ENGAGING YOUNG CHILDREN IN INFORMAL STATISTICAL INFERENCE

Efi Paparistodemou[1] and Maria Meletiou-Mavrotheris[2]
[1]Ministry of Education, Cyprus
[2]European University, Cyprus
e.paparistodemou@cytanet.com.cy

At the lower levels of schooling, students' exposure to statistical concepts has been restricted to basic descriptive statistics. In recent years, however, leaders in mathematics education have advocated a much wider and deeper role for statistics in school mathematics. Reflecting the recent shift in statistics education research from a focus on specific skills and procedures towards a greater focus on statistical reasoning and thinking embedded in the process of a statistical investigation, the current study was designed to investigate ways in which the foundations of inferential reasoning can be laid at a very young age. This paper reports on how a group of young students formulated and evaluated data-based inferences. The term informal inference is used here to describe the drawing of conclusions from data that is based mainly on looking at, comparing, and reasoning from distributions of data.

INTRODUCTION

Since formal statistical inference ideas and techniques are beyond the reach of young learners, an informal approach to statistical inference is necessary in the early years of schooling (Ben-Zvi, 2006). According to Watson (2007), informal statistical inference represents a continuum of experience from the point when students start to pose questions about datasets to the point when they are about to meet formal inferential statistics. Along the way from informal to formal inference, a number of important ideas are added to the student package. Rubin et al. (2006) define informal inference as reasoning that involves consideration of the following related ideas: (i) properties of aggregates rather than properties of individual cases, (ii) sample size and its effect on the accuracy of population estimates or on process signals, (iii) controlling for bias, and (iv) tendency, distinguishing between claims that are always true and those that are often or sometimes true. Moreover, Makar and Rubin (2007) regard informal inferential reasoning in statistics as a reasoned but informal process of creating or testing generalizations from data that extend beyond the data collected. They consider the following three principles to be essential to informal inference: (i) making generalizations (predictions, parameter estimates, conclusions) that extend "beyond the data," (ii) using data as evidence for these generalizations, and (iii) using probabilistic language in describing the generalizations, including references to levels of certainty about the conclusions drawn.

Recent advances of technology provide schoolteachers and college instructors with new tools for adopting informal, data-driven approaches to statistical inference that can help lay the conceptual groundwork for formal inferential reasoning (Rubin et al., 2006). The appearance, in particular, of dynamic statistics learning environments (e.g., TinkerPlots® and Fathom®), which are designed explicitly to facilitate the visualization of statistical concepts, provides enormous potential for making inferential reasoning accessible to students. These new technological tools provide a medium for the design of activities that integrate experiential and formal pieces of knowledge, allowing students to make direct connections between physical experience and its formal representations (Paparistodemou, Noss & Pratt, 2008). Students can experiment with statistical ideas, articulate their informal theories, use the theories to make conjectures, and then use the experimental results to test and modify these conjectures. Several researchers have, in recent years, been exploiting the affordances provided by these modern technologies for promoting learners' ability to reason and argue about data-based inferences, with very encouraging results (e.g., Bakker, 2004; Ben-Zvi, 2006; Rubin et al., 2006). The conceptual "Framework for Teaching Statistics within the K-12 Mathematics Curriculum" (GAISE report, 2005), guided the program design. This framework focuses on building learners' conceptual understanding of the statistical process by emphasizing and revisiting, with increasing sophistication through the grade levels, a set of central statistical ideas. It uses a spiral approach to the statistics curriculum, so that instructional programs from early childhood through high school encourage students to gradually develop

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

understanding of statistics as an investigative process that involves the following components:

(i) clarifying the problem at hand and formulating questions (hypotheses) that can be answered with data,

(ii) designing and employing a plan to collect appropriate data,

(iii) selecting appropriate graphical or numerical methods to analyze the data: summarizing the data, making conjectures, drawing conclusions, making generalizations, and

(iv) interpreting the results of the analysis and relating the interpretation to the original question.

Consistent with a theoretical perspective on statistics instruction at the school level, the study was focused on building young learners' informal inferential reasoning by enabling them to experience and develop the "big ideas" of statistics through the collection and exploration of real data of interest to them.

THE STUDY

The study reported in this article is based on a previous research (Paparistodemou & Meletiou, 2008) which aimed at fostering third-grade students' informal notions of inference through adopting a hands-on, project-based approach to statistics using the dynamic statistics software TinkerPlots® as an investigation tool and social activity focusing on making data-based arguments (Stohl & Tarr, 2002). The term informal inference is used here to describe the drawing of conclusions from data that is based mainly on looking at, comparing, and reasoning from distributions of data (Pfannkuch, 2006).

