CHILDREN'S EMERGENT INFERENTIAL REASONING ABOUT SAMPLES IN AN INQUIRY-BASED ENVIRONMENT

Dani Ben-Zvi[1], Katie Makar[2], Arthur Bakker[3], & Keren Aridor[1]

[1]University of Haifa, [2]The University of Queensland, [3]Utrecht University

Research on informal statistical inference has so far attended little to sampling. This paper analyzes children's reasoning about sampling when making informal statistical inferences in an inquiry-based environment. Using data from a design experiment in Israeli Grade 5 (age 11) classrooms, we focus on the emergent reasoning of two boys working with TinkerPlots on investigations with growing sample size. They turn out to have useful ideas about whether inferences can be made from samples of different sizes. Initially, they oscillate between deterministic and relativistic conclusions, but they come to reason in more sophisticated ways with increasing awareness of what is at stake when making inferences from samples.

INFORMAL STATISTICAL INFERENCE

One of key things that statistics allows us to do is to draw inferences from samples. Doing so with formal techniques such as estimation, confidence intervals or hypothesis testing goes well beyond what most students will have the opportunity to learn, yet it seems important to give them a sense of the power of statistics by making such inferences informally (Garfield & Ben-Zvi, 2008). For these reasons, statistics educators have studied *informal statistical inference*, characterized as a generalized conclusion expressed with uncertainty and evidenced by, yet extending beyond, available data (Makar & Rubin, 2009). Two special issues have already been dedicated to this theme (Makar & Ben-Zvi, in press; Pratt & Ainley, 2008).

In earlier work we addressed the question of how informal inferential reasoning, the reasoning processes leading to informal statistical inference, can be nurtured and supported (Makar, Bakker, & Ben-Zvi, in press). Supporting elements include statistical concepts and tools, knowledge of the problem context, inquiry drivers such as doubt, explanation, and resolution of cognitive conflicts. We proposed that an inquiry-based learning environment with suitable tasks and tools as well as teacher scaffolds is especially suitable to support students' informal inferential reasoning.

SAMPLES AND SAMPLING

The concept of sample is a central concept in statistics yet it has received limited attention in the research literature compared to other concepts such as average, variation, and inference. Concepts and issues surrounding sampling are complex and require coordinating several ideas at once. Researchers note that students and teachers often conflate samples with their population when working with data (Lavigne & Lajoie, 2007; Pfannkuch, 2008; Pratt, Johnston-Wilder, Ainley, & Mason, 2008). Others caution that students may hold extreme beliefs about relationships between samples and their population: Those focusing on sampling representativeness might believe that a sample provides complete information about a population, while students focusing on sampling variability might believe that a sample provides no information (Rubin, Bruce, & Tenney, 1990). Watson and Moritz (2000) found that children in their study (Grade 3, age 8-9) had fairly primitive notions of samples and were typically comfortable making claims from small samples with little concern about bias, while the older children (Grade 9, age 14-15) generally attended to both sample size and representativeness in making claims, recognizing potential problems of bias and variability of small samples. Students between these ages (Grade 6, age 11-12) held a diversity of beliefs about sample size and sampling, suggesting they are at a critical age in their development of concepts of sampling.

Recent interest has arisen about the potential of informal statistical inference as an organising principle in learning statistics. Several aspects of informal statistical inference have been addressed in the literature, but the role of sampling has received surprisingly little attention given its centrality in inference. Several researchers (e.g., Arnold & Pfannkuch, 2010; Konold & Kazak, 2008) have used hands-on activities, visualisations, and simulations in helping students coordinate the complex issues of sampling in inferential reasoning. In this paper we focus on the question: *How does children's reasoning about sampling emerge when making informal statistical inferences in an inquiry-based environment?* We use 2010 data from Ben-Zvi's Connections Project (Ben-Zvi, Gil, & Apel, 2005) to respond to this question by examining the work of two boys (aged 11) participating in a teaching experiment as they grapple with drawing inferences from samples of increasing sizes.

GROWING SAMPLES

A key idea behind the design is that of growing samples—an instructional idea mentioned by Konold and Pollatsek (2002), worked out by Bakker (2004) and elaborated by Ben-Zvi (2006). Starting with small data sets (e.g., n=8), students are expected to experience the limitations of what they can infer from them about the whole class. They are next asked to draw conclusions from the whole class and speculate on what can be inferred about the whole grade in the school. Bakker (2004) found that such an approach is helpful in supporting coherent reasoning with key statistical concepts such as data, distribution, variability, tendency, and sampling. What-if questions proved particularly stimulating.

