work of the statistics and mathematics education communities of researchers.

4 Teaching students to reason about samples and sampling

One important implication from the collection of articles in this special issue as well as the research literature reviewed above is that it takes time to help students develop the ideas related to sampling, longer than just a few class sessions, which is the amount of time typically allotted in various grade levels (cf. Wild, Pfannkuch, Regan & Horton, 2011). Students need authentic, situated, and rich experiences in taking samples and learning how samples do and do not represent the population prior to formal higher studies of statistics. These experiences may include collecting data through surveys and experiments, where they learn characteristics of good samples and reasons for bad samples (e.g., bias), and creating models using simulation tools (e.g., *Tinkerplots*) to study the relationship between sample and population. These experiences may help students develop a deeper understanding of sampling and (informal) inference, as they repeatedly deal with taking samples, repeated samples, and simulations.

Based on the available research, we think that it is important to introduce ideas of sample and sampling to students early in their statistical learning. By the time students are ready to study the formal ideas of sample, sampling, and sampling distributions, they should have a good understanding of the foundational concepts of sample, variability, distribution, and center (cf. Bakker & Derry, 2011). As students learn methods of exploring and describing data, they should be encouraged to pay attention to ideas of samples and to consider sampling methods (e.g., Where did the data come from? How was the sample obtained? How do different samples vary?). By the time students begin the formal study of sampling variability, they should understand the nature of a random sample and the idea of a sample being representative of a population. They should understand how to choose a good sample and the importance of random sampling.

The study of sampling variability typically focuses on taking repeated samples from a population and comparing a sample statistic, either the sample mean or the sample proportion. Chance and Rossman (2001) recommend that much instruction time should be spent on sampling distributions so that students will be able to use these ideas as a basis for understanding statistical inference.

The research suggests that when students view simulations of data, they may not understand or believe the results, and instead watch the simulations without reasoning about what

the simulation represents. Therefore, many instructors find it effective to first provide students with concrete materials (e.g., counting candies or pennies) before moving to an abstract simulation of that activity (Bakker & Frederickson, 2005).

Many web applets and simulation programs have been developed and are often used to visually illustrate the sampling ideas, sampling distributions and the CLT. Programs such as *Fathom* (Finzer, 2014) can also be used to illustrate the sampling process, allowing for parameters to be varied such as sample size, number of samples, and population shape. Despite the numerous software tools that currently exist to make these difficult concepts more concrete (Biehler, Ben-Zvi, Bakker & Maker, 2013), and despite the innovative learning environments (cf. Garfield & Ben-Zvi, 2009) and curricula that are offered nowadays, there is still not enough research on the ways they can be used to support the emergence of students' statistical reasoning effectively and efficiently and how to assess them. Neither do we know much about teachers' understanding of these issues nor how they can assist their students in developing these ideas. These issues are waiting to be investigated further.

References

Aberson, C. L., Berger, D. E., Healy, M. R., Kyle, D., & Romero, V. L. (2000). Evaluation of an interactive tutorial for teaching the Central Limit Theorem. *Teaching of Psychology, 27*(4), 289–291.

Bakker, A. (2004). *Design research in statistics education*: *On symbolizing and computer tools*. Utrecht, the Netherlands: CD Beta Press.

Bakker, A. (2007). Diagrammatic reasoning and hypostatic abstraction in statistics education. *Semiotica, 164*, 9–29.

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning, 13*(1–2), 5–26.

Bakker, A., & Frederickson, A. (2005). Comparing and distributing growing samples by hand and with a computer tool. In W. J. Masalski (Ed.), *Technology-supported mathematics learning environments* (pp. 75–91). Reston, VA: National Council of Teachers of Mathematics.

Bakker, A., & Gravemeijer, K. P. E. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147–168). Dordrecht: Kluwer Academic Publishers.

Bakker, A., Kent, P., Noss, R., & Hoyles, C. (2009). Alternative representations of statistical measures in computer tools to promote communication between employees in automotive manufacturing. *Technology Innovations in Statistics Education, 3*(2).

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*(1–2), 127–155.

Ben-Zvi, D. (2002). Seventh grade students' sense making of data and data representations. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*. Voorburg, The Netherlands: International Statistical Institute.

Ben-Zvi, D. (2004). Reasoning about data analysis. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 121–145). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman and B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics, 45*(1–3), 35–65.

Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM - The International Journal on Mathematics Education, 44*(7), 913–925.

Ben-Zvi, D., & Garfield, J. (2004a). *The challenge of developing statistical literacy, reasoning, and thinking.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

Ben-Zvi, D., & Garfield, J. (2004b). Research on reasoning about variability. *Statistics Education Research Journal, 3*(2), 4–6.