Paparistodemou & Meletiou (2008) suggest three categories of informal inference: (a) data-based argumentation, (b) data-based argumentation and generalization, and (c) data argumentation and chance. Data based argumentation refers to children's conclusions based on the data they had collected; Data-based argumentation and generalization refers to children's conclusions about their data and using the data to draw inferences about a larger population without engaging the idea of chance; and Data argumentation and Chance refers to children's conclusions about and using the data to draw inferences about an unknown population by engaging the idea of chance (e.g., using expressions such as 'more likely,' 'might be,' 'more possible to'). The particular study investigates these categories with another group of students with focus on the analysis of more than one variable.

The study took place in Spring 2009, working with a group of twenty-nine, 9 year-old, students. The students' reasoning about informal inference was studied through genuine statistical endeavours using the dynamical statistics environment TinkerPlots® (Konold & Miller, 2005).

Students participated in data-centered activities, in contexts familiar to them. After collecting real data about themselves, they worked in small groups to explore the data and to formulate and evaluate data-based inferences, using TinkerPlots® as an investigative tool. These children had some previous experiences with graphing 'by hand,' like collecting data from their class (e.g., the colours that their classmates like), making simple graphs (usually object graphs or bar graphs) by using grid paper, and drawing conclusions like finding the most popular vegetable amongst the children of their class and the number of children that liked to eat this vegetable. They had no previous exposure to TinkerPlots® other than going through a short tutorial which introduced them to the main features of the software. They became acquainted with TinkerPlots® while analyzing data in this study.

RESULTS

The snapshots represented in the following sections come from students' interactions with the software and with each other while analyzing the data and creating Figure 1 and Figure 2. These snapshots show children's informal statistical inferential reasoning while trying to derive conclusions from data. The emphasis is placed on data-based argumentations, data-based argumentation and generalization and data argumentation and chance. On Figure 1 and Figure 2 represented below, children used English characters for Greek words, as there were not Greek characters on the software.

In Figure 1 'LAX_HMERA' means HOW MANY VEGETABLES I EAT IN ONE DAY, 'LAX_FYTEYO' means VEGETABLES I PLANT; 'NAI' means YES, 'OXI' means NO;

'KANENA' means NOTHING and 'PER_4'means MORE THAN 4. The question was how many vegetables the children of their school eat in relation with the fact that the children plant vegetables at home. The children collected answers from the whole school (128 children).

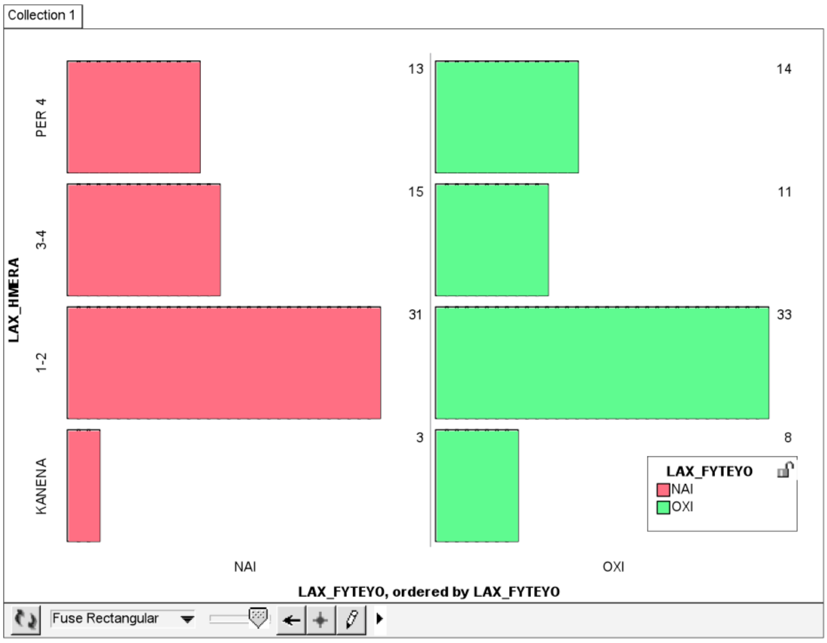

Figure 1. The number of vegetables that the children eat
in relation to the fact that they plant vegetables at home

The following group of three students worked on Figure 1. It is interesting how their statistical inference is developed through the categories that we have already presented in a previous research (Paparistodemou & Meletiou, 2008). Moreover, it is interesting that the students try to find links between the amount of vegetables that the students eat and whether they plant vegetables. At the end, the children decided also to find out how the amount of eating vegetables is related with the gender of children.

