# STUDENTS' EMERGING INFERENTIAL REASONING ABOUT SAMPLES AND SAMPLING

THEODOSIA PRODROMOU

The University of New England

tprodrom@une.edu.au

This paper investigates students' emerging inferential reasoning about samples and sampling through observation of 13- to 14-year-olds, challenged to infer aspects of an unknown population in an inquiry–based environment. This paper reports on how students working with *TinkerPlots* focus on changing aspects of the samples as the sample size grew larger. Students made connections to key statistical concepts during the process of growing samples and quantified the level of confidence about their informal statistical inferences. They generally recognized the relationship between the sample size and the confidence interval for a given confidence level.

## Introduction

Over the past decade there has been an increasingly strong call for statistics education to focus more on statistical literacy, reasoning, and thinking. The Australian Curriculum and Reporting Authority [ACARA] (2010) advocates the broadening of probability and statistics in the school curriculum. "Statistics and probability initially develop in parallel and curriculum then progressively builds the links between them" (p. 2). In particular, these two topics are connected in the study of inferential statistics, in which one makes inferences that are based on data and qualified using probability. The curriculum anticipates that "students recognise and analyse data and draw inferences. They develop an increasingly sophisticated ability to critically evaluate chance and data concepts and make reasoned judgements and decisions, as well as building skills to critically evaluate statistical information and develop intuitions about data" (p. 2).

Related to this, by the end of primary school students are expected to "develop an understanding of sampling" (p. 32) and "consider the need of sampling and recognizing when a census of an entire population is not possible or necessary, and identifying examples of sampling in the media" (p. 33). Year 10 students are expected to be able to "evaluate the appropriateness of sampling methods and sample size in reports where statements about a population are based on a sample" (p. 48).

Developing a sophisticated reasoning about sample data, and sampling "may be associated with developing literacy and social reasoning skills rather than developing numeracy skills" (Watson, 2004, p. 279) because the target reasoning is the cornerstone of drawing conclusions about populations in our society. Such reasoning is embedded in decision-making under uncertainties in different contexts and fields.

## Literature review

Selecting samples of data and using samples to draw inferences about unknown populations lie at the heart of statistics. The concepts of "sample" and "sampling" are structurally complex and require the coordination of several concepts including graph interpretation, spread, distribution, randomness, and likelihood (Ben-Zvi, Makar, Bakker, & Aridor, 2011).

Although, literature is replete with research on college students' conceptions of sample size and representativeness (Tversky & Kahneman, 1971), limited research was undertaken until recently on school students' conceptions of samples and sampling.

Research on students' conceptions of sampling by Watson and Moritz (2000a, 2000b), has shown that children as young as 8- and 9-years-olds have relatively naïve conceptions about samples. According to Watson and Moritz, the children of this study were typically comfortable drawing conclusions about a population based on small samples without recognizing any potential problems of bias. Early middle school students (age 13–14) understood the concept of samples in real world situations, but they had difficulties making the transition to the formal statistical meaning and using appropriate associated terminology. Watson and Moritz (2000b) showed that older students (age 14–15) were concerned about potential errors arising from small samples. The observations in the research study of Watson and Moritz show the importance of making explicit the differences between taking a small sample from a homogeneous entity (for example, a small sample of blood) to make generalisations about the larger entity from which it was drawn, with taking a sample from a heterogeneous population that has much variability (for example, a sample from a population of students) to estimate a specific characteristic such as weight. The ideas inherent in sampling from homogenous entities do not generalize to the notion of sampling variation and the need for large samples in making inferences from data. Watson and Moritz have also emphasized the importance of the notions of variation and representativeness when students engaged in a sampling related task.

