12th International Congress on Mathematical Education
Program Name XX-YY-zz (pp. abcde-fghij)
8 July – 15 July, 2012, COEX, Seoul, Korea (This part is for LOC use only. Please do not change this part.)

LAYING FOUNDATIONS FOR STATISTICAL INFERENCE

Maxine Pfannkuch

The University of Auckland

m.pfannkuch@auckland.ac.nz

Chris J. Wild

The University of Auckland

c.wild@auckland.ac.nz

In this paper we give an overview of a five-year research project on the development of a conceptual pathway across the curriculum for learning inference. The rationale for why statistical inference should be part of students' learning experiences and some of our long deliberations on explicating the conceptual foundations necessary for a staged introduction to inference are described. Implementing such a pathway in classrooms required the development of new dynamic visualizations, verbalizations, ways of reasoning, learning trajectories and resource material, some of which will be elucidated. The trialing of the learning trajectories in many classrooms with students from age 13 to over 20, including some of the issues that arose, are briefly discussed. Questions arising from our approach to introducing students to inferential ideas are considered.

Keywords: Secondary-university students; Sampling variability; Visualisations; Verbalisations

INTRODUCTION

Traditionally statistical inference is the focus of the final year of high school with previous learning experiences featuring constructing plots and describing them. Research over many years (e.g., Chance, delMas, & Garfield, 2004) has consistently demonstrated that for the majority of students formal statistical inferential reasoning eludes them. One conjectured reason is that inference is grounded in mathematics and is presented as a procedure to follow. The resultant mathematical manipulations and calculations then act as obstacles to understanding the thinking behind inference. The second conjectured reason is that concepts underlying inference such as sample, population, and sampling variability should be developed over time in the curriculum rather than presenting such a complex network of integrated ideas in the final school year.

In the last decade, with increased access to technology, some curricula have adopted an exploratory data analysis (EDA) approach allowing students to investigate real data to look for interesting patterns and trends. But as Konold and Kazak (2008, p. 1) explain:

> As researchers began studying settings in which students were introduced to EDA, an unsettling picture soon emerged. Students given considerable exposure to and instruction in data-analysis techniques nevertheless had difficulties performing one of the most basic

tasks in analyzing data — judging whether two groups appeared different by comparing their averages or the approximate centers of their distributions.

Pfannkuch (2006) in her research also noted that students had difficulties in comparing two groups. One reason was students did not know what game they were playing: Game One involving reasoning only about the sample or Game Two involving reasoning about the population from the sample. If students were meant to be playing Game Two then another reason for their difficulties was that they had no learning experiences of sampling variability and other underpinning concepts such as the notion of a population distribution. In further research Pfannkuch (2008, 2011) noted that even when students experienced sampling variability for quantitative and qualitative data the basic problem of how they should judge whether one group tended to have larger values than another group still remained. These required understandings to make a judgment were conceived as *informal inference* by an International Forum of Statistical Reasoning, Thinking and Literacy led and founded by Joan Garfield and Dani Ben-Zvi (see special issue of Statistics Education Research Journal, Pratt and Ainley, 2008). Biehler (2011), when reflecting on the research presented at the 2011 Forum on how students were making inferences, argued that the responses of students across the year levels from junior to middle to senior needed to provide some evidence of growth in conceptual understanding and he raised questions about how that growth might be characterized and the type of learning approaches that might contribute to such growth.

Noting that problems were occurring with statistical inference using both the traditional and EDA approaches and the fact that comparing two groups was a problem for New Zealand teachers with respect to national assessment of 15 year-olds, we embarked on a journey to find a pathway for realizing statistical inference ideas and in particular to place substantive conceptual foundations into the statistics curriculum.

OVERVIEW OF RESEARCH PROJECT

Our five-year project used a four-phase design research cycle incorporating identification of the problematic situation, promulgation of the conceptual foundations to inform the design process, designing the learning trajectories, and testing them in the classroom (Hjalmarson & Lesh, 2008). It is an interactive cycle whereby, for example, the design of learning trajectories can raise conceptual issues, which need to be revisited. The cycle is continued as new problems are identified, hypotheses are generated and theories are conjectured. The instructional materials are designed in an attempt *to engineer and support a new type of learning and reasoning.* Design research engages researchers in improving education and provides results that can be readily used by practitioners (Bakker, 2004a; Cobb & Gravemeijer, 2008; Schwartz, Chang, & Martin, 2008). The methodology employed in the project used a mixed methods approach of pre- and post-tests, interviews, observations, and reflections.