Researcher: What do you think about this graph?
George: You *see* here! We have the more students to eat 1-2 vegetables per day, and also there is no difference between planting and not planting their own vegetables.
Melina: Hm..You see that in our school most of the students do not plant their own vegetables but there are also many children that they plant vegetables at home. [*Data-based argumentation*]
Basil: I do not believe that this is something that happens *generally in our town*...I mean that people in town do not plant their own vegetables. They might have parsley, but not too many of them plant their own cucumbers!
George: Yes! But again, you see...as here, I don't think that this influenced the amount of vegetables they eat per day. If somebody thinks that is important to eat vegetables or like them, he will eat it! And this is exactly what our graph shows [*Data-based argumentation and generalization*].
…
Researcher: So, if you find a child that does not eat vegetables in our school what would you say? Does she/he plant her/his own vegetables?
Melina: I would have said that she *probably* does not have a garden with vegetables at home. This is something that we also get from the graph.
George: But, if she eats vegetables, we cannot *make a guess* whether or not she plants vegetables at home! We cannot draw any conclusions [*Data argumentation and chance*].

Melina: I would like to find out what is happening with boys and girls in our school [*Data-based argumentation*].
[The children created Figure 2, where 'FYLO' means GENDER, 'A' means BOY and 'K' means GIRL.]

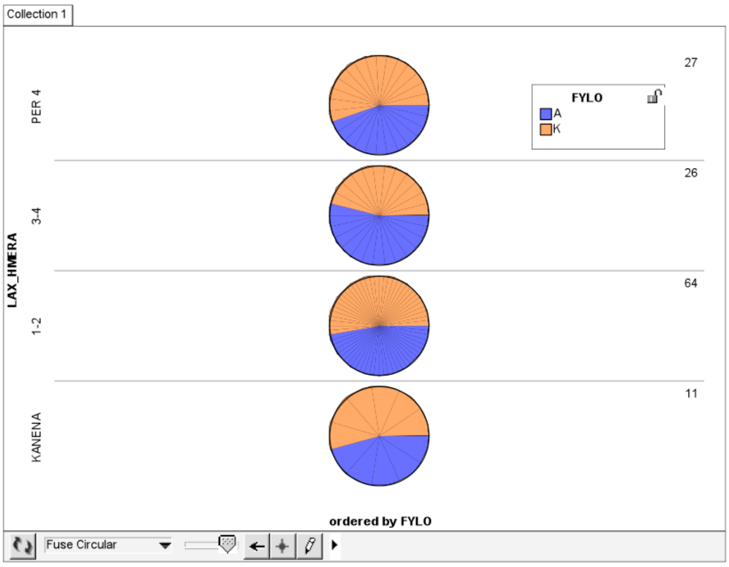

Figure 2. The number of vegetables that the children eat in relation to gender

Melina: Hmm…again not many things to say…There is not a relation between gender and how many vegetables a child eats per day (see Figure 2). [*Data based argumentation and generalization*].
Basil: We see that a person who eats more than 4 vegetables per day *might* be a girl [*Data argumentation and chance*]….you know…girls take care of calories and vegetables do not have many (calories)!

The young children in our study expressed statistical informal inference in three distinct ways: (a) data-based argumentation, (b) data-based argumentation and generalization, and (c) data argumentation and chance. Children drew their conclusions based on the data they had collected, using the data to draw inferences about a larger population without engaging the idea of chance, and using the data to draw inferences about an unknown population by articulating uncertainty (e.g., using expressions such as 'probably,' 'might be,' 'make a guess').

DISCUSSION

It seems that informal inference provides new opportunities to introduce powerful statistical concepts early in the school curriculum. This study shows that statistics can be used as a tool for gaining insights into understanding problems, rather than only as a collection of graphs, calculations, and procedures (Sorto, 2006). Findings from this study show that young learners begin to reason about informal inference when their interest in the task is high. The children in this study were very much involved with data and the conclusions drawn from the data were important for them in order to understand what was happening at their school. At this age, personal experience and interest play a key role in children's interactions with data. Personal interest is important for children's involvement in reasoning about informal inference.

Moreover, the study is an example of an approach to improving students' use of statistical reasoning and thinking by embedding statistical concepts within a purposeful statistical investigation that brings the context to the forefront. It is not just making a conclusion about data that provides the conceptual muscle to draw inferences, but a conclusion about the situation that the data are meant to represent or signify (Makar & Rubin, 2007). The focus is on understanding the situation (Makar & Confrey, 2007), rather than examining decontextualized data. Perhaps a focus on an interesting problem and engaging context may influence students' inclination to look beyond the data they have. Making a conclusion about the situation suggests that students need a particular level of complexity to engage with in order to consider possible avenues to connect the data with the context.