Watson (2004), who summarizes outcomes of research on reasoning about sampling, notes that students often pay attention to fairness and distrust random sampling methods as a process producing unbiased samples. According to Watson, students prefer biased sampling methods, such as voluntary samples. Other researchers have documented that students and teachers often have difficulties in distinguishing samples from populations when working with data (Pratt, Johnston-Wilder, Ainley, & Mason, 2008; Pfannkuch, 2008). In response, there has been a recent research effort to understand how better to approach the topic from a pedagogic perspective. One response has been *informal statistical inference*, characterised as a process of drawing generalised conclusions expressed with uncertainty from data, which extend beyond the data collected (Makar & Rubin, 2009). Two international research forums on statistical reasoning, thinking and literacy (SRTL-5 and SRTL-6) have been dedicated to the study of how students might make sense of informal inferential processes and reason about inference related tasks. In particular, the definition of *informal inferential reasoning* provided for SRTL-6 in 2008 was "the cognitive activities involved in drawing conclusions with some degree of uncertainty that go beyond the data and having empirical evidence for them". Three fundamental principles of informal inference were provided: generalising beyond data, using data as evidence of generalisations, and expressing the degree of certainty (due to variability) for the generalisation. The main types of generalisations indicated were predictions, parameter estimates, and conclusions. Making such inferences informally, gives students the sense of the power of statistical techniques in making reasoned judgements and decisions about data from real-world contexts.

Another response has been *Growing Samples*—an instructional idea suggested by Konold and Pollatsek (2002), but then developed by Bakker (2004) and used by Ben-Zvi et al. (2011). Bakker helped eighth grade students who engaged with a sequence "of growing samples" activities to see stable patterns generated by larger samples, thus students understood that larger samples are less variable and better represent a population. Bakker suggested that asking students to make conjectures about the growing samples build students' reasoning about sampling in the context of variability and distribution. Such an approach is helpful in supporting coherent reasoning, based exclusively on the integration of key statistical concepts such as sampling, data, distribution, variability, and tendency. Ben-Zvi et al. used data from a design experiment in Israeli Grade 5 classrooms to show how 11 year-olds develop inferential reasoning about sampling while working with *TinkerPlots*. This research was in line with the literature of growing samples beginning from a sample of size eight from their class (including themselves), and moving to a bigger sample (a whole class) and then to the whole grade in the school. The students not only experienced the limitations of small samples when making inferences about a larger population, but also an emerging quantification of confidence in making such inferences, interconnections of concepts of sampling, and informal statistical inference with key concepts such as spread, distribution, likelihood, randomness, average, and graph interpretation

In this paper, the focus on informal statistical inference and children's reasoning about sampling, emerges out of aspects of the work of Ben-Zvi et. al (2011). Two of the questions for future research as suggested by Ben-Zvi, et. al guide this research study. First, this research investigates how ideas about sampling in relation to informal statistical inference can be further developed in the next stage. It is expected through asking this question that some insights might be gained into the conceptual struggle that needs to take place for 13- to 14-year-olds to engage in inferential reasoning about samples and sampling. In doing so, a constructivist stance is used to search for naïve conceptions that might serve as resources in deploying more sophisticated strategies. Second, this might shed some light on how the instructional idea of growing samples can be further improved and used.

## Method

This research study falls into the category of design experiments (Cobb, Confrey, diSessa, Lehrer, & Shauble, 2003). Typically, design experiments require several iterations. This article reports on a pilot study that examined students' exploration of a dataset using *TinkerPlots* (Konold & Miller, 2005).

The learning sequence was built around two sessions of extended data investigations of a student-administered survey from Years 7−9 in the previous school where the researcher taught. The survey gathered information about students' weight and weight of students' backpacks[1]. Afterwards, the weight of a student's backpack was divided by the student's weight, and the calculated percentages were compared with the doctor's recommendations.

Students used *TinkerPlots* to analyse the data collected from the student-administered survey for approximately two hours. Sliders and filters which control the increase and decrease of the sample size and formula-defined attributes were implemented in order to allow the students to have more control over the sample size that they select from the dataset. During the activities, students observed how the animated plots they were studying varied when the sample size slider was used to add cases to the graph.

The design of activities evolved around the idea of growing samples, starting from a sample of size 10, moving to about 30, then 100, and finally the entire population. Using a sequence of "growing sample" activities was a pedagogical design conjecture to help students understand that larger samples better represent the population, progressively developing their inferential reasoning about samples, and a level of confidence students place in their inference.