In the first year of the project in 2008 a statistics education researcher and three statisticians debated, argued, and mapped out a potential pathway for introducing inferential ideas from the beginning of secondary school to the first year of university. In the next two years a team of two statisticians, two statistics education researchers, and nine teachers collaboratively worked through two design research cycles. Four classes in both years (a total of about 200 students), whose teachers were in the project team, participated. These classes covered a range of school socio-economic levels, student abilities, ethnicities and ages (13-16). In the final two years, 2011 and 2012, a team of 33 people consisting of statisticians, education researchers, Year 13 (last year of high school) teachers, and university statistics lecturers collaborated on identifying conceptual underpinnings and developing learning trajectories for statistical inference using bootstrapping and randomization methods. About 3000 university students and 200 Year 13 students participated in the implementation. The main data collected were: pre- and post-tests from all students, pre- and post-interviews of a sample of these students, videos of some classes implementing the teaching unit, and teacher reflections.

CONCEPTUAL FOUNDATIONS: INFORMAL STATISTICAL INFERENCE

A staged introduction to the conceptual foundations of statistical inference in the Year 10 (aged about 14) to Year 13 (aged about 17) New Zealand secondary school curriculum was constrained by time (four weeks per year for Years 10 to 12), access to technology (limited to one computer and a data projector per classroom), resources (very limited school budgets), and national assessment at Years 11, 12, and 13 with its requirement to show progression of growth in skills, concepts, and thinking. With these constraints in mind we chose to start with teachers' current problem in national assessment on how to compare two groups using box plots. As Biehler (1997) had already noted, a rich conceptual repertoire underpinned these comparisons such as sample, population, sampling variability, distribution, and sample size effect. We deliberated on and debated many conceptual foundations in response to the literature, analyses of data from the research project, and our own reflections. In this section we briefly discuss three issues, namely, making a call, sample-population ideas, and sampling variability, give some examples from our research and highlight some pivotal moments in our thinking.

Making a call

First, we deliberated on how students could make a call directly from the comparison of two box plots that had a statistically sound basis. Starting with current class practice and what was possible using hands-on simulations we determined that initially the sample size should be fixed at about 30 and through running simulations found a quick rule-of-thumb for making or not making a call in terms of shift of the boxes (the middle 50%) and location of the medians (Fig. 1). To progress from these notions the next stage was to consider that spread and sample size matter when making a call, which is then taken further to consider an informal confidence interval based on the work of Tukey (see Wild, Pfannkuch, Regan, and Horton, 2011 for more detail). At the next level, still using the idea of taking a random sample from the population the confidence interval is formalized using the bootstrap method. In addition, students are introduced to comparative experiments where random samples are not taken from populations rather convenient samples are used and units are randomly allocated to one of two treatment groups and hence the randomization method is used for making or not making a claim.

