

Learning to reason from samples:

Commentary from the perspectives of task design and the emergence of ‘big data’

Janet Ainley¹, Robert Gould² and Dave Pratt³

1. University of Leicester School of Education, UK
2. Dept. of Statistics, University of California, Los Angeles, USA
3. UCL Institute of Education, University College London, UK

Abstract

This paper is in the form of a reflective discussion of the collection of papers in this Special Issue on *Statistical reasoning: learning to reason from samples* drawing on deliberations arising at the Seventh International Collaboration for Research on Statistical Reasoning, Thinking, and Literacy (SRTL7). It is an important part of the structure of the academic work of SRTL community that at the end of each conference a small group of discussants are given the space to present their reflections and reactions with the aim of raising questions and issues which may carry forward into the future work of the community of researchers. Traditionally they have the freedom to choose the perspectives from which they do this, and this paper has been developed in the same spirit. At SRTL7, the authors of this paper addressed issues on which they have been working for some time, namely Task Design and the emergence of Big Data and are now able to offer a commentary from these two perspectives on what might be learnt from the papers in this special issue.

Corresponding author: Janet Ainley, janet.ainley@le.ac.uk, +44(0)116 252 3690

Key words: task design, big data, samples,

Learning to reason from samples:

Commentary from the perspectives of task design and the emergence of ‘big data’

Janet Ainley¹, Robert Gould² and Dave Pratt³

1. University of Leicester School of Education, UK
2. Dept. of Statistics, University of California, Los Angeles, USA
3. UCL Institute of Education, University College London, UK

Abstract

This paper is in the form of a reflective discussion of the collection of papers in this Special Issue on *Statistical reasoning: learning to reason from samples* drawing on deliberations arising at the Seventh International Collaboration for Research on Statistical Reasoning, Thinking, and Literacy (SRTL7). It is an important part of the structure of the academic work of SRTL community that at the end of each conference a small group of discussants are given the space to present their reflections and reactions with the aim of raising questions and issues which may carry forward into the future work of the community of researchers. Traditionally they have the freedom to choose the perspectives from which they do this, and this paper has been developed in the same spirit. At SRTL7, the authors of this paper addressed issues on which they have been working for some time, namely Task Design and the emergence of Big Data and are now able to offer a commentary from these two perspectives on what might be learnt from the papers in this special issue.

Task Design: What do we learn about the nature of the development of statistical reasoning from an analysis of the design and use of tasks in these papers?

Our perspective on task design is framed by the related notions of *utility* and *purpose* (Ainley, Pratt, & Hansen et al., 2006). We argue that an important, but often overlooked, dimension of learning about a mathematical or statistical idea is understanding what it is useful for and

what power it offers in addressing problems. We call this the *utility* of the idea. One question we would wish to ask in analysing any task design is *does this task give learners opportunities to appreciate the utility of the ideas they are learning?* We further argue that to provide such opportunities tasks should have a clear *purpose* for the learner. *Purpose* here does not refer to a real-world application, but to an engaging challenge for the learner within the classroom context. A second question that we wish to ask is therefore *is this task purposeful for learners?*

In the context of the topic for this special issue, learning to reason from samples, this leads us to raise a question which we suspect is rarely voiced, but is potentially puzzling for learners, “If you want to know about the population, why look at a sample?” In other words, what are the *utilities* of sampling; why and how sampling is useful?

In fact, we are aware that there can be confusion amongst children and sometimes teachers about the role of sampling, which demarcates descriptive and inferential statistics. For example, Pratt et al (2008) have referred to the distinction between Game 1, in which the population is fully known, such as in a census, and Game 2, where the population is only partially known and what is available is a sample of data. Whereas in Game 1 it is sufficient to describe the population, in Game 2 there is a need to infer from the sample conclusions about the population with a degree of uncertainty about the correctness of those conclusions. Teachers might not always make this distinction clear to learners, who in consequence might think they are playing Game 1 when in fact the teacher’s agenda relates to Game 2. Of course, teachers of statistics are drawn not only from mathematics but from many disciplines such as science, social science and the humanities. In such circumstances, it would not be surprising if the distinction were also unclear to the teacher, and this is likely to be reflected in the design, choice or implementation of tasks.

