STUDENTS' VISUAL REASONING AND THE RANDOMIZATION TEST

Stephanie Budgett and Chris J. Wild
The University of Auckland, New Zealand
s.budgett@auckland.ac.nz

The traditional approach to teaching statistical inference limits students' reasoning to mathematical manipulations. Cobb (2007) advocated the use of simulation methods for introductory statistical inference. In this paper, the dynamic visual simulations we use for introductory statistics for experiment-to-causation inference are briefly described. Using data from recent research, two students' reasoning is analysed; one in response to being asked to use visual inference tools to analyse data from a randomized experiment, the other being asked to visualize the simulation process. Our findings suggest that the dynamic visualizations are becoming part of these students' cognitive processes for understanding experiment-to-causation inference. Remaining issues that arose in students' reasoning will be discussed.

INTRODUCTION

Statistical inference has long been recognized as a problematic area for students of introductory statistics. Abstract mathematical concepts such as the normal distribution, the Central Limit Theorem, and sampling distribution of estimators have acted as obstacles in the path of student understanding while the unfamiliar logic associated with statistical argumentation further inhibits student understanding. The result is that the majority of students emerge from traditional introductory courses lacking a sound conceptual understanding of the statistical inference process and are often prone to misconceptions (Sotos, Vanhoof, Van den Noortgate, & Onghena, 2007). Thus many statistics education researchers believe that the way forward is to change the way in which the statistical reasoning process is taught (Cobb, 2007; Rossman, 2008). Motivated by Cobb's (2007) challenge to place the logic of inference at the heart of the introductory statistics curriculum, and aligned with recent moves in statistical practice (Hesterberg, Moore, Monaghan, Clipson, & Epstein, 2009), we introduced the randomization and bootstrap methods into introductory statistics. Part of the appeal in incorporating these methods at the introductory level is that they lend themselves to visual processes that have the potential to clarify some of the more abstract underpinning concepts of statistical inference that have remained elusive in the traditional courses. The focus of this paper is to explore the visual reasoning employed by beginning students when they are given an exploratory task and access to purpose-built *Visual Inference Tools* (VIT) software (www.stat.auckland.ac.nz/~wild/VIT).

LITERATURE REVIEW

For many years, verbal learning was the focus of educational research, with scant attention paid to visual learning. Despite research on visual and mental imagery being commonplace in the field of psychology as long ago as the 19th century, it wasn't until the 1980s that investigation of the role of visual representations in the field of mathematics education gained momentum (Presmeg 2006). Since then, owing to the proliferation of data and the technological advances that have occurred over more recent times, both static and dynamic visual representations have gained traction in the field of mathematics and statistics education with educators combining both verbal and visual information in their teaching. According to Clark and Paivio (1991), the use of visual representations offers the opportunity for generation of mental images which can have a positive impact on learning. Unlike verbal processes, which deal with information sequentially, mental, or nonverbal, images have the ability to encode information simultaneously, thus allowing illumination of more complex ideas. Clark and Paivio also note that there exists the ability to manipulate such mental images and to note the resultant effects, a property that is not possible with verbal representations. Visual and graphical representations have the potential to clarify previously inaccessible concepts by providing new ways of engaging with these often abstract ideas (Arcavi, 2003). As Ware (2008) suggests, harnessing human's "active vision" by designing representations that allow us to form mental images, allows engagement in problem solving using a process of visual reasoning. The role of sensory cognition is also important in the learning process,

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

particularly for novices (Radford, 2009). This lends weight to the argument of other researchers, who suggest that tactile experiences such as hands-on activities form an integral stage in any learning trajectory designed to facilitate novice students' statistical thinking (Rossman, 2008).

While acknowledging the advantages in employing visual representations in education, it is important to anticipate potential problems. It is vital to reflect on the many and varied ways in which learners might interpret visualizations, both static and dynamic, and to acknowledge that these differences may result in novices failing to perceive the visualizations in the manner intended (Arcavi, 2003). Cognitive load theory is a learning theory that requires attention when developing new visual representations (Sweller, 1988). This theory suggests that consideration of learners' working memory is imperative, and since learners' ability to process new information is limited, the way in which this new information is presented is crucial. Research has shown that despite an initial belief that dynamic visualizations had great potential for developing understanding of abstract concepts, they did not appear to be any more effective than static visualizations (Hegarty, 2004). Drawing on cognitive load theory, which identifies germane load as the result of mental activity that is directly relevant to learning and that contributes to the construction and automation of knowledge in long-term memory, visualizations used in this research were designed to reduce extraneous factors, together with providing the optimal conditions for student learning.