METHOD

We address the research question by drawing on findings from a design study (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003) carried out in three Grade 5 classrooms in Israel. This study is part of the Connections Project—a longitudinal development and research project (2005-2010) aiming to develop an inquiry-based environment for learning statistics in grades 4-6 using *TinkerPlots* (Konold & Miller, 2005).

The setting and participants

The learning sequence was built around five cycles of extended data investigations (2-3 lessons of 90 minutes each) of a student-administered survey across several grades in their school. The survey gathered student information about dimensions of body parts, free time activities, pets ownership, etc. (33 variables, n=270), creating a rich and interesting database for investigation. In each cycle, students posed a research question, organized their sample data using *TinkerPlots*, and made sense of it to draw informal inferences.

In line with the literature on growing samples, the design of activities evolved around the idea of starting from a sample of size 8, moving to about 30 (a whole class), then 90 (a grade level), and finally 270 cases (entire cohort) (see Figure 2 in Ben-Zvi, 2006). Starting with a small sample size was a pedagogical design decision to draw students' attention to the limitations of small samples, gradually developing their reasoning about samples, confidence level in their inferences, and "what-if" questions about larger sample sizes (e.g., "If you had a sample size of X, would the inference you just made still hold?").

Students worked in pairs through a scaffolded, open-ended inquiry of the data, with some pairs presenting their investigations in front of the class for further discussion. Independent investigations were videotaped using Camtasia to capture both their computer screen and faces. In this paper we focus on one pair of academically successful and articulate boys—Liron and Shay—in their first two investigations.

The episodes

In the first episode—their first independent investigation with *TinkerPlots*—the pair studied issues about free time (e.g., what students do in their free time, preferred communication method). They were given a sample of 8 students from their class (including themselves) with eleven variables to analyze and make inferences beyond the data at hand and a handout with instructions and questions about sampling issues (e.g., "Would the conclusions you have reached apply also to half of the class? Please explain."). In the second episode, the sample was increased to 29 (whole class) and they were asked to see if their conclusions still held for the larger sample.

Data analysis

Videos were observed, transcribed, translated from Hebrew to English and annotated for further analysis of the development of their reasoning about samples in relation to inferences. Interpretations were discussed until consensus was reached. Differences between Hebrew and English connotations of words (inference, conclusion, sample) were discussed extensively. Episodes were selected to illustrate the boys' developing reasoning about samples when making informal inferences.

RESULTS

First investigation, n=8

After orienting to *TinkerPlots*, Shay and Liron organized the small data set with the software. In their first graph, the eight data points spread across six categories of students' free time activities, with most categories having only one point (Figure 1).

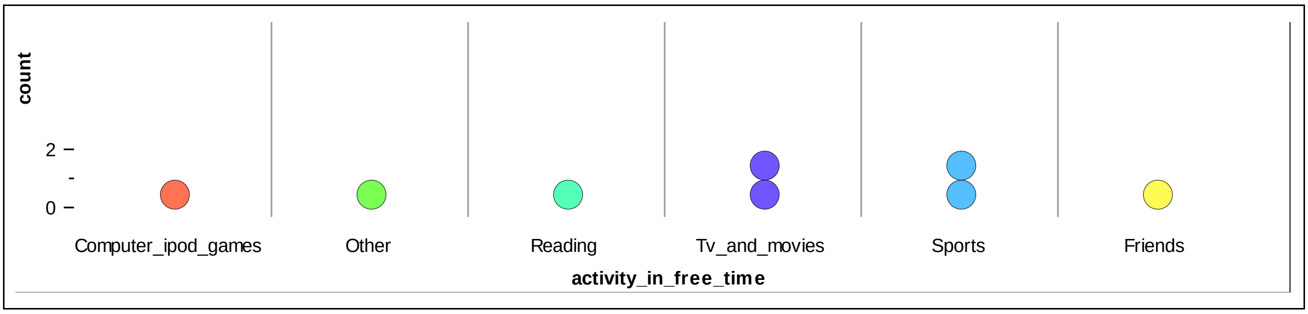

Figure 1: Liron and Shay's initial graph of 8 data points spread across 6 categories

They expressed initial dissatisfaction with working from such a small set of data: "We only have 8 kids, we don't have enough data! … We don't have enough to know things properly!" (Shay, lines 9-12) and found it frustrating to draw conclusions. With only eight data points, they considered them not to be "real data":

25 Shay: So let's see. Still, we don't have enough data that we can see because we have only 8 kids and it kind of spreads out. So we'll try to see something else, and then we will see if we have enough real data.