In Game 1, the descriptive game, there is no utility for sampling. But, in Game 2, the inferential game, we are able to categorise distinctive sampling situations:

A. Situations in which the population is material, finite and countable:

- i) Typical ‘market research’ situations, where the population is finite but large and not easy to access so only countable in the imagination;
- ii) Typical ‘sweets in a bowl’ situations, where the population is finite, but contained (and so could be counted in practice).

B. Situations in which the population is a mathematical formulation:

- i) Throwing a die or tossing a coin, where the ‘population’ is a probability distribution containing a set of possible outcomes and an associated probability; here the population does not have a material existence (unless transformed into the Aii situation, as say when drawing balls from an urn);
- ii) A scientific experiment, where the resulting measurements might be imagined prior to the experiment as a sample from an infinitely-densely packed continuum in the form of a probability density function.

From these four categories we might expect interestingly different kinds of answers to the question *why look at a sample?*

- For Ai), looking at the whole population is not possible, and so looking at a sample is the only access we have: we can imagine that sampling in this case may seem self-evidently purposeful and necessary, and its utility clear, even to a relatively naïve learner.
- For Aii), sampling seems less purposeful: we could empty out the whole bowl and separate out the different colours. Sampling here might seem artificial, and its utility is not clear. However, as a pedagogic task the fact that the whole population can be accessed provides opportunities to highlight features of the sampling process.

- For Bi), theoretically we know about the population, and so there appears to be no purpose for sampling. Sampling here may also have a pedagogic purpose, but in order to give the task purpose for learners, and allow them to appreciate the utility of sampling, some additional factor must be introduced, such as conjecturing a biased die.
- For Bii), the population is genuinely not known, but the situation might not be seen as essentially statistical. What is the purpose of taking more than one measurement under the same conditions? We know that in many UK science classrooms ‘do it three times and take an average’ is a mantra for teachers; a pragmatic response to the ‘problem’ of variability. However, the purpose for sampling that this represents might be no more than an automatic response that is not founded upon a true sense of the utility for sampling.

From this categorisation, we make two important and connected observations. The first is that the two categories which appear to provide opportunities to appreciate the utility of sampling are Ai) and Bii), whilst those most likely to be used pedagogically to teach about sampling are Aii) and Bi).

Our second observation is that whilst the statistician may see all four of these categories as instances of sampling, and indeed see them as equivalent to each other, this similarity may not be apparent to learners. Arguably, professional statisticians have become so enculturated into sampling that they are able to focus exclusively on sampling as an entity. However, it is reasonable to suppose that learners’ knowledge is more likely to be situated in the reasons for sampling activity. We make this assertion from the perspective of Constructionism (Harel & Papert, 1991) and Inferentialism (Brandom, 2002, Bakker & Derry, 2011). The former would advocate that tasks encourage using statistics in order to come to know statistics so that students are from the start engaged with the reasons for the activity. The latter would position knowledge in the inferential actions of the students as they give and

ask reasons, thus prioritising the inferential role over that of representations. We argue that learners might in fact identify quite different purposes for the activity in the above types of task and therefore foreground differences that statisticians have, in a sense, forgotten or at least learned to ignore.

We notice in our reading of the papers in this volume that where tasks are described in detail, the majority can be described in our categorisation as A_i , where the context is some form of sampling from a large population. In other cases, for example in Pfannkuch, Wild and Arnold's paper in which the focus is on comparing pairs of data sets, we assume that the data has arisen from A_i type situations. In other cases, where the design of the whole tasks is indicated, we notice that a variety of situations, from the perspective of the above analysis, is used, which could be potentially confusing for students.

Meletiou-Mavrotheris and Paparistodemou report on the use of a sequence of tasks, based on a hypothetical learning trajectory, in which the emphasis is clearly on pupils learning about 'the purpose¹ and usefulness of sampling' (this Special Issue). The task sequence is composed of a mixture of type A_i and type A_{ii} tasks, and also a coin tossing activity (B_i) which is introduced to highlight ideas about random samples. Interestingly, despite the emphasis on sampling, when the children were given the freedom to design their own investigation, some chose to investigate an issue within their own school and, quite reasonably, thought that the best approach was to survey the whole population of all the children.

