ENHANCING STUDENTS' INFERENTIAL REASONING: FROM HANDS ON TO "MOVIE SNAPSHOTS"

Pip Arnold[1] and Maxine Pfannkuch[2]
[1]Cognition Education and Department of Statistics, The University of Auckland, New Zealand
[2]Department of Statistics, The University of Auckland, New Zealand
PArnold@cognition.co.nz

Computer simulations and animations for developing statistical conceptions are often not understood by beginners. Hands-on physical simulations that morph into computer simulation images are teaching approaches that can build students' concepts. In this paper we describe an instructional sequence, from hands on to "movie snapshots", which was trialed in a Grade 9 class. The instruction focused on developing students' sampling variability concepts and on making inferences about populations from samples. Responses from three students' interviews and two assessment items are explored, including the images they worked with when they reasoned and made a call from box plots. The findings suggest that students can use sampling variability ideas to support their inferential reasoning.

INTRODUCTION

Zieffler, Garfield, delMas and Reading (2008, p. 44) characterize informal inferential reasoning as "the way in which students use their informal statistical knowledge to make arguments to support inferences about unknown populations based on observed samples." Such inferences, however, are based on arguments that are particular to a class or student's development and depend on what is acceptable to a teacher. In New Zealand a problem arose when the new curriculum (Ministry of Education, 2007) and subsequent national assessment required students to make informal inferences about populations from samples in Grade 10. In such a situation the arguments to support inferences can no longer be an agreement between a teacher and her class but must be based on criteria that are transparent to all teachers. Therefore Wild, Pfannkuch, Regan and Horton (2009) proposed a developmental pathway for comparative situations from Grade 9 to 12 for justifying how to "make a call" or make a decision about whether condition A tends to have bigger values than condition B. Enabling students to "make a call" depends on building their understanding of a network of underlying interrelated concepts. To build students' conceptions of sample, population, sampling variability, and the sample size effect Wild et al. designed and produced computer "movie snapshots." Aware that learners do not often understand computer simulations, a research project team devised a teaching sequence of physical simulations that would morph into the "movie snapshot" images. This paper reports on an initial trial of this teaching sequence in a Grade 9 class.

LITERATURE REVIEW

According to Franklin et al. (2007, p. 6) "statistical problem solving and decision making depend on understanding, explaining, and quantifying the variability in data." Despite the importance of considering variation in statistics, researchers have only in the last decade begun to document students' conceptions of variability. Since there are many types and sources of variability, there is only some research on sampling variability and it is mainly limited to chance settings. In chance situations Shaughnessy (2007, p. 982) notes that for students to consider variability among samples they need to develop distributional reasoning. That is, students need to grow beyond their propensity to focus on the expected value and develop intuitions "for a reasonable amount of variation around an expected value." He also reports that there is some evidence students' distributional reasoning can be improved if they conduct hands on simulations. Applying these findings to the statistics setting it would seem that drawing out random samples of a fixed size by hand from a population and noting the sample median might start to develop students' intuitions about sampling variability. But at some stage hands on simulations become laborious for students and carefully designed technology tools that attend to students' current knowledge of statistics can help to bridge them from "naïve conceptions to richer, more powerful understandings of statistical concepts" (Shaughnessy, 2007, p. 995).

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

For the development of new cognitive technology-based tools delMas (1997) provides guidelines for software and teaching. Some of his suggestions for fostering understanding in statistics are: recognize the roles encoding, prediction, and feedback play in the development of understanding; integrate physical activities with computer simulations; emphasize the interplay among verbal, pictorial, and conceptual representations when using instructional activities designed around simulations; use representations that are familiar to students; and make it clear to students which features in the software environment are important. Wild at al. (2009) also emphasize the importance of hands on activities, visual imagery and the discourse of the teacher to bring meaning to pictorial representations. Their tools present the concept of sampling variability as an integral part of the familiar box plot representation. Color is used to focus students' attention on the properties and structure of sampling variability in order for students to encapsulate it as part of the representation. The "movie snapshots" simulations are dynamic representations (http://www.censusatschool.org.nz/2009/informal-inference/WPRH/). Greer's (2009, p. 701) comments about an earlier version of their tools are:

> the sample values are not represented numerically, which may well be significant since numerical values could cue computation, whereas the visual counterpart invites comprehension. Consider also how the process unfolds in time, and leaves a history … that could stimulate episodic memory of the process that gave rise to it.

He notes that provided classrooms promote sense making and flexibility of thought students can draw on prior episodic memories that assist them to solve problems in new settings. The technological tools and the methods for "making a call" devised by Wild et al. (2009) are novel. They incorporate facets suggested by other researchers and build on the chance research of Shaughnessy (2007), but only further research will tell whether the sequence of instruction from hands on to the "movie snapshots' will enhance students' inferential reasoning.