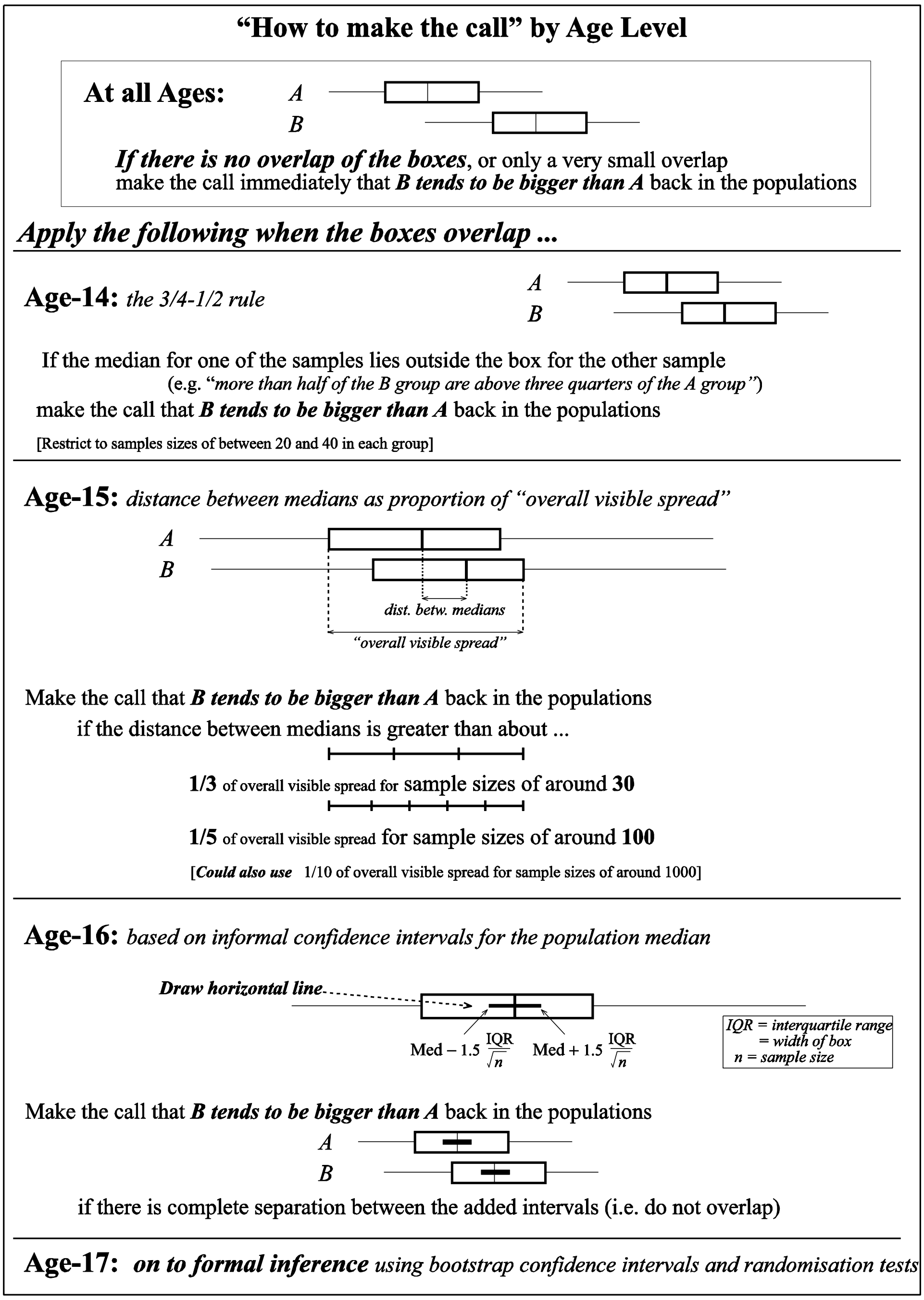


Figure 1. Guidelines for making a claim

Introducing secondary students to making a claim or call through the direct comparison of box plots necessitated the development of new dynamic visualizations and verbalizations. In particular, a breakthrough in determining some of the verbalizations occurred when we were looking at the example in Figure 2 from the GAISE K-12 Report (2007, p. 47).

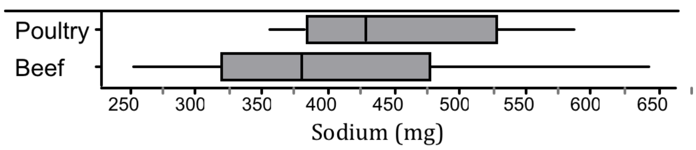


Figure 2. Sodium content of hot dogs

> The median sodium content for poultry hot dogs is 430 mg, almost 50mg more than the median sodium content for beef hot dogs. The medians indicate that a typical value for the sodium content of poultry hot dogs is greater than a typical value for beef hot dogs. The range for the beef hot dogs is 392 mg, versus 231 mg for the poultry hot dogs. The ranges indicate that, overall, there is more spread (variation) in the sodium content of beef hot dogs than poultry hot dogs.