There is clearly a pedagogic dilemma here. The researchers' intention in introducing an open-ended opportunity for the children to design and carry out their own study was to engage them in the experience of genuine data collection and analysis, which would 'support students' emerging views of samples and sampling' (this Special Issue). Focusing on a question of real interest to them, the children then designed a study which could be carried

¹ 'Purpose' is used here in a different sense to the one we explain above. At other time they also refer to the 'utility' of sampling, but it is unclear whether this usage of the word matches our own.

out without sampling, and the teacher had to impose artificial restrictions on their activity in order to retrieve the focus on samples. Meletiou-Mavrotheris and Paparistodemou recognise this dilemma and address it by including a comment on the need to scaffold children's choices towards studies involving larger populations in the revised learning trajectory. Whilst this offers one way to ensure that pedagogic opportunities to focus on sampling are not lost, we suggest an alternative perspective. Part of what is involved in understanding the *utility* of sampling is recognising situations in which it may not be the most appropriate way to collect data if you really want to know about a population. Structuring all experiences of data-based inquiry to include the use of samples may therefore be counter-productive. The careful design of tasks might therefore support students in recognising the utility of sampling, or indeed to recognise situations in which sampling has no utility.

Garfield, Le, Zieffler and Ben-Zvi consider task design from the perspective of the novice and the expert on sampling. Their reading of the novice-expert literature leads the authors to the conclusion that 'learning environments that are designed according to the expert–novice theories should foster an atmosphere in which it is safe for learners to make mistakes and express their partial understandings' (this Special Issue). Without doubt, this feature is one of the foundational ideas in microworld design from the constructionist perspective as advocated by Harel and Papert (1991). A task that is seen as purposeful by students is likely to allow space for the student to make decisions in order to take some ownership over the activity. When the student has some control over their activity, they make mistakes but, in the sort of settings envisaged by the authors and by constructionists, mistakes become positive steps towards more sophisticated understandings. Such mistakes enhance rather than detract from the student's sense of purpose.

Later, the authors argue that 'having a model of expert thinking can serve as an ideal that guides the planning of a learning trajectory or the use of a particular a tool' (this Special Issue). Part of the design challenge that leads to a purposeful task is to connect that design to utilities that might be constructed by students. This is far from trivial but, in considering

utility, a starting point is the epistemological analysis of the knowledge domain, which requires a model of expert thinking. One of the reasons that it is non-trivial for the designer to connect utility to purpose is that some components of the expert model are more utility-oriented than others and so a careful selection has to be made from the array of possibilities. Furthermore, the designer will often need to invent a novel representation of the chosen component from the expert's model. The novelty in the representation might lie in how it differs from the conventional representation in order to make the underlying idea more accessible within the likely learning trajectory that the student follows in pursuing their purpose. The authors noted Konold's design principles (2002) for TinkerPlots™ (Konold, 2011), whereby the teaching of data analysis should be structured to fit the students' development rather than targeting specific graphical representations. One of the reasons that TinkerPlots is so effective is that it incorporates novel representations that are more accessible by relatively naïve students than would be more conventional representations.

In this issue, Konold, working with Higgins, Russell and Khalil, offers a glimpse of how deep analysis of children's thinking can inform the creation of tools. They note how children seem to hold different perspectives when looking at the presentation of data. They have observed what they call pointer, case value, classifier and aggregate perspectives to capture how children seem to focus respectively on, for example, the activity as a whole, a specific piece of data, a frequency total and the shape or position of the distribution. The design of a task, including the specific question asked, might draw children towards one or other perspective. Although the authors do not want to place these perspectives in a hierarchy of levels, they also note that some perspectives are often apparent amongst younger children and that the aggregate perspective seems sometimes inaccessible even when the task might appear (to the designer) to require it. This suggests both the complexity of the challenges of task design and some limitation in the power of task design to challenge children's intuitive ways of thinking. They also argue that it would be a mistake to move children too quickly towards the aggregate view, since other perspectives are "well suited to the interests of many

young students” (this Special Issue) and keep them connected to the meanings of the data. Through this sentence, the authors make explicit how consideration of children’s interest and purpose is fundamental for them.