26 Liron: Pets is very easy. Let's check this first [scanning the pets data].

27 Shay: Well OK, we see it's [also] too spread out, and since we have only 8 kids, we don't have much to see. So let's try to see, err, what shall we try to see, Liron?

In looking for other variables to investigate, the boys seemed to search for an apparent pattern. With most variables having low frequencies in each category (due to the small sample), they characterized the data as too "spread out" to draw conclusions. This implies that their reluctance to draw conclusions involved both sample size and its relation to frequencies (in this case, large spread and a "flat distribution", cf. Ben-Zvi et al., 2005) and suggests a need to acknowledge students' statistical conceptions more broadly (e.g., variability and distribution) when focusing on sampling. When exploring the number of after-school activities per week (Figure 2), the boys finally felt able to draw a conclusion from numerical data. These data were not spread out like in the variable *activity in free time*:

39 Liron: 3 is the biggest. It is the most common.

40 Shay: According to what we see, the 'most mode one' is 3. What we see is that it can also be said that the average here is 3. The average is 3.

45 Shay: … Ok, wait, we have found something very interesting – we have the average number of children [activities] in a week is three. Let's save it.

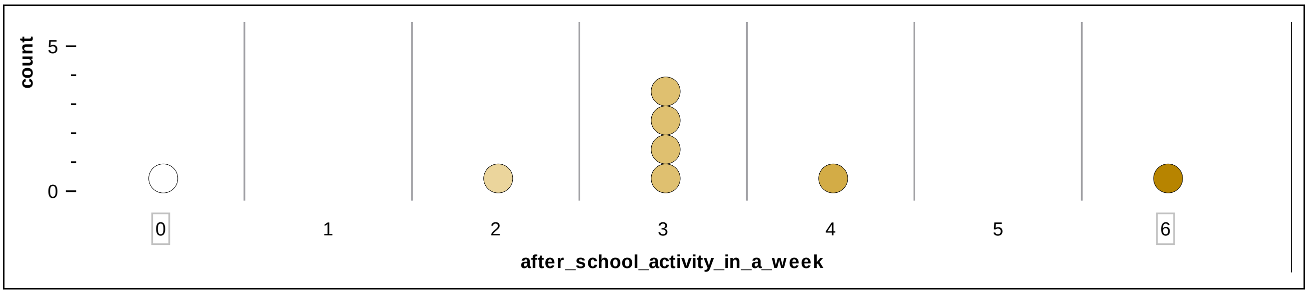

Figure 2: Liron and Shay's distribution with higher frequency in one value

As they searched for interesting and sensible stories in their data, they reviewed several additional variables (e.g., methods of communication with friends). However, after having done these analyses, Shay dismissed them almost immediately:

56	Shay:	So usually you use the telephone at home, some use the cell phone and chats on the internet. [But] as we said before, we don't have enough kids in order to have something, a proper result as needed.

Shay realized again that the small sample size was a flaw in the validity of the inference. The conflict of making valid conclusions from a small sample arose several times during the investigation. At other times, they set the issue aside and made fairly strong claims (e.g., "Boys like computers more than girls."). This oscillation between stating there is nothing to say and stating claims deterministically continued throughout this first investigation. We should cautiously note, however, that it is not clear whether Shay had changed his mind or went through the analysis with the impression that the results would be improper results. To overcome this concern, Liron proposed to collect more data.

One aim of the growing samples activities is to develop, as well as gain insight into, students' inferential reasoning about the relationship between conclusions drawn with a small sample and their inferences to larger samples (e.g., limitations of small samples, confidence about their inferences). When asked whether their claims applied more generally to the class, their initial response was quite confident that they would be. When pressed, however, their responses were qualified. "According to *these data for now*, only based on these, boys like computers more than girls and girls like to spend time with their friends more than boys" (line 198). Shortly thereafter, their responses changed again:

207	Int.:	And what if I ask you about the whole class? What do you think your inferences look like?
208	Shay:	It could be completely different!
210	Int.:	… What do you mean by "completely different"? Could it be, say, that girls like computers more than boys?
211	Shay:	It will not be *that* different.
213	Int.:	… Won't it surprise you? If you see for example that the girls in the whole class, there will be more girls that like computers than boys?