We realized that the verbalizations in this example were mixing up descriptive and inferential thoughts. That is, describing what could be seen in the plots and inferring what might be happening back in the populations. We noticed the subtleness of the language with the use of the definite article when referring to the sample and non-use of the definite article when referring to the population. We wondered how students not exposed to sample and population ideas would know which statements were descriptive and which were inferential. We realized that for learners we would need to be careful in separating out the descriptive from the inferential but often the boundary between them was blurred. Another consideration was contextual thoughts that also needed to be invoked when making a claim. (For the full discussion see Pfannkuch, Regan, Wild, and Horton, 2010.)

Sample-population ideas

A second conceptual foundation we deliberated on was the notion of sample and population and the links between them. From previous research (Pfannkuch, 2006), we knew that students thought they were reasoning about the sample not the population. To assist students to conceptualize that they were reasoning about the population from the sample, we gave students population bags of over 600 datacards from which they drew a sample. Each datacard had 13 pieces of information about a student such as height, year level and gender. The datacard information about each student was drawn from a CensusAtSchool online survey of over 30,000 New Zealand students. To reduce confusion about the New Zealand school student population, the subpopulation that answered the survey and the sub-subpopulation of the datacards, we chose, after much debate, not to refer to all these populations but simply present the bag of datacards as the population. Students were gradually introduced to the notion that they were reasoning about a population from a sample through individually physically sampling from the population bag, plotting the data, and comparing their plots to other students' plots to see whether the message in the data was similar. Teachers in the study reported that a powerful reminder to students that they were reasoning about the population was the population bag, to which they frequently drew students' attention.

The notion of a population in a bag and a sample as a collection of individuals on which measurements are taken, however, does not give students a sense of the population distribution of the variable of interest. A retrospective analysis of the 2009 student data alerted us to the fact that the students had an impoverished understanding of and language for distributional shape and underlying plausible shapes of population distributions (Pfannkuch, Arnold, & Wild, 2011). We regarded the situation as problematic, since in practice contextual knowledge and statistical experience is used to conjecture an expected population distributional shape and if the sample distribution is at variance with the expected shape then further investigation of the data is warranted. Furthermore, we considered that students' statistical knowledge should include images of population distributions of everyday situations such as height and reaction times. The question was raised about how students could build the contextual knowledge necessary for thinking about population distributions and consequently sample distributions. The breakthrough occurred when we realized that *the story in the data was often in its distributional shape*; that shape is a key foundational concept. In a conversation with Cliff Konold in 2009 it emerged that there was little research in this area but Konold had realized from his own research, and Bakker (2004b) also, that students needed assistance in *seeing shape*. Closely allied with such a problem is the well-documented difficulty for students to conceptualize and reason from an aggregate perspective of data (Konold, Higgins, Russel, & Khalil, 2004). Consequently we devised new learning trajectories to build students' conceptions, contextual and statistical knowledge, and the language for engaging with shape. One learning task involved students matching data plots to context, while in another students were given the context and asked to draw the shape of the plot. Our research findings on conceptualizing shape are currently being reported (see Arnold and Pfannkuch, 2012).

To understand some of our research in this area, consider the plot in Figure 3. Before our teaching intervention many students sketched the shape of the plot by drawing an outline or "skyline" and described in detail how the plot went up and down. After the intervention consider the responses from interviews of two students aged 14 who now sketched curves over the plot. One student thought the underlying population shape would be normal with perhaps a right skew. On being questioned further she said she expected this shape based on her general knowledge but this particular sample just happened to be bimodal. The other student believed the underlying population shape was bimodal because of fit and unfit people. Both these students seem to be beginning to realise that shape is closely aligned with unlocking the story in the data, the contextual part, with the first student beginning to consider also the notion of sampling variability, the statistical part.