The difficulty of the challenge in designing for both purpose and utility is further demonstrated in this issue in the writing of Pfannkuch, Arnold and Wild. They raise an interesting dilemma. There are without doubt key statistical concepts that can be derived from the model of expert thinking. There are also procedures through which those concepts are deployed, and which therefore must also be regarded as important. This paper discusses how it might be possible to teach students relatively informal approaches to statistical decision-making, what they refer to as ‘how to make the call.’ They do not give detail of the context in which specific tasks are set, but it appears that they relate to situations which we would categorise as Ai (‘market research’), and in which two data sets are being compared. Along the way, the students are introduced, again informally, to visual images of sampling and sample statistics. To teach these concepts and procedures, even informally, the learning trajectory involved exploring on-screen behaviour that was not contextualised in the world of the student but in the teacher’s agenda. In other words, a pedagogic decision has to be made to focus tightly on the specific statistical concept and rely on the student’s willingness to engage with the teacher’s agenda, even though the purpose and utility may be obscure for the student.

That demand on the student appears to be exacerbated when their approach sets out rules about ‘how to make the call’ in the decision with little justification that is accessible to the students, certainly not at the theoretical level. We do not make these points to criticise the authors’ approach, which is surely imaginative and holds great promise. Indeed, teachers the world over engage students in activity around specific concepts in ways that require the students to have faith that the learning will one day be helpful to them. Rather, we wish to make the point that, despite the importance of connecting purpose and utility in task and tool design, there does appear to be some limitation in their scope. Nevertheless, recognition that

consideration of purpose and utility are being ignored at any particular point in a learning trajectory might help to alert the teacher to the pedagogic risks of disengagement by the student.

Big Data: What do these papers tell us about preparing students to live in a world of Big Data?

The term “Big Data” means different things to different people. For the purposes of this discussion, we prefer to adopt a definition by Lane et al (2014), who use Big Data as a “paradigm” that represents a culture in which data play an historically large role. In other words, data are not “big” because of the size of the datafile, but because they belong to a new class of data that differ in structure and source from “traditional” data that have inspired institutional changes in how we learn from data. This culture has produced a class of data that are not necessarily covered in aforementioned types A (sample data that emerge from situations in which the population is material, finite and countable) and B (sample data that emerge from situations in which the population is a mathematical formulation) above. As opposed to A and B data, which are typically collected by humans with a specific purpose in mind, at least some types of Big Data are collected algorithmically. A sensor, for example, may automatically take readings every millisecond. An app running on a personal computer might record every song that is played through the computer’s music system. A camera chip may store meta-data on every picture taken. In addition, many civic organizations, such as police departments, routinely collect data on every event; the event of an arrest triggers a data collection protocol.

What is interesting about Big Data from an educational perspective is not (just) the size of the data collected, but its ubiquity. Most of the data discussed in these papers could be described as “professionals’ data.” These are data that professionals, market researchers, scientists, administrators, politicians, routinely rely upon. Many Big Data, on the other hand, directly impact students’ lives. Big Data can be, therefore, “students’ data.” Students’ data is

potentially a large and diverse category of data-types, and might include data accumulated on smartphones, data collected by the students using smartphones, data presented in html format on websites, data from personal health sensors, government web sites, etc. The point is that data are not necessarily objects encountered by students in the formal context of the classroom, but are objects they see every day (Gould, 2010).

How, then, do we prepare students to understand these data? Because data description is an important component of reasoning in the context of Big Data, Konold, Higgins, Russell and Khalil offer some valuable insights. An important question with many Big Data is “What has happened?” – a question requiring students to find and describe trends in data that might very well represent a population. Although the context of Konold et al is understanding how students interpret univariate distributions, their finding that students’ views generally fall into four perspectives, not all of which are useful for identifying trends and patterns, remind us that students need careful scaffolding to interpret data visualizations. Many Big Data are now accessed and understood through sometimes complex data visualization portals, for example, Google Maps, or the New York Times data visualization pages², and so computer scientists and statisticians who design these visualizations should include aids to assist in aggregation (e.g., Ridgway & Smith, 2013).

Garfield, Le, Zieffler, and Ben-Zvi raise many interesting issues, particularly about the importance of understanding how experts engage with data. A complete portrait of statistical thinking should include experts’ work with Big Data, particularly since a complete understanding of how this might differ from more traditional views is not known. The Garfield et al.’s curriculum includes early experience with designing simulations, and we conjecture that this experience would lead to deeper computational thinking skills, which have been identified by some as an important component of modern statistical thinking (Nolan & Temple Lang, 2010; Undergraduate Guidelines Workgroup, 2014). Although their

² <http://www.nytimes.com/newsgraphics/2013/12/30/year-in-interactive-storytelling/#dataviz>

discussion is software-agnostic, we are reminded that little is known about how experts reason with statistical software, and whether statistical reasoning skills can be improved through a comparison of how students work within a software environment.