Informal inferential reasoning in statistics has recently become a focus of research (Pratt & Ainley, 2008). But problems arose when students were making inferences or claims in comparison situations. Pratt et al. (2008) noted that students and teachers being studied did not know whether they were reasoning about the data as if it were the whole population or about an underlying population from which the data were a sample. At the International Forum for Statistical Reasoning Thinking and Literacy in 2007 statistics education researchers collectively agreed that students from middle school onwards should be generalizing beyond the data in hand to a population or process. Konold and Kazak (2008) argue that the recognition by the Forum that students need deeper understandings of inference had the potential to result in an acceptance that chance or sampling behavior must be addressed. This research is specifically focused on facilitating students to make a decision in comparison situations, to draw inferences about populations from samples, and to take sampling variability into account.

The research question for this particular paper is: In comparison situations how do three Grade 9 students support inferences about unknown populations from samples?

METHOD

The research method follows design research principles (Roth, 2005) for a teaching experiment in a classroom. In the preparation and design stage a research project team, consisting of teachers, statisticians, and researchers, worked together to develop the teaching and learning materials to use in the teaching experiment. The activities were evaluated and critiqued in a series of six meetings with follow up discussions continuing over many days. In the implementation stage both a teacher and a researcher from the project team trialed the teaching sequence in the teacher's Grade 9 class. The teacher and researcher were involved in a reflective discussion following each lesson, which allowed adjustments to be made to the hypothetical learning trajectory. The 26 students in the class were average in ability and were from a mid-size (1300), multicultural, middle socio-economic inner city girls' secondary school. Students were given a pre and post-test and the lessons were videotaped. A group of four girls were observed specifically and three of these four girls had pre and post-interviews about their responses to their tests.

The retrospective analysis for this paper focuses on students' understanding of three lessons (lessons 12 to 14 in a 15-lesson teaching sequence) where students learn about sampling variability for samples of size 30 and what “calls” or claims they can make about populations from samples when comparing box plots. Prior work in the teaching sequence focused on: posing different types of questions; describing summary and comparative distributions; learning about taking samples from populations; constructing box plots from dot plots; and conducting investigations involving the comparison of dot plots and box plots. To understand the type of learning experiences the students had in lessons 12 to 14, they will now be described.

The lessons had three main phases. The first phase was the preparation phase. In a bag were 616 students (datacards) from a fictitious college, but the data recorded on the cards for the students came from the CensusAtSchool 2009 database. Each pair of students had a bag from which they selected samples of size 30 to explore the following two questions: Do the heights of boys at Karekare College tend to be greater than the heights of girls at Karekare College? Do Karekare College students who walk to school tend to get there faster than Karekare College students who take the bus? The students recorded the “box” part of the information from the box plots onto pre-prepared graph outlines. It was decided to use just the “box” to concentrate students’ attention on comparing shift and overlap and on noticing persistent or inconsistent patterns across different samples. Altogether 14 different samples were taken for each question.

The second phase was looking for patterns across the collection of samples for the two questions. Each group of students was given a copy of all the graphs from the class samples. The students were asked to sort the samples for the heights question, sort the samples for the time to school question, and to describe what they noticed. The teacher stapled the two sets of graphs onto the wall as well, but these were not sorted (Figure 1). During this phase the class collectively came up with the idea that there were two different situations, which became simply situation one and situation two. The two situations are defined by the shift or location of the boxes, the amount of overlap between the two boxes and the placement of the medians relative to the overlap. In the following excerpt the teacher explores the differences between the two situations.

T: So in our first situation we’ve got the boxes they’re all overlapping some of them are going this way and some of them are going the other way. The medians are very close together and the medians are also within the overlap of the boxes. In the second situation how is it different? What’s different about the overlap here? Is there no difference between the overlap on these boxes and these boxes?
S: They’re not overlapped so much.
T: They’re not overlapped so much. No, they’re not. Okay do they all overlap?
S: No.
T: No, so when they do have an overlap they don’t overlap much and otherwise they don’t overlap at all. What can you tell us about the medians in this one?
S: They’re not overlapped.
T: They’re not in the overlap.