The researcher conducted clinical interviews with small groups 13- to 14-year-old students, in Year 8 of an Australian secondary school. The researcher worked with students from one class, covering a range of attainment. Students worked in pairs.

While the students were working, Camtasia software was used to video record the computer screen output and audio record the students' voices. The data collected were analysed using progressive focusing (Robson, 1993). At the first stage, the audio recordings were simply transcribed and screenshots were incorporated as necessary to make sense of the transcription. Subsequently, the transcript was turned into a plain account with no explicit interpretation other than through selection of the most promising sections. The less interesting sections were replaced with discursive descriptions of what happened. At the third stage, an interpretative account was written. Episodes were selected to illustrate students' evolving informal reasoning when making inferences about samples and sampling.

The findings are presented below through the case of Rafael (Ra) and Gina (G). Analysis of the data from other students is ongoing.

## Results

### Stage 1: First investigation with 10 data points

Rafael and Gina expressed dissatisfaction with working from only ten data points (Figure 1) and formed an initial reaction:

---

[1] The activity was inspired by a report written by students at Hermantown academy, available at www.ga.k12.pa.us/Academics/LS/5$^{TH}$/Backpck/Index.htm

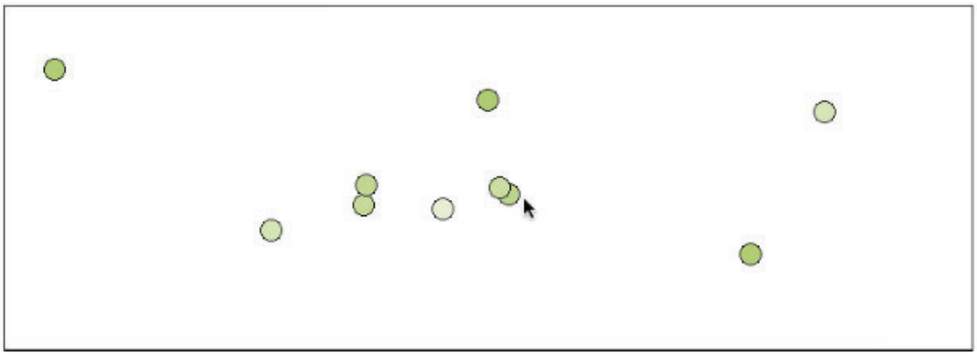

*Figure 1. Distribution of 10 data points.*

1. G: I don't think we can draw any conclusions about this data because we do not have enough data.
2. Ra: The data we have are too spread out to make inferences about all the students' backpacks.

The students then began to look more closely at the distribution of the data trying to identify an apparent pattern. Rafael characterised the data as too spread out to make inferences about all the students' backpacks (line 2).
Rafael and Gina organised the ten data points to spread across eight categories of backpacks weights, with most categories having zero to three points (Figure 2).

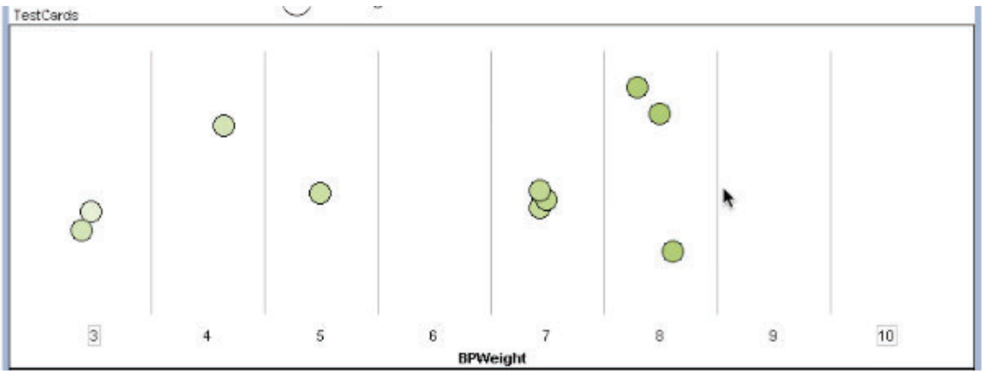


*Figure 2. 10 data points spread across 8 categories.*

When exploring more carefully the weights of backpacks:

3. Ra: I can see just two packs weighing 3 lbs, one 4 lbs, one 5 lbs, three 7 lbs, three 8 lbs… Most of the packs are 7-8 lbs.
4. Re: Do you think if we weight all the backpacks from all the students from Years 7–9, we will be able to draw this conclusion?
5. Ra: I do not think so. Maybe.
6. Re: Can we talk about all the backpacks from looking only at this data?
7. G: No, there are not enough backpacks to say this represents the entire school.
   I want to see the weights of more backpacks.

Rafael seemed to be able to draw conclusions from numerical data (line 3). The relatively high frequency of backpacks weighing seven to eight lbs attracted Rafael's attention but he seemed to be uncertain whether he could base any inferences about the weights of all the backpacks of all the students from Years 7–9 upon the current sample of ten (lines 2, 5). Similarly, Gina appeared to be reluctant to draw any conclusions about the population from this sample size (line 7), so she suggested investigating the weights of more backpacks.

## Stage 2: Second investigation with 30 data points

Students were given data for a class from Year 7 (30 students). Gina and Rafael's immediate reaction was to engage enthusiastically with the investigation of the data points (Figure 3).

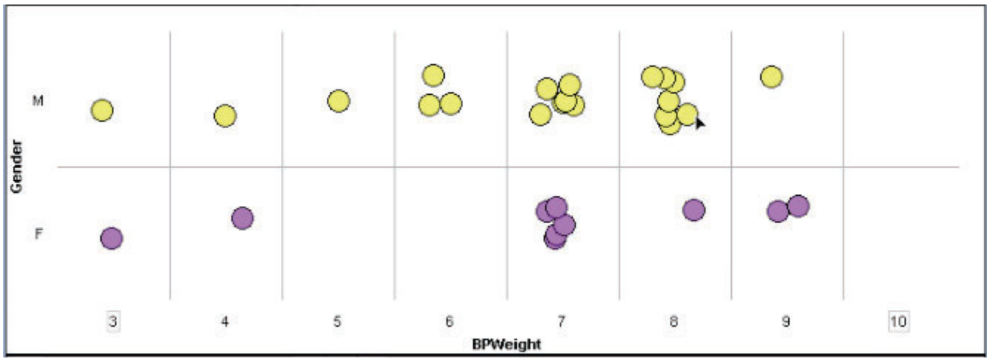


*Figure 3. Distribution of 30 data points.*

| | | |
|---|---|---|
| 8. | Ra: | My initial observation still seems true. Most of the packs are heavy. There are more packs weighing seven to eight lbs then there are three-six lbs. I see that most of the boys carry packs that weigh seven to eight lbs. |
| 9. | G: | This is interesting. We can see clearly that most of the packs are still on the heavy side. We only have one boy that his backpack weighed nine lbs. But we have two girls that carry pack weighing nine lbs… |

Rafael's immediate reaction was one of "surprise" when he realised that the prediction he made earlier for the sample of ten hold true more broadly (line 8). A recurring feature of students' investigation was their focus on the changes or similarities that occurred in the appearance of the distribution of data as they compared the new data (sample size 30) with those from the previous investigation (sample size 10).

On the one hand, Rafael did not seem to experience any kind of conflict when drawing conclusions from a small sample. The researcher wished Rafael could see that the small sample size was a flaw in the validity of his inference due to the vagaries of sample variability. On the other hand, Gina found it "interesting" because she did not anticipate that her conclusions would be similar to those she made for the sample of ten. Even though Gina recognised the unexpected similarities (line 9), she did not seem to understand the reasons underlying them.

| | | |
|---|---|---|
| 10. | R: | Do you think that in general the boys carry heavier packs which weigh 7 to 8 lbs more than girls? |
| 11. | Ra: | I guess so. |
| 12. | G: | I cannot tell. The backpacks of 30 students cannot represent all the backpacks of students, but we can better draw conclusions now about bigger samples than when we were given the data for a sample of 10. |