Purpose-built software, incorporating dynamic visual imagery, was designed to reveal the processes behind the randomization test and formed a key component of the learning trajectories which were implemented. The visual imagery was closely connected to hands-on activities that formed an important stage in the learning sequence. The VIT module for randomization is composed of three panels: a control panel, a data panel and a graphics panel, with the graphics panel having three linked sections (see Figure 1). The vertical arrangement of the graphics panel was designed so that learners could link the original data in the top section with the results of re-randomizations in the middle section, and finally to the re-randomization distribution in the bottom section. For further details, see Budgett, Pfannkuch, Wild and Regan, 2013.

METHOD

The interviews that form the basis of this paper come from a large collaborative New Zealand-wide research project involving 33 team members and over 2700 students. The majority of these students (2553) were first year university introductory students. Another 198 were final year school students, with the remaining 14 being workplace learners. Depending on the teaching environment, the timetable of the instruction varied. In the university setting, students were introduced to experimental design in one 50-minute lecture and to the randomization test in two 50-minute lectures, in a class of approximately 450 students. In keeping with normal practice for workplace professional development, the three 50-minute lectures on experimental design and the randomization test were delivered in one half-day session. Regardless of the setting, lectures comprised hands-on activities, followed by lecturer demonstrations of the VIT modules. The university students were required to download and use the VIT modules in a graded assignment which was handed in approximately one week after the teaching sequence, while workplace learners were allocated time for using the VIT modules and assignment work the day after their teaching session. Interviews with the university students were carried out approximately one week after the assignment was handed in, while workplace learners were interviewed the day after completion of the assignment. All interviews were video-recorded and transcribed prior to analysis.

Two interviews are reported in this paper. One was carried out with a workplace learner (S1) who had taken one first year university statistics paper several years beforehand, while the other was carried out with a university student (S2) who had taken statistics in his final year at school. Both had experienced a traditional curriculum with exposure to the Central Limit Theorem, confidence intervals and sample-to-population inference. They did not have any previous experience of experimental design or the randomization test. We have chosen to concentrate on these two students since they had the same interviewer and were articulate. The first student, S1, was provided with raw data on decreases in blood pressure for 21 black male volunteers with high blood pressure who were randomly assigned to either a calcium tablet group or a placebo group, and asked whether the added calcium intake reduced blood pressure. This task was adapted from Hesterberg et al. (2009). The second student, S2, was asked to visualize the randomization process.

RESULTS

S1 was asked to provide a commentary on her use of the VIT module in order to explain what she was doing. The second student, S2, was asked what images he had in his head of the randomization VIT. We report on the respective discussions below.

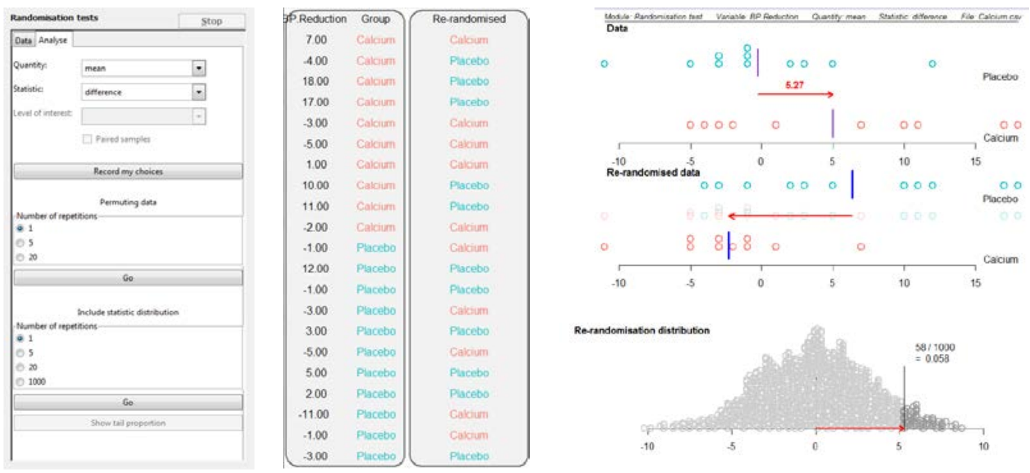
Student 1

S1 explained that in order to decide whether or not calcium reduces blood pressure, she would use randomization. When prompted as to what she meant, she responded by saying:

Randomization is the technique that allows us to look at the differences between two groups and to see if we think this would normally happen by chance or if there is probably some other explanation

Thus S1 appears to know the purpose of the randomization test, recognizing that the requirement is to establish whether or not an observed difference between two groups may be attributed to chance explanations, that is, to *chance acting alone*. After entering the data, S1 then interpreted the observed data display in the top section of the graphics panel (see Figure 1) in the following way:

So it is saying that those in the calcium group have a mean… decrease in blood pressure that is 5.3 points …greater than those in the control group… So what we're going to see is if this 5.3, if that difference observed between the two was likely to happen by chance or not. And to do this we repeat the process, lots and lots of times, by reshuffling the labels on the data. So I'll show you one.



Control panel Data panel Graphics panels

Figure 1. VIT software panels for performing randomization test
(Note that the graphics panels are dynamically linked and appear sequentially)

This statement provides evidence of S1 thinking ahead, anticipating visually what might happen when the blood pressure decreases are randomly reassigned to either the calcium or placebo group. The notion of reshuffling the labels suggests that she is recalling the hands-on activities where this language was used, and in addition she recognizes that this reshuffling needs to happen many times in order to experience the extent of the variability in the differences between group means that might occur under chance alone. She then selected one repetition from the *Permuting data* section of the control panel, resulting in the data points in the top section of the graphics panel moving down into the middle section, combining to form one group and then being randomly split into two groups resulting in the dot plot in the middle section of the graphics panel (Figure 1).

S1: *When it does the relabeling process it makes the new plot of the data points… for the control and for the calcium, and then it recalculates the difference in the mean.*

I: *So what is happening here? Why is that arrow…*

S1: *Pointing in the opposite direction? Well, because it randomly assigns the group to the number. So in this case, some of the high values, which possibly were initially in the calcium group, have been assigned now to the control group.*

In her response to the interviewer's question, S1 demonstrates an appreciation of the way in which the treatment groups are allocated, that is, randomly. We conjecture that the visuals are assisting her reasoning and helping her construct the notion of random assignment through visually being able to see the data values combine to form one group and then being randomly split into a calcium group and a placebo group. They also allow her to experience and reason about the extent of the variability in the difference between the means, represented by the red arrow which dynamically switches back and forth and increases and decreases in magnitude.

S1 then stated: "… so if we do this lots and lots of times it will give us a nice distribution of the difference between the two groups when we randomly assign the condition." Again, S1 is anticipating a visual representation, this time of the distribution of the differences in means between the two groups. We believe that she is harnessing visual reasoning in order to construct one of the abstract underpinning concepts of statistical inference, the notion of a distribution under *chance acting alone*. In referring to repeating the reshuffling process "lots and lots of times", S1 demonstrates the capacity to think in terms of a repeatable process which, over the long run, will produce a stable distribution of outcomes. The VIT module provides images that allow her to visualize and experience the phenomenon of random re-assignment and to appreciate that the resulting differences between the group means form a distribution, thus facilitating comprehension of these two abstract concepts. Such concepts are difficult to grasp with the traditional approach, relying as it does on symbolic reasoning rather than visual reasoning. Also of note is the fact that S1 recognizes that it is a distribution of differences in group means that will be built up over time. Thus she understands that this difference in group means is the statistic of interest, not the individual data points. S1 then proceeded to the *Include statistic distribution* section of the control panel, selected 1000 and gestured to the distribution being built up dynamically in the bottom section of the graphics panel (Figure 1).

That's calculating a whole bunch of new differences of means, and building up this distribution here. What we are really interested in is if our original mean difference of 5.3 was likely to happen by chance…We want to know if 5.3 is a number that we expect to happen if we just randomly relabel it, how many of the times when we relabeled, the difference was greater than that. So I'll just tell it to show us.