214 Shay: Yes, it will surprise me, but there is a chance it will happen.

215 Int.: Why will it surprise you?

216 Shay: Because I know from personal experience that boys like computers more than girls, but there is a chance that my personal experience is wrong! There is a chance.

Although Shay thought the results in the entire class might be completely different than the sample results, his answer in line 211 is more nuanced. This remark may suggest that Shay anticipates that the results in the sample compared to the entire class could be similar and he sees a small chance that the results will be completely opposite. This is the first time the role of chance is mentioned in the discussion, not in explicit relation to the sample size but rather to the boy's personal knowledge of the context. In the end of this discussion, the interviewer explored the limits of what the boys were willing to conclude from different sample sizes:

243 Int: Can you say based on the conclusion from these data that they are true for a larger group of kids?

244 Liron: No. Every child has a different opinion.

245 Int: Shay, what do you think?

246 Shay: I agree. Each child has their own characteristics.

247 Int: How many kids in your class?

248 Liron: 29.

249 Int.: What if we ask 15 kids? Do you think then we could conclude something?

250 Liron: Yes.

We saw this same reasoning about the ability to infer from half of the population in discussions among other students as well. Even at the end of this activity, the boys still oscillated between not being able to say anything ("Every child has a different opinion") and being able to infer something regarding bigger samples. Their confidence in drawing conclusions oscillated during their investigation. They repeatedly asked for more data, offering even to go to the class and collect it themselves; at the same time, they often made claims based on a frequency of one (e.g., girls like to talk to friends more than boys). When probed, they were able to qualify the claims, but were uncertain whether their claims would hold more broadly.

Second investigation, n=29

In the second investigation, students were given data for their entire class (29 students) as well as five additional variables (e.g., various body measurements, additional pet ownership data). Shay and Liron's immediate reaction was one of excitement "Wow! … It's so much fun now!" (Lines 255-257). Before investigating the data, Liron anticipated that the conclusions would be different from those they made for the sample of eight. This suggests one way that the growing samples sequence stimulated students to think about sample-population relationships:

258 Int.: Do you think they will be similar or different for the entire class and why?

259 Liron: Different.

260 Shay: Not just different. We should say in what [way] they will be different.

261 Liron: Eight kids can’t represent an entire class!

When the boys were given the data for the entire class, they compared their new findings (sample size 29) with those from the previous investigation (sample size 8).

265 Shay: I knew! Girls love to be with friends more than boys. Boys like computers more than girls. Here, here, my hypotheses are materializing!

266 Liron: Because?

267 Shay: Look, it's really beautiful. We really have conclusions!

Although the boys easily recognized the similarities, they didn’t understand why.

289 Liron: Most of our hypotheses were confirmed. …

292 Shay: … But I don't understand. It is the same data. It is unfair. I don't get it.

293 Liron: Neither do I, but what do you know?

294 Shay: I was sure it would be entirely different data.

295 Liron: So was I, but here it is. You can see it.

The boys expressed surprise by the unexpected similarity then questioned its validity. Their language became more subtle, focusing on the uncertainty.

307 Shay: But you know it happened by chance.

308 Liron: It happened by chance. They didn't do it on purpose.

And later:

316 Shay: This quite surprised me, and I thought that if I take more kids, it will change, but who knows, apparently it is the same thing.

317 Liron: It must have been a coincidence.

318 Shay: Maybe, not for sure.

When asked for an explanation, Liron said he had no idea and Shay repeated that it happened by chance. Later, following the design idea of growing samples, they were asked again whether they expected the result to be similar for the entire 5^{th} grade.

423 Shay: Now I think that if you take like, what happen is that it was [first only] 8. Now we took the entire class, it was exactly the same properties. So I think now that if you take the entire 5^{th} grade, it will also be the same conclusions.

424 Int.: Why? Try to explain once more why? You just said something and I am not sure I got you.

425 Shay: Because from the conclusions here, we saw that once we expanded it, there were the same conclusions. Now what do I infer from this?

426 Int.: What?

427 Shay: That it is really true. That it will keep being true, also.

When probed, the boys informally quantified their level of confidence about the conclusions, both now and from the original sample of eight, but cautioned that the results would not necessarily extend to other ages in the school.

441 Int.: In the same age, if we take a scale from 1 to 10, … how much are you certain from 1 to 10 that the results will remain the same?