Ben-Zvi, D., & Garfield, J. (2004c). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 3–15). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Biehler, R., Ben-Zvi, D., Bakker, A., & Maker, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–690). New York: Springer.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education, 1*(1).

Chance, B. L., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 295–323). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Chance, B. L., & Rossman, A. J. (2001). Sequencing topics in introductory statistics: A debate on what to teach when. *The American Statistician, 55*(2), 140–144.

delMas, R. C. (2001). *Sampling SIM* [Computer software]. Available: http://www.tc.umn.edu/~delma001/stat_tools.

delMas, R. C., Garfield, J., & Chance, B. (2004). *Using assessment to study the development of students' reasoning about sampling distributions.* Paper presented at the annual meeting of the American Educational Research Association, San Diego.

Finzer, W. (2014). *Fathom Dynamic Data™ Software* (Version 2.2) [Computer software]. Available: http://concord.org/fathom-dynamic-data-software.

Garfield, J., & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 397–409). Dordrecht, The Netherlands: Kluwer Academic Publishers/Springer.

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal, 4*(1), 92–99.

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75*(3), 372–396.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice.* Dordrecht, The Netherlands: Springer.

Garfield, J., & Ben-Zvi, D. (2009). Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment. *Teaching Statistics, 31*(3), 72–77.

Gil, E., & Ben-Zvi, D. (2011). Explanations and context in the emergence of students' informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1–2), 87–108.

Jacobs, V. R. (1999). How do students think about statistical sampling before instruction? *Mathematics Teaching in the Middle School, 5*(4), 240–246.

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education, 2*(1).

Konold, C., & Miller, C. D. (2014). *Tinker Plots: Dynamic data exploration* (Version 2.2) [Computer software]. Available: http://www.srri.umass.edu/tinkerplots.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259–289.

Lane-Getaz, S. J. (2006). What is statistical thinking, and how is it developed? In G. F. Burrill (Ed.), *Thinking and reasoning about data and chance: Sixty-eighth NCTM Yearbook* (pp. 273–289). Reston, VA: National Council of Teachers of Mathematics.

Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning, 13*(1–2), 1–4.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82–105.

Mills, J. D. (2002, March). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education, 10*(1).

Pfannkuch, M. (2006). *Informal inferential reasoning. Proceedings of the Seventh International Conference on Teaching Statistics.* Voorburg, The Netherlands: International Statistical Institute.

Pfannkuch, M. (2008, July). *Building sampling concepts for statistical inference: A case study.* Paper presented at the International Congress on Mathematical Education, Monterrey, Mexico.

Pfannkuch, M., & Reading, C. (2006). Reasoning about distribution: A complex process. *Statistic Education Research Journal, 5*(2), 4–9.

Pratt, D., & Ainley, J. (2008). Introducing the special issue on informal inference. *Statistics Education Research Journal, 7*(2), 3–4.

Rubin, A., Bruce, B., & Tenney, Y. (1990). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. 1, pp. 314–319). Voorburg, The Netherlands: International Statistical Institute.

Saldanha, L. A., & Thompson, P. W. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*(3), 257–270.

Schwartz, D. L., Goldman, S. R., Vye, N. J., & Barron, B. J. (1998). Aligning everyday and mathematical reasoning: The case of sampling assumptions. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (pp. 233–273). Mahwah, NJ: Lawrence Erlbaum.

Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical implications*. Hillsdale, NJ: Erlbaum.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making, 10*(1), 33–51.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Charlotte, NC: Information Age Publishing.

Smith, M. H. (2004). A sample/population size activity: Is it the sample size or the sample as a fraction of the population that matters? *Journal of Statistics Education, 12*(2).

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105–110. (Reprinted in D. Kahneman, P. Slovic, & A. Tversky (1982) *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.)

Watson, J. M. (2002). Doing research in statistics education: More than just data. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a Statistically Literate Society*. Voorburg, The Netherlands: International Statistical Institute.

Watson, J. M. (2004). Developing reasoning about samples. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 277–294). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education, 31*(1), 44–70.

Well, A. D., Pollatsek, A., & Boyce, S. J. (1990). Understanding the effects of sample size on the variability of the mean. *Journal of Organizational Behavior and Human Decision Processes, 47*(2), 289–312.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–265.

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society, 174*(2), 247–295.

Zieffler, A., Ben-Zvi, D., Chance, B., Garfield, J., & Gould, R. (2015). *Reasoning about uncertainty: Learning and teaching informal inferential reasoning*. Minneapolis: Catalyst Press.