Again note that S1 knows that each data point in the re-randomization distribution represents a difference in means, and that the distribution represents an underlying repeatable process which will stabilize over time. Thus the visual representations allow S1 to reason with and to verbally interpret the process of re-randomization and build a conceptual understanding of the distributional aspect of the re-randomized outcomes.

By selecting *Show tail proportion* from the control panel, S1 continues:

5.8% of the time we would get a result of 5.3 or larger by chance. They call this the tail proportion. You could not claim that calcium was any better than the placebo drug that you have given people. [5% guideline given to workplace students]

In response to a question about why she would include all values greater than 5.3 when reading the tail proportion from the re-randomization distribution, S1 stated:

Any difference over here [pointing to values greater than 5.3] is a more extreme difference. It is a better result for the efficacy of your calcium… they would say the calcium is working heaps better than the control because the difference between the two groups is larger.

We conjecture that the visual representation of the tail proportion, incorporating a darker colour shading the area to the right of the observed difference in means (Figure 1), assists S1 in reasoning that differences in group means of greater than 5.3 would provide more evidence against a chance alone explanation.

Towards the end of the interview, S1 was asked what images she had of randomization.

The funny thing is when I first think of randomization I actually think of ripping the labels off and shuffling them and reassigning them. I have to do a few steps in my head to get the distribution, okay to get the new means, to then get the differences to put in the distribution to then get the tail, so that takes a few steps.

In this short excerpt, S1 describes the entire randomization process, from the hands-on activities through to the three dynamically-linked screens of the graphics panel (Figure 1) and thus seems to have incorporated the visualizations into how she thinks and reasons with the randomization test.

Student 2

S2 was asked what images he had of the randomization test. He seemed able to visually reconstruct the randomization test making use of his own scenario, described in the following excerpts.

S2: *I guess the example we'd use is people taking aspirin for heart and people that aren't and so we have the two groups and then we record the data over the trial period and say we see a difference between the results between the control and non-control. We want to kind of see if in fact the aspirin is effective in this experiment. We want to test for chance alone and by testing for chance alone we get rid of the non-control. We just completely disregard what group they're from, mix the two populations together and then I think we take samples. Do we? I'm just trying to think.*

I: *So you've got your aspirin and non-aspirin together.*

S2: *I know the end result is to establish a difference between like say two means or medians and at the end of the process I know we establish a tail proportion, that's what we call it and if the tail proportion is less than 10% [10% guideline given to university students] then we know that chance is probably not acting alone and there's another variable involved.*

As S2 was talking, he gestured the form of a distribution and indicated where he might locate a tail proportion.

I: *So you've got the top screen with observed difference in the means… you've got the distribution in your head. What was the middle screen showing you?*

S2: *I think the middle screen showed the re-allocation of the data and then making new groups.*

S2 was able to reconstruct the process of re-randomization by recalling the visualizations and, through recreating what was happening in the middle screen, making a link between the original observed data and the distribution of the re-randomized differences. We believe that he demonstrated distributional reasoning through visualization since he was able to articulate, through words and gestures, that in order to establish whether or not an outcome was unusual, he needed to look at the tail proportion.

Responding to a question about making use of mental imagery in order to recreate the randomization test, he said: "That definitely helps, especially with the tail proportion. I just remember that arrow. I remember like key numbers that just come out in red. Red's a great colour yeah [it means] listen, watch." S2's statement supports our belief that colour is an important consideration when designing visualizations. The use of a vivid colour such as red, together with the vertical motion of the arrow from the observed data to the re-randomization distribution, provides a powerful way of conveying the idea of a tail proportion by linking the concrete (the observed data) to the abstract (the notion of a distribution). From his overall interview we believe that S2 was able to mentally reconstruct the randomization procedure by recreating the visual imagery he had experienced with the VIT modules. We believe that visual representations, including the red arrow moving down vertically from the plot of the observed data in the top section of the graphics panel to the re-randomization distribution in the bottom section, enabled S2 to reconstruct key concepts of statistical inference.