Rafael seemed to be more certain than Gina (line 11). On the contrary Gina expressed her lack of confidence in drawing conclusions from only 30 data points although she recognised that the increase of the sample size gave a better basis for their inferences (line 12). When asked to informally quantify their level of confidence about their conclusions from the current sample of thirty:

13. R: Can you give me an interval from 1 to 10 how certain you are the distribution of the data points will remain the same if we carry out the investigation with more data?
14. Ra: 7 to 9.
15. G: This is too high. I'm giving 5 to 6.

Rafael's confidence in his inferences was stronger than Gina. Gina's expressed a lack of confidence in making inferences about the population because she was expecting to observe even small changes on the distribution of data.

## Stage 3: Third investigation with 100 data points

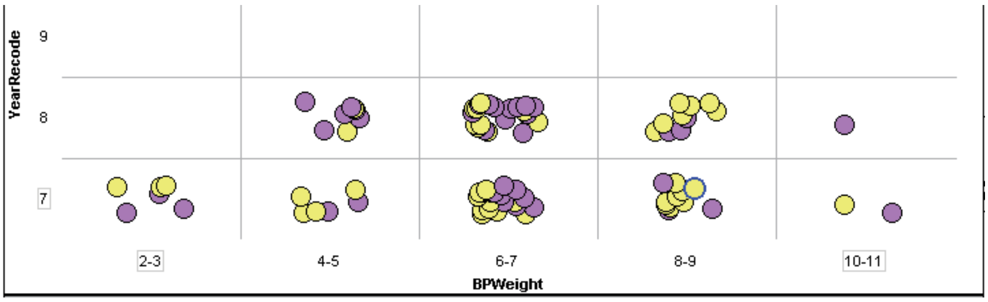


*Figure 4. Distribution of 100 data points.*

In the third investigation, students were given data for the entire Year 7 and Year 8.

16. Ra: The data points are spread out from four to ten lbs, the data points for Year 7 are closer together, between 4–11 lbs and Year 8 is spread out evenly from 8–14 lbs
17. R: So, do you think these results shows what is happening in the whole school?
18. Ra: Yes. I am giving an 8–9 interval of how certain I am.
19. R: Why not 9-10?
20. Ra: I can not be so sure. You see … The more backpacks we weigh, there will be less room for mistakes or unknown backpacks weights.
21. G: I'm not quite sure too, it is better than what it was before, but I'm not sure if I can talk about all the students in the school. The bigger the sample is, the better the results are.

The students engaged in interpreting the graph (line 16) and tried to focus on even small changes. Of course the process of growing samples produced consecutive images of the distribution of data points that allowed Rafael and Gina to gain some sense of the sample/sample size and population relationships (lines 18-21). Rafael articulated that "the more backpacks we weigh, there will be less room for mistakes or unknown backpacks weights". He seemed to acknowledge the importance of large samples and the uncertainty caused by unexplained variation in the weight of backpacks (line 20).

However, there is perhaps another way to think about what the students were expecting to experience: that is the idea of an ideal sample that can perfectly represent the parent population. This might have proved very convincing, regardless of the size of the sample. Gina suggested:

22. G: Why don't we try something else?

23. R: What do you want to try?
24. G: I want to see how many students' backpacks we need to weigh in order to decide the size of the sample upon which we can draw valid conclusions.
25. R: What do you want to try?
26. G: I want to gradually move the slider of the sample size from 0 to 180 (size of the population) and then move it back from 180 to 0.
27. R: Why do you want to do with that?
28. G: I want to see how spread out the data are for Years 7−9 and how their distributions change as more data values are added or taken out.

The students then spent time adjusting the slider, moving it forward and backwards and looking simultaneously at the representations of data. As they progressed, they expressed their preference to engage in data investigations that involved activities of growing and reducing the size of samples (lines 26–28).