442 Shay: 7.

443 Liron: I am also 7. …

447 Int.: How sure were you from [the sample of] 8 about the entire class? …

448 Shay: About 2. Maybe 3.

DISCUSSION

The brief excerpts from Liron and Shay's investigations give us some insight into initial ways that children can reason about samples in an inquiry classroom designed to provoke their informal statistical inferences. The analysis of Liron and Shay's inferential reasoning about sampling issues shows a development from fairly extreme and seemingly contradictory views of what can be concluded from a small sample to more nuanced statements about the strength of their later claims and emerging quantification of confidence in making inferences. In the first investigation, the boys repeatedly expressed their lack of confidence in conclusions drawn from only eight data points while concurrently making fairly strong claims (e.g., boys like computers more than girls do) based on frequencies of only one or two data points. As they progressed, they qualified their claims as only holding for their limited data and were rightly conflicted about whether the sample would provide any information about a larger population. In the second investigation, the larger data set confirmed many of their previous conclusions, which surprised them. This confluence appeared to provoke them to question the way the data were gathered.

In addition to working with concepts of samples and informal statistical inference, strong links to other statistical key concepts arose during the investigation, such as average (average of 3 activities per week), spread and distribution, likelihood, randomness, and graph interpretation. The excerpts further underline Bakker and Derry's (in press) discussion of the importance of a holistic approach to concept development as these concepts were not encountered in isolation, but emerged collectively as relevant tools by provoking their reasoning in the context of inquiry. The opportunity to probe the students' reasoning was helpful to explore the scope, flexibility and robustness of students' concepts in action. This aligns with inferentialist approaches to have students reason within the context of complex problems and through that process find the scope and limit of their conceptions in the act of reasoning with them.

We end with a few questions for future research:

1. How can ideas about sampling in relation to informal statistical inference be further developed in the next grade?
2. What was the role of the activities design and the inquiry-based environment in the development of students' reasoning about samples and sampling? And in particular how can the instructional idea of growing samples can be further improved and used?
3. How can new tools (e.g., *TinkerPlots2*, see Konold et al., in press) probe students' inferential reasoning?

ACKNOWLEDGEMENTS

We gratefully acknowledge Hana Manor who interviewed the two boys, and Einat Gil who supervised the activities design. The Freudenthal Institute funded part of Dani Ben-Zvi and Katie Makar's visit to Utrecht where the videos were analyzed.

REFERENCES

Arnold, P., & Pfannkuch, M. (2010). Enhancing students' inferential reasoning: From hands on to "movie snapshots". In C. Reading (Ed.), *Proceedings of the 8th International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.

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

Bakker, A., & Derry, J. (in press). Lessons from inferentialism for statistics education. To appear in *Mathematical Thinking and Learning, 13*(1&2).

Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Proceedings of the 7th International Conference on Teaching of Statistics* (CD-ROM), Salvador, Bahia, Brazil, 2-7 July, 2006. Voorburg, The Netherlands: International Statistical Institute.

Ben-Zvi, D., Gil, E., & Apel, N. (2005). What is hidden beyond the data? Helping young students to reason and argue about some wider universe. In D. Pratt & J. Ainley (Eds.), *Reasoning about informal inferential statistical reasoning: A collection of current research studies*. Proceedings of the Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-5), University of Warwick, UK, August, 2007.

Cobb, P., Confrey, J., diSessa, A. A., Lehrer, R., & Schauble, 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.

Konold, C., Madden, S., Pollatsek, A., Pfannkuch, M., Wild, C., Ziedins, I., Finzer, W., Horton, N. J., & Kazak, S. (in press). Conceptual challenges in coordinating theoretical and data-centered estimates of probability. To appear in *Mathematical Thinking and Learning, 13*(1&2).

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

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.

Lavigne, N., & Lajoie, S. (2007). Statistical reasoning of middle school children engaged in survey inquiry. *Contemporary Educational Psychology, 32*, 630-666.

Makar, K., Bakker, A., & Ben-Zvi, D. (in press). The reasoning behind informal statistical inference. To appear in *Mathematical Thinking and Learning, 13*(1&2).

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

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., & Ainley, J. (Eds.) (2008). Informal inferential reasoning. Special Issue of the *Statistics Education Research Journal, 7*(2), 3-4.

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

Rubin, A., Bruce, B., & Tenney, Y. (1990). *Learning about sampling: Trouble at the core of statistics*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

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