The remaining three papers focus almost exclusively on sampling and, from our perspective, are more likely to concern professionals' data than students' data. Even so, the concepts of sample representativeness, the relation of samples to population, sampling variability, and bias are essential for students to understand the limits of algorithmically gathered data. The issue of representativeness is particularly interesting. Data gathered by a personal health sensor (such as the FitBit, the Nike+ FuelBand, or the new Apple Watch) might be perfectly representative of the owner's activity for the last few days. Or it might have left out important activities if, for instance, the owner left his Fitbit on the table all day. However, it is not representative of other students' activities. Even the collection of all data from all students in a classroom is not representative of all students in all classrooms. A random sample, however, is representative of its population by definition (with some caveats), although the size of the sample might mean that sampling variability is too high to make a useful inferential claim. The prevalence of sensor data means that students are faced with additional challenges to their understanding of how and when we can make inferences to a larger population.

We suspect that part of the appeal of studying how students reason from samples is that this is an inherently mathematical topic, and most statistics education is developed within mathematics' curricula. However, the intersection between Statistics and Mathematics, while rich and important, is shrinking as scientists learn to make sense of the growing class of data collected from non-random samples. One might be tempted to write off this struggle as one belonging to specialists and far beyond the concerns of the classroom were it not for the fact that these non-random samples produce data (and statistics) that students see every day of their lives. We strongly encourage the mathematics education research community to turn

their attention to better understand how students think and reason with algorithmically collected data.

Epilogue

In the introduction to this special edition, the Editors draw attention to the complexities of learning to reason about samples and sampling, and the need for both extended time and rich experiences to allow students to develop secure understandings of what may otherwise appear to be counter-intuitive statistical ideas and procedures. In our reflections on these five papers, we see many important contributions to the development of effective teaching in this area, as we have highlighted in the discussions above. In conclusion we draw attention to three inter-related issues.

The need for thoughtful task design which not only emphasizes the need for a range of experiences to build conceptual understanding, including exploiting technology to support new visualizations, but also ensures that learning is located in meaningful contexts, in which students can experience the power and usefulness of sampling.

The need to listen carefully to students' voices as their reasoning develops, so that teaching, task design and developments in technology are informed by the perceptions of novices as well as those of experts.

The need to locate teaching and learning about reasoning about samples and sampling in the changing social context, so that what students are learning today will equip them to function effectively in the future.

References

- Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal*, 32(1), 23-38.
- Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, 13(1-2), 5-26.

- Brandom, R. (2002). Overcoming a dualism of concepts and causes: A unifying thread in empiricism and the philosophy of mind. In R. Gale (Ed.), *The Blackwell guide to metaphysics*, (pp. 263-281). Hoboken, NJ: Blackwell Publishers Ltd.
- Gould, R., (2010). Statistics and the modern student. *International Statistical Review*, 78(2), 397-315.
- Harel, I., & Papert, S. (1991). *Constructionism*. Norwood, New Jersey: Ablex.
- Konold, C. (2002). Teaching concepts rather than conventions. *New England Mathematics Journal*, 34(2), 69–81.
- Konold, C. (2011). *TinkerPlots™: Dynamic data exploration*. Emeryville, CA: Key Curriculum Press.
- Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (2014). Conceptual framework. In J. Lane, V. Stodden, S. Bender and H. Nissenbaum (Eds.), *Privacy, big data, and the public good: Frameworks for engagement*, (pp. 1-3). New York, NY: Cambridge University Press.
- Nolan, D., & Temple Lang, D. (2010). Computing in the statistics curriculum. *The American Statistician*, 64(2), 97-107.
- Pratt, D., Johnston-Wilder, P., Ainley, J., & Mason, J. (2008). Local and global thinking in statistical inference. *Statistics Education Research Journal*, 7(2), 107-129.
- Ridgway, J., & Smith, A., (2013). Open data, official statistics and statistics education: Threats, and opportunities. Keynote talk: *Proceedings of the First Joint International Association for Statistics Education and the Association for Official Statistics, Statistics Education for Progress*, Macau. Retrieved from <https://www.dur.ac.uk/smart.centre/publications/>
- Undergraduate Guidelines Workgroup (2014). *American Statistical Association Draft Guidelines for Undergraduate Programs in Statistical Science*. October 1, 2014.