Attributes of TinkerPlots® like the ability to operate quickly and accurately, to dynamically link multiple representations, to provide immediate feedback, and to transform an entire representation into a manipulable object enhanced students' flexibility in using representations and provided the means for them to focus on statistical conceptual understanding. Technology has an important role to play in promoting the development of informal inferential reasoning in the early grades. Innovative educational software such as TinkerPlots® allows children to explore informal inferential ideas in contexts that are both rich and meaningful to them. Use of such software, in combination with appropriate curricula and instructional settings, can help students develop a strong conceptual base on which to build a more formal study of inferential statistics later.

The expanding use of data for prediction and decision-making in almost all domains of life makes it a priority for mathematics instruction to help all students develop their inferential statistical reasoning. As the current study and several other studies have illustrated (e.g., Ben-Zvi, 2006; Makar & Rubin, 2007), when given the chance to participate in appropriate instructional settings that support the development of informal inferential reasoning, even very young children can develop intuitions about fundamental statistical concepts related to statistical inference. More research is needed on designing instruction and on building teachers' and students' concepts and reasoning about informal inference.

The challenge for future research should be to develop teachers' and students' sampling conceptions in terms of learning to reason about populations from samples using informal inference. The focus should shift from examining how students learn and use the different statistical tools and procedures, to how they understand the statistical investigation cycle as a process of making data-based inferences (Makar & Rubin, 2007). The study shows that students should come to appreciate the real reasons for which statistical tools and procedures have been developed–to help humans understand underlying phenomena and make informed decisions.

REFERENCES

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal, 3*(2), 64-83. Online: www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_Bakker.pdf

Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education: Proceedings of the Seventh International Conference of Teaching Statistics (ICOTS-7)*, Salvador, Brazil. Voorborg, The Netherlands: International Statistical Institute. Online: www.stat.auckland.ac.nz/~iase/publications/17/2D1_BENZ.pdf

GAISE Report (2005). *Guidelines for assessment and instruction in statistics education: A Pre-K-12 Curriculum Framework*. Alexandria, VA: The American Statistical Association. Online: http://www.amstat.org/education/gaise

Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic Data Explorations*. Emeryville, CA: Key Curriculum Press.

Makar, K., & Confrey, J. (2007). Moving the context of modeling to the forefront. In W. Blum, P. Galbraith, H-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education*. New York: Springer.

Makar, K., & Rubin, A. (2007, August). *Beyond the bar graph: Teaching informal statistical inference in primary school*. Paper presented at the *Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-5)*, Warwick, UK.

Paparistodemou, E., & Meletiou-Mavrotheris, M. (2008). Enhancing Reasoning about Statistical Inference In 8 Year-Old Students. *Statistics Education Research Journal, 7*(2), 83-106.

Paparistodemou, E., Noss, R., & Pratt, D. (2008). The Interplay Between Fairness and Randomness in a Spatial Computer Game. *International Journal of Computers for Mathematical Learning, 13*(2), 89-110.

Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education: Proceedings of the Seventh International Conference of Teaching Statistics, Salvador, Brazil.* [CDROM] Voorburg, The Netherlands: International Statistical Institute. Online: www.stat.auckland.ac.nz/~iase/publications/17/6A2_PFAN.pdf

Rubin, A., Hammerman, J., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics,* [CDROM]. Salvador, Brazil. Voorburg, The Netherlands: International Statistical Institute. Online: www.stat.auckland.ac.nz/~iase/publications/17/2D3_RUBI.pdf

Watson, J. (2007, August). Facilitating beginning inference with TinkerPlots for novice grade 7 students. Paper presented at the *Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-5)*, Warwick, UK.

Shaughnessy J. M., Ciancetta M., Best K., & Canada D. (2004, April). Students' attention to variability when comparing distributions. Paper presented at the *82nd Annual Meeting of the National Council of Teachers of Mathematics*, Philadelphia, PA.

Sorto, M. A. (2006). Identifying content knowledge for teaching statistics. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM] Voorburg, The Netherlands: International Statistical Institute. Online: www.stat.auckland.ac.nz/~iase/ publications/17/C130.pdf

Stohl, H., & Tarr, J. E. (2002). Using multi-representational computer tools to make sense of inference. In D. Mewborn (Ed.), *Proceedings of the Twenty-fourth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Athens, GA: Columbus, OH: ERIC Clearinghouse for Science Mathematics and Environmental Education.