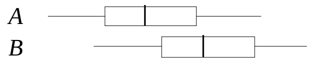
Figure 1. The two situations

Figure 2. Students in class raising hands

The third phase involved using “movie snapshots” to reinforce the message from the two situations. Situation one occurs when the boxes are located in the same place with little or no shift. There is a lot of overlap, and the medians are within the overlap and can swap positions, that is, one median might be higher in one sample, but the other median higher in the next sample. In this situation ideas are consolidated that when there is a large overlap and the medians are within the overlap, the message across many samples is mixed and inconsistent about the pattern back in the populations. For situation two, the boxes are located apart with a large shift. The overlap is small or non-existent, and at least one of the medians is outside the overlap. In this situation the location of the boxes and the position of the medians relative to one another stays consistent across many samples and therefore the message about the pattern back in the populations stays consistent. In this third phase students were shown many more samples using "movie snapshots." This entailed an almost movie-like view of the multiple samples to emphasize either the mixed message (situation one) or consistent message (situation two) about the populations. Students raised either their left or right hand depending on which median was higher (Figure 2). In situation one they were swapping their raised hand constantly, for situation two the same hand remained raised.

The hypothesized next phase, which time did not allow, is to engage the students' attention on the "movie snapshots" that track the history of the variability for these two situations. Students may then fully appreciate and develop the visual imagery for the extent of sampling variability with samples of size 30. We conjecture that in this phase an informal "decision rule" for "making a call" will be realized as in Figure 3. Future learning experiences will include the sample size effect.

Age-14: *the 3/4-1/2 rule*



If the median for one of the samples lies outside the box for the other sample
(e.g. *"more than half of the B group are above three quarters of the A group"*)
make the call that ***B tends to be bigger than A*** back in the populations

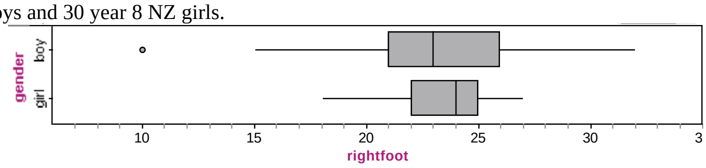
[Restrict to samples sizes of between 20 and 40 in each group]

Figure 3. Goal for "making a call" for Grade 9 (age 14) students (Wild et al., 2009)

RESULTS

This section presents the responses of three students to two questions (Figure 4) from the pre and post-tests, which comprised many items. Question one was in both the pre and post-tests, question two was in the post-test. Responses from the post-test interview are also included.

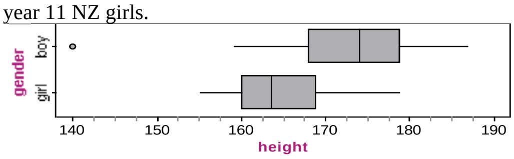
Sample of 30 year 8 NZ boys and 30 year 8 NZ girls.



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.

1. Would you make the same claim as Emma? Why?

Sample of 30 year 11 NZ boys and 30 year 11 NZ girls.



Emma looks at her graph and claims that the heights of Year 11 NZ boys tend to be greater than the heights of Year 11 NZ girls.

2. Would you make the same claim as Emma? Why?

Figure 4. Abbreviated versions of two assessment questions

Even though only three students' responses are described, the analysis of all students responses to the pre and post-tests against an assessment framework indicated a shift was made in their ability to agree or disagree with claims made by another person based on box plots as evidence for the claim. These three students responses are indicative of the shift made by the class.

In the pre test* two of the three students (S1, S2) made the claim for question one based on the size of the box, that is, they compared an incorrect feature to agree or disagree with the claim. The third student (S3) made the claim by referring to the center or middle. (*Graph in the pre test was printed without gender labels and students assumed upper box plot was girls.)

Question 1:

S1: No I would not make the same claim as Emma because the Year 8 NZ girl's right foot lengths are spread out across the graph whereas the Year 8 NZ boys right foot lengths are found close together at a place where a normal bell curve would be found.

S2: Yes, because the box is bigger.

S3: No because the graph doesn't agree with Emma. It tells that Year 8 NZ boys have bigger average length with their right feet.

For their post-test responses shown below, note the following. S1 moved to agreeing or disagreeing with the claim by referring to the shift and/or overlap, the position of the medians relative to the overlap, and acknowledging what another sample might look like. S2 moved to agreeing or disagreeing with the claim by referring to the center or middle and in question two indicated location of one median relative to the other. S3 moved to agreeing or disagreeing with the claim by referring to the shift and/or overlap and position of the medians relative to the overlap.
Question 1:

S1: 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.

S2: Yes the boys box is bigger but the girls median seems to be higher than the boys.

S3: No because in the graph there is a HUGE overlap showing hardly any difference, the medians are also inside the overlap and the median of the boys is actually further left. I would not make a claim at all.

Question 2:

S1: Yes I would make the same claim as Emma because on the box plots above there isn't much overlap and both the medians are outside the overlap. I know that if another random sample was taken it most probably would show the medians the same way. Therefore I would make the same claim as Emma.