## Discussion

The above analysis of students' excerpts sheds some light on the developmental process of students' inferential reasoning about samples and sampling issues. The findings demonstrate that the two students placed highest emphasis on the distribution of sample data to make inferences about the population. It is also evident that students forged new connections about the interplay of sample size and population, and they further linked those concepts to other statistical fundamental concepts during their investigations, such as spread, distribution, (explained) variation in data, unexplained variation, uncertainty, randomness and graph interpretation.

In this paper, there is evidence that the students perceived the importance of large samples (line 21). It is likely that they had a global resource such as the Law of Large Numbers available to them. Nevertheless, students felt comfortable to explore the impact of the sample size on data representations when they engaged in data investigations which involved activities of growing and reducing the size of samples. This shows that the students needed to have a broader experience of the interrelationship of sample size and data representation. The emergent new idea of "reducing the size of samples" or "shrinking the size of samples" needs to be further improved and elaborated.

This paper gives some light into students' emerging quantification of confidence intervals in making informal inferences. The above activity demonstrates students' changes in thinking towards a situated abstraction, which was schematised as "I am giving an 8–9 interval of how certain I am … I can not be so sure … The more backpacks we weigh, there will be less room for mistakes or unknown backpacks weights" (lines 18–20). Rafael seemed to acknowledge how variation (in the weights of backpacks) arises, and the uncertainty caused by unexplained variation in the weight of backpacks (line 20) such as measurement errors. Such understanding of variation in a real situation is prerequisite in making informal statistical inferences.

As mathematics educators, we need to ask ourselves about the level of confidence students place in their informal inferences. It would be interesting to explore the level of confidence students place in drawing informal conclusions about a population based on sample data. Should we be satisfied with increasing our understanding of how such decisions are made or should we consider this evidence as a pedagogic challenge to find ways to support changes in our students' thinking towards an abstraction, which might be schematised as “the bigger the sample we have, the more confidence we could place in our informal inferences”? This is potentially an unusual question for the mathematics curriculum.

## References

Australian Curriculum, Assessment and Reporting Authority. (2010). *Australian Curriculum: Mathematics*. Version 1.1. Retrieved March 15, 2011, from http://www.acara.edu.au

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64–83.

Ben-Zvi, D., Makar, K., Bakker, A., & Aridor, K. (2011, February). *Children's emergent inferential reasoning about samples in an inquiry-based environment*. Paper presented to the 7th Congress in Mathematics Education. Rzeszow: Poland.

Cantania: Tec Smith Corporation (2000). *Catania studio* (Version 6.0) [Computer software]. Okemos, MI: Tec Smith Corporation. Retrieved October 20, 2009, from http://www.techsmith.com/camtasia.asp

Cobb, P., Confrey, J., diSessa, A. A., Lehrer, R., & Shauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*, 9–13.

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

Konold, C., Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education, 2*(1), Article 1. Retrieved October 20, 2009, from http://respositories.cdlib.org/uclastat.cts/tise/vol2/iss1/art1

Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic data exploration* (Version 1.0) [Computer software]. Emeryville: CA: Key Curriculum Press.

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

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

Pfannkuch, M. (2008, July). *Building sampling concepts for statistical inference: A case study*. Paper presented at the 11th International Congress on Mathematics Education, Monterrey, Mexico.

Pratt, D., Johnston-Wilder, P., Ainley, J., & Mason, J. (2008). Local and global thinking in statistical inference. *Statistics Education Research Journal, 7,* 107–129.

Robson, C. (1993). *Real World Research*. Oxford: Blackwell.

Tversky, A., & Kahneman, D. (1971/1982). Belief in the Law of Small Numbers. *Psychological Bulletin, 76,* 105–110. (Reprinted in D. Kahneman, P. Slovic, & A. Tversky [1982] *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press.)

Watson, J. M. (2004). Developing an awareness of distinction. 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. (2000a). Developing concepts of sampling. *Journal for Research in Mathematics Education, 31,* 44–70.

Watson, J. M., & Moritz, J. B. (2000b). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behaviour, 19,* 109–136.