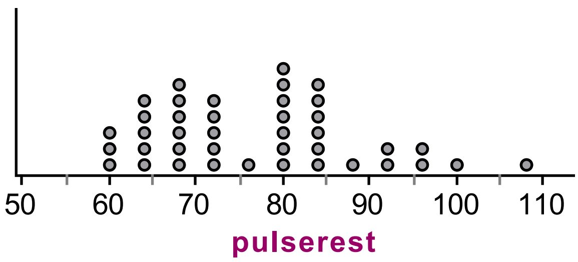


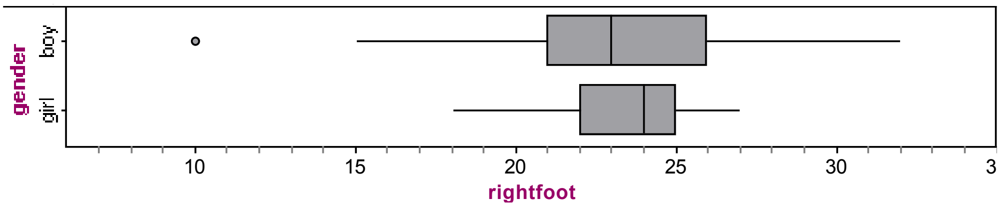
Figure 3. Resting pulse rate of a sample of 40 Year 13 NZ students

Sampling variability

A third conceptual foundation was building students' concepts about sampling variability. The breakthrough occurred when Chris Wild recognized the potential to develop dynamic animations to visualize sampling variability when he was listening to how Pfannkuch (2008) conducted a small-scale teaching intervention on developing students' notions of sampling variability. She explained how plots of random samples were projected on to a whiteboard and each time the sample median in the case of quantitative data or the sample percentage in the case of qualitative data was recorded manually with a horizontal line. In this way a sampling-variation band was built up and students started to become aware of sampling variability including the sample size effect. By taking this idea and realizing that the visualization of sampling variability was more effective when connected directly with the plot rather than increasing cognitive load by referring to a sampling distribution of a statistic, Wild developed a number of dynamic visualizations based on the research literature to assist learning. In order for students to fully conceptualize sampling variability related hands-on activities were developed with attention on developing verbalizations, and imagery, which included gestures (see Arnold et al., 2011).

With regard to developing students' sampling variability reasoning, we believe, that our learning trajectories are assisting students' awareness when they are confronted with static sample distribution plots. For example, we will describe one 14 year-old student's reasoning before and after the teaching intervention. In Figure 4 is a question that was in the pre- and post-test.

Emma is interested in comparing the right foot lengths (in cm) of Year 8 NZ boys and girls. Emma takes a *random sample* of 30 Year 8 NZ boys and a *random sample* of 30 Year 8 NZ girls. She can find out their right foot lengths without shoes on in cm. She plots her graph correctly.



Emma looks at her graph and claims that the right foot lengths of Year 8 NZ girls tend to be bigger than the right foot lengths of Year 8 NZ boys. Would you make the same claim as Emma? Why?

Figure 4. Pre- and post-test question

In the pre-test the student gave the following written response:

> No I would not make the same claim as Emma because the Year 8 NZ boys right foot lengths are spread out across the graph whereas the Year 8 NZ girls right foot lengths are found close together at the place where a normal bell curve would be found.

Note that she is not making the same call as Emma and makes her judgment on the spread of the data. Also some prior knowledge about how the data would be distributed in a normal curve is used, which she said she had learnt in science. In her interview she was asked about drawing another random sample, to sketch what the plots might look like, and whether she would make the same call. Intuitively she knew that another sample would give different plots. She sketched them with a focus on the spread and said:

> I think that, well the overall results would still be the same but I've just spread out the girls more across the graph and made the boys a bit more close to one point on the graph.

In the post-test her written response was:

> No I would not make the same claim as Emma. I would not be prepared to make this claim because on Emma's box plots both the medians are in the overlap. This makes it hard to make an accurate claim because I know that another random sample could easily show the medians the other way round.

Note that her focus is now on the medians and she is aware of the variation in them due to sampling and that she cannot make a call. To check out her imagery of a sampling-variation band for the medians the interviewer asked her to use her hands as the box plots to show her how the plots might change if she took many random samples, which she demonstrated in a similar manner to the dynamic visualizations and said:

> Okay so maybe they would go like this and then maybe like this and the next one would be like this again and then maybe the next one would be like this. There's not much difference but …

Hence from our research data we believe that this student and many others are gaining a sense of sampling variability (see Pfannkuch, Arnold, and Wild, 2012 for a fuller discussion).