S2: Yes because the boys median is in a higher area than the girls.

S3: Yes because there is a tiny overlap and each medians are outside at the overlap the boys are further right than girls.

From questioning the students in the post-test interviews all three students appear to have two different images in their mind to represent the two different situations. Below are examples of the images they have for situation one and situation two.

S1 interview response to question 1:

I: So if you think that if this is a situation one, if just using your hands like the boxes, if I was to do the repeated sampling from the population, what would the graphs look like, what sort of image do you have of what would happen with the graphs?

S: 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 (Figure 5)

I: There'd be a tendency for them to move slightly backwards and forwards?

S: Yeah.

I: Okay and will the girls' median right foot length always be higher?

S: No.

I: And would you expect the boxes to overlap?

S: Yes.

S3 interview response to question 2:f

I: So do you think you could again with your hands just show what would happen if we repeated the sample from the population?

S: I think it would stay like that or maybe go a little closer. (Figure 6)

I: But you'd always have the boys further to the right? So will the boys' median height always be higher?

S: Yeah.

I: And would you expect the boxes to overlap in another sample?

S: Maybe just a little. But yeah.

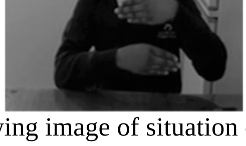


Figure 5. Showing image of situation one with hands

Figure 6. Showing image of situation two with hands

The research question for this paper is interested in how students support inferences about unknown populations from samples. From the student test transcripts S1 and S3 show that they have a reasonable understanding of the 3/4–1/2 rule (Figure 3), whereby they make the claim based on the shift or location of the boxes, the amount of overlap between the two boxes, and the relative location of the medians to the overlap. S1 has a slightly deeper understanding than S3 as she recognizes that another sample may give a different (situation one) or a similar (situation two) result. Neither S1 nor S3 were able to take this description to the next level in the assessment framework and embed it within the context of either year 8 right foot lengths or year 11 heights. S2 progressed from making a call on an inappropriate measure to making a call using the medians. All three students when interviewed showed an understanding of the difference between situation one and situation two. They were able to demonstrate how the two situations appear using their hands. Their demonstration links strongly to the “movie snapshots”, which give a sense of constant changing of location of median within the overlap of the boxes (Figure 5) or boxes staying consistent relative to one another with little or no overlap with a jiggle (Figure 6).

CONCLUSION

Our initial trial of a teaching sequence, which aimed to build Grade 9 students’ concepts of sampling variability and on making inferences about populations from samples, suggests these ideas are well within the grasp of New Zealand students. By addressing sampling behavior (Shaughnessy, 2007; Konold & Kazak, 2008) students can begin to deepen their understanding of inference. Hands-on activities that morph into “movie snapshots” (Wild et al., 2009) seem to provide a rich learning environment but what is just as important is the underlying principles that foster such understanding and comprehension (delMas, 1997). We conjecture that some of these principles are the focus on the visual rather than the numerical (Greer, 2009), integrating sampling variability ideas into dynamic box plot imagery (Wild et al., 2009), and drawing students’ attention to the properties of sampling variability via an interplay among visual, verbal, physical, and use of gesture for representations. Suggested recommendations for further research include working with students to embed their inferential reasoning within the context given and to support them to make a call on a single sample rather than repeated samples. These ideas will be explored further in the next iteration of the research project.

ACKNOWLEDGEMENT

This work is supported by a grant from the Teaching and Learning Research Initiative (www.tlri.org.nz). Thanks to Chris Wild for his helpful comments on this paper.

REFERENCES

delMas, R. (1997). A framework for the evaluation of software for teaching statistical concepts. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 75-90). Voorburg, The Netherlands: International Statistical Institute.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: a pre-k-12 curriculum framework*. Alexandria, VA: American Statistical Association.

Greer, B. (2009). Representational flexibility and mathematical expertise. *ZDM Mathematics Education*, 41, 697-702.

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/

Ministry of Education (2007). *The New Zealand curriculum*. Wellington, New Zealand: Learning Media Limited.

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

Pratt, D., Johnston-Wilder, P., Ainley, J., & Mason, J. (2008). Local and global thinking in statistical inference. *Statistics Education Research Journal*, *7*(2), 107-129.

Roth, W. (2005). *Doing qualitative research: praxis of method*. Rotterdam: Sense Publishers.

Shaughnessy, M. (2007). Research on statistics learning and reasoning. In F. Lester (Ed.), *Second handbook of research on the teaching and learning of mathematics*, (Vol. 2, pp. 957-1009). Charlotte, NC: Information Age Publishers.

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2009). *Precursor statistical inferences*. Unpublished paper.

Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 40-58.