Interpretation of the tail proportion

While we believe that the dynamic visualizations have assisted these two students in developing a better conceptual understanding of statistical inference, the interpretation of the tail proportion is still problematic. In response to being asked what conclusion she would make if the tail proportion was small, S1 stated: "You could conclude that it is quite likely that the calcium is actually affecting blood pressure more than the control. You could conclude that calcium is actually lowering the blood pressure." For his aspirin example, S2 concluded: "Chance is probably not acting alone and the aspirin is effective." Neither student articulated the idea that chance is always acting, albeit alongside treatment in the case of a small tail proportion. In terms of interpreting a large tail proportion, typically more prone to misconceptions, S1 stated: "You could not claim that calcium was any better than the placebo drug that you have given people", while S2 commented: "We still don't know if chance is acting alone but it probably is I think." Neither interpretation is entirely correct. Both S1 and S2, along with many other students, failed to provide a full argumentation of what a large tail proportion represents. A better interpretation would be "Nothing can be concluded. Chance may be acting alone or perhaps other factors such as the calcium could be acting alongside chance."

CONCLUSION

It has been suggested that visual representations have the potential to reveal previously inaccessible concepts (Arcavi, 2003) and may facilitate more engagement in the problem solving process (Ware, 2008). In addition, sensory experiences designed to complement visual imagery may contribute to a richer conceptual understanding (Radford, 2009). With this in mind, the dynamic visualization software developed to convey the underlying processes of the randomization test aimed to reduce extraneous cognitive demands by allowing students to view the entire re-randomization process as it developed. Hands-on activities were designed to provide direct experience of re-randomizing to two groups and were closely linked to simulation of the identical process using the VIT software. Through an analysis of S1's interaction with the randomization VIT module and accompanying commentary, we conjecture that visual reasoning has become an integral part of her cognitive processes. Relating the randomization process to previous experiences with hands-on activities appears to have contributed to her ability to link the various steps of the randomization procedure. Interpreting the commentary of S2, we believe that his prior experience of the dynamic visualizations enabled him to construct an appropriate scenario with which to visually and verbally describe the behavior of the randomization test and facilitated conceptual understanding of the underlying concepts. Thus it would appear that dynamic visualizations, preceded by appropriate hands-on activities, have facilitated in these students a consolidation of abstract inferential concepts using a process of visual reasoning and that these visualizations are becoming part of the way that they think. Although the randomization process using the VIT modules provides a visual image for a tail proportion, the interpretation of the tail proportion was not part of the visualization and hence remains difficult for many students.

ACKNOWLEDGEMENT
Work in this paper is supported in part by a grant from the Teaching and Learning Research Initiative (http://www.tlri.org.nz/sites/default/files/projects/9295_summary%20report.pdf).

REFERENCES

Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *Educational Studies in Mathematics, 52*, 215-241.

Budgett, S., Pfannkuch, M., Regan, M., & Wild, C. J. (2013). Dynamic Visualizations and the Randomization Test. *Technology Innovations in Statistics Education, 7*(2), http://escholarship.org/uc/item/9dg6h7wb

Clark, J., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149-210.

Cobb, G. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1), 1-15.

Hegarty, M. (2004). Dynamic visualizations and learning: Getting to the difficult questions. *Learning and Instruction, 14*, 343-351.

Hesterberg, T., Moore, D., Monaghan, S., Clipson, A., & Epstein, R. (2009). Bootstrap methods and permutation tests. In D. Moore, G. McCabe, B. Craig, D. Moore, G. McCabe, & B. Craig (Eds.), *Introduction to the practice of statistics* (pp. 16-1 - 16-60). New York, NY: Freeman.

Presmeg, N. (2006). Research on visualization in learning and teaching mathematics. In A. Gutierrez, & P. Boero (Eds.), *Handbook of Research on the Psychology of Mathematics Education* (pp. 205-235). Sense Publishers.

Radford, L. (2009). Why do gestures matter? Sensuous cognition and the palpability of mathematical meanings. *Educational Studies in Mathematics, 70*, 111-126.

Rossman, A. (2008). Reasoning about informal inference: One statistician's view. *Statistics Education Research Journal, 7*(2), 5-19.

Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review, 2*, 98-113.

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science, 12*(2), 257-285.

Ware, C. (2008). *Visual thinking for design.* Burlington, MA: Morgan Kaufmann Publishers.