CONCEPTUAL FOUNDATIONS: FORMAL STATISTICAL INFERENCE

Statistics students at Year 13 and first-year university levels are introduced to formal statistical inference. Since statistical practice is rapidly changing to simulation methods we developed learning trajectories using the bootstrap method for sample-to-population inference and the randomization method for causal inference. A key idea we use with these methods is to mimic the data production process (Hesterberg, 2006). We again especially developed dynamic visual inference tools that can be used for learning and analysis (see VIT –visual inference tools – http://www.stat.auckland.ac.nz/~wild/VIT). In this section we will briefly highlight one conceptual issue among many issues that arose for each method based on data from a pilot study in 2011.

Bootstrap method

From previous work on sampling variability, our learning trajectory incorporated the prior visual imagery of a sampling-variation band, which transformed into a re-sampling distribution of a statistic so that the confidence interval was quantified (Fig. 5). All the students in the pilot study seemed to be able to articulate how the bootstrap method worked to compute a confidence interval and to explain all the components of the dynamic visualizations for the bootstrap. When faced with a word-only question in the post-test about confidence intervals, however, the concept and image of a re-sampling distribution of a statistic seemed to slip from their minds. We then appreciated that we did not pay enough attention to a key conceptual transition. In primary school students view data displays as distributions of data while at secondary school our conceptual pathway led students to view them as distributions of sample data from which they make inferences about populations. With the bootstrap method students needed a new conception of a familiar distribution as a distribution of a statistic. We realized our verbalizations, learning trajectories, and resources were remiss in addressing such a key conceptual transition (see Parsonage, Pfannkuch, Wild and Aloisio, 2012 for more detail).

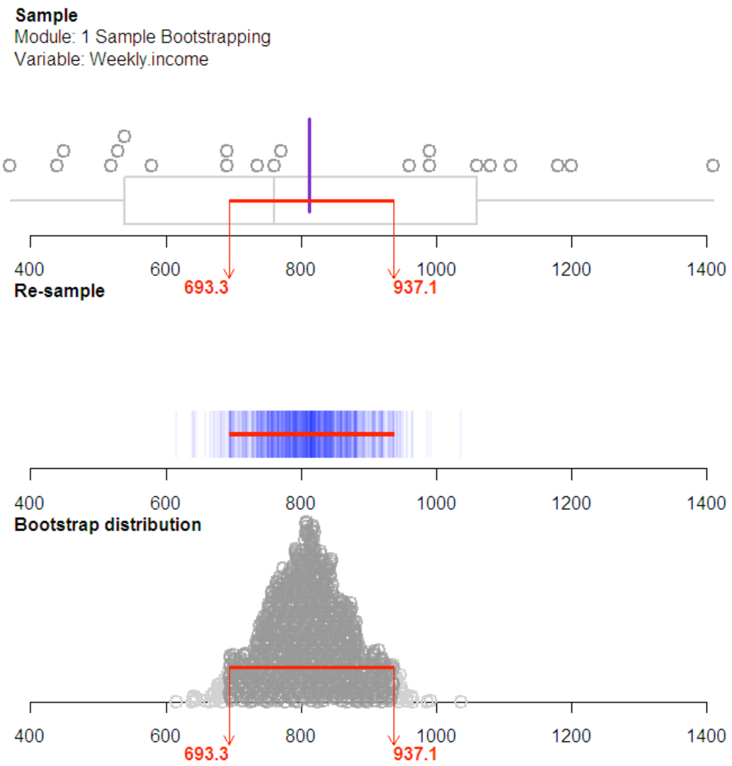


Figure 5. Graphics panel of software for performing bootstrap process

Randomization method

Since we worked on the principle for beginning formal inference that the method must mimic the data production process, students were introduced into a new world of experiments and the very new idea of *causal* inference via the randomization method. For example, in comparative experiments volunteer participants are randomly allocated to one of two groups, the treatment or the control. The experiment is then conducted and the data are plotted. One question for us was how to explain the observed difference in the two groups. We decided to use "the treatment is effective" or "chance is acting alone" as the two possible explanations for the observed difference (see Pfannkuch et al., 2011 for full discussion on many other issues with respect to verbalizations and language). In the pilot study some students reported difficulty in understanding the concept of "chance alone" (see Budgett and Pfannkuch, 2012, for more detail on findings from the study). Furthermore, the concept that the "treatment is effective" comprises the two components of chance and treatment. The problem of how we might explicate such concepts started to be resolved when we developed dynamic visualizations to show, for example, the observed difference in weights, which could be obtained through random allocation to two groups. Such difficult concepts, however, need more thinking and debate.

CONCLUSION

Our research is concentrated on improving the *quality* of the content of the statistics that is taught to students. Overall, we believe, that our attempt at providing students with learning experiences to develop their inferential thinking seems to be improving their understanding of inference at a conceptual level (Arnold et al., 2011; Pfannkuch et al., 2012). In terms of curriculum-time constraints we think that by automating graphics students can spend more time on learning to interrogate and reason from data and less time on learning how to construct plots. Ideally we envisage students learning statistics within an EDA environment, building inferential concepts, and being enculturated into a statistical way of thinking.

In this paper we discussed some of the conceptual foundations necessary for statistical inference that we have wrestled with and debated. We also highlighted some pivotal moments in our research that led us to reconsider some conceptual foundations. Our journey into laying foundations for statistical inference involved us in exploring and thinking about our own conceptual understandings at a very deep level. Through explicating concepts, which were previously implicit, we have come to appreciate the complex nature of inferential thinking in statistics. It has also made us more aware of the many gaps in curricula and textbooks such as no attention to shape of distributions, apart from naming the shapes, the imprecise use of language, and the mixing up of descriptive and inferential thoughts. To reveal the complexity of inference to students requires careful attention to learning trajectories within and across year levels. Our learning trajectories incorporated a coherent set of experiences and ideas of concepts such as sample, population, and sampling variability through interconnecting hands-on simulations, dynamic visualizations, verbalizations, gestures, visual imagery, language, and discourse.

To build this conceptual pathway for laying foundations for inference we started with a desired learning goal in mind that students would be able to make formal statistical inferences for observational studies and experiments. Working from a problematic situation for New Zealand teachers with the comparison of box plots we theorized a pathway and devised learning activities and software. As our journey progressed with discussion, implementation, and modification through the period 2008 to 2012, we learnt a great deal. The question we now ask ourselves is whether the pathway would look the same with what we now know. The answer is probably not. We believe that there could be other pathways for building students' inferential reasoning such as developing paths for sampling variability using both quantitative and qualitative data and for inferences from observational studies and experiments. Or perhaps the randomisation method path could be used for all types of data and study (e.g., Holcomb, Chance, Rossman, Tietjen, & Cobb, 2010) although this does violate the Hesterberg principle of mimicking the data production process.

At the moment, however, there is too big a gap between statistical practice and statistics education (Cobb, 2007; Efron, 2000). To not try to advance the quality of statistics learnt so that it is more closely aligned to practice through laying foundations for inference at secondary school and incorporating methods such as bootstrapping and randomization at the upper levels, is to leave it, paraphrasing Efron's (2000, p. 1295) words "stuck in the 1950s." We hope that our work will contribute to a paradigm shift in what is learnt in statistics.

Presenting statistics at the secondary school level without attention to inferential conceptual foundations is no longer viable.

Acknowledgement

This research is partly funded by a grant from the Teaching and Learning Research Initiative (www.tlri.org.nz).

References

Arnold, P., & Pfannkuch. M. (2012). The language of shape. *In 12th International Congress of Mathematics Education Proceedings*, *Seoul, Korea.*

Arnold, P., Pfannkuch, M., Wild, C.J., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: From hands on to "movies." *Journal of Statistics Education*, 19(2). http://www.amstat.org/publications/jse/v19n2/pfannkuch.pdf

Bakker, A. (2004a). *Design research in statistics education: on symbolizing and computer tools*. Utrecht, The Netherlands: CD-β Press, Center for Science and Mathematics Education.

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

Biehler, R. (1997). Students' difficulties in practicing computer-supported data analysis: Some hypothetical generalizations from results of two exploratory studies. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 169-190). Voorburg, The Netherlands: International Statistical Institute. Online: http://www.stat.auckland.ac.nz/~iase/publications.php

Biehler, R. (2011). *Five questions on curricular issues concerning the stepwise development of reasoning from samples*. Presentation at the seventh International Forum on Statistical Reasoning, Thinking and Literacy, 17-23 July, 2011, Texel, The Netherlands.

Budgett, S., Pfannkuch, M., Regan, M., & Wild, C.J. (2012). *Dynamic visualizations for inference.* Paper presented at the International Association of Statistcial Education Roundtable Conference July 2012, Cebu, Philippines.

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi and J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-324). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Cobb, G. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education, 1* (1), 1-15. Online: http://escholarship.org/uc/item/6hb3k0nz

Cobb. P., & Gravemeijer, K. (2008). Experimenting to support and understand learning processes. In A. Kelly, R. Lesh, & K. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching* (pp. 68-95). New York, NY: Routledge.

Efron, B. (2000). The bootstrap and modern statistics. *Journal of the American Statistics Association*, 95(452), 1293-1296.

GAISE K-12 Report (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: a pre-k-12 curriculum framework*. Alexandria, VA: American Statistical Association. Online: http://www.amstat.org/education/gaise

Hesterberg, T. (2006). Bootstrapping students' understanding of statistical concepts. In G. Burrill (Ed.), *Thinking and reasoning with data and chance. Sixty-eighth National Council of Teachers of Mathematics Yearbook* (pp. 391-416). Reston, VA: NCTM.

Hjalmarson, M., & Lesh, R. (2008). Engineering and design research: Intersections for education research and design. In A. Kelly, R. Lesh, & K. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching* (pp. 96-110). New York, NY: Routledge.

Holcomb, J., Chance, B., Rossman, A., Tietjen, E., & Cobb, G. (2010). Introducing concepts of statistical inference via randomization tests. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics,* Ljubljana, Slovenia. Voorburg, The Netherlands: International Statsitical Institute. Online: www.stat.auckland.ac.nz/~iase/publications.php.

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education*, 2(1), http://repositories.cdlib.org/uclastat/cts/tise/vol2/iss1/art1/

Konold, C., Higgins, T., Russel, S., & Khalil, K. (2004). Data seen through different lenses. Unpublished paper. Online: http://www.srri.umass.edu/publications/konold-2004dst

Parsonage, R., Pfannkuch, M., Wild, C.J. & Aloisio, K. (2012). Bootstrapping confidence intervals. *In 12th International Congress of Mathematics Education Proceedings*, *Seoul, Korea.*

Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal*, 5(2), http://www.stat.auckland.ac.nz/serj.

Pfannkuch, M. (2008). Building sampling concepts for statistical inference: A case study. *In 11th International Congress of Mathematics Education Proceedings*, Monterrey, Mexico. Online: http://tsg.icme11.org/tsg/show/15

Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, 13(1 &2), 27-46.

Pfannkuch, M., Arnold, P., & Wild, C.J. (2011). *Statistics: It's reasoning not calculating.* Summary research report for New Zealand Council of Educational Research on Building students' inferential reasoning: Levels 5 and 6. Online: *www.tlri.org.nz*.

Pfannkuch, M., Arnold, P., & Wild, C.J. (2012). *What I see is not quite the way it really is: Students' emergent reasoning about sampling variability*. Unpublished paper.

Pfannkuch, M., Regan, M., Wild, C.J., Budgett, S., Forbes, S., Harraway, J., Parsonage, R. (2011). Inference and the introductory statistics course. *International Journal of Mathematical Education in Science and Technology,* 42(7), 903-913.

Pfannkuch, M., Regan, M., Wild, C.J., & Horton, N. (2010). Telling data stories: essential dialogues for comparative reasoning. *Journal of Statistics Education*, 18(1), http://www.amstat.org/publications/jse/v18n1/pfannkuch.pdf

Pratt, D., & Ainley, J. (2008). Introducing the special issue on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 3-4, http://www.stat.auckland.ac.nz/serj

Schwartz, D., Chang, J., & Martin, L. (2008). Instrumentation and innovation in design experiments. In A. Kelly, R. Lesh & J. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching* (pp. 45-67). New York: Routledge.

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