Students' Understanding of Randomness after an Introductory Tertiary Statistics Course

Robyn Reaburn
University of Tasmania
<Robyn.Reaburn@utas.edu.au>

Random sampling and random allocation are essential processes in the practice of inferential statistics. These processes ensure that all members of a population are equally likely to be selected, and that all possible allocations in an experiment are equally likely. It is these characteristics that allow the validity of the subsequent calculations that use probabilistic reasoning. This paper suggests that despite the importance of these processes, students may poorly understand the characteristics of these processes, and the reasons for them. The paper concludes with suggestions for the improvement of teaching these topics.

This paper describes students' understanding at the end of an introductory tertiary statistics unit at an Australian University, of the importance and reasons for random sampling and random allocation. The discipline of inferential statistics arises because researchers are usually forced to make decisions on the basis of samples, not entire populations. Due to ubiquitous variation, it is extremely unlikely that any two samples will be exactly alike or have exactly the same characteristics of the original population. Therefore, probabilistic reasoning is used to make conclusions about the populations and assign levels of uncertainty to these conclusions. If this probabilistic reasoning is to be valid, then either random sampling or random allocation will have had to be used. These processes ensure that either all potential members of the sample are equally likely to be selected, or, in the case of experiments, that all possible allocations between treatments and control are equally likely. For example, it would not be valid to estimate the mean length of fish in a holding pond if the sampling method resulted in the smaller fish being more likely to be chosen. Neither would it be valid to test for differences between treatment groups based on the hypothesis of no difference between the groups (such as occurs in the calculation of the *P*-value), if the allocation made it more likely that certain characteristics ended up in one group over another. In addition, random allocation minimises the effects of any unknown confounding factors. If such a factor is present, then random allocation makes it likely that it will be evenly distributed between the different groups.

It is because of the importance of these processes that Lecoutre, Rovira, Lecoutre, and Poitenvieau (2006) stated that "Variants of the concept of randomness are at the heart of probabilistic and statistical reasoning" (p. 21). Despite this importance, however, previous research has shown that students can be confused about the meaning of randomness in mathematics and its use in inferential statistics.

Part of this confusion arises because of the way the word *random* is used in general conversation. In everyday speech, random may refer to anything that is without a pattern, that is haphazard, that is without a known cause, or is unpredictable (Batanero, Green, & Serrano, 1998; Petersen, 1998). In mathematics and statistics, however, *random* can refer to a process where an overall structure exists even though individual outcomes cannot be predicted.

Previous research (Kaplan, Fisher, & Rogness, 2009), carried out on students at the beginning of a tertiary statistics unit, found that these students had views on randomness similar to the general population, that is, that the term referred to anything unplanned, unexpected or haphazard. These students also indicated that randomness applied to any selection that was done without pre-set criteria or a plan or was carried out without bias.

2019. In G. Hine, S. Blackley, & A. Cooke (Eds.). Mathematics Education Research: Impacting Practice *(Proceedings of the 42nd annual conference of the Mathematics Education Research Group of Australasia)* pp. 572-579. Perth: MERGA.

Lecoutre et al. (2006) presented high school students, mathematics researchers and psychology researchers with various real and stochastic scenarios (e.g., You meet a friend you have not seen for 10 years; an even number was obtained from the role of a die) and asked them to select which ones were random. The majority of participants stated that the stochastic items were random because probabilities could easily be computed. The participants were divided when it came to real situations: some participants stated that because probability was involved these events were random, and others said that, because causal factors could be identified, randomness was not involved. Some of the psychology researchers stated that because the probabilities could be easily computed the stochastic situations these were not random. Interestingly, the researchers found that the mathematics and psychology researchers were no more accurate in their answers than the students in secondary school who had not studied probability before. This indicates that the concept of randomness may be as confusing for academics as well as the rest of the population.

When investigating the understanding of students at the end of a first year tertiary statistics unit del Mas, Garfield, Ooms, and Chance (2007) found that approximately 60% of the students understood that random allocation supports causal inference. However, only approximately 12% of the students could demonstrate understanding of the purpose of randomisation in an experiment. This paper examines students' beliefs about randomness at the end of a first year tertiary statistics unit at an Australian university, and gives further insight into their understandings and difficulties.

Method

Selection of Participants

The research described in this paper was part of a larger study examining students' understanding of inferential statistics at the end of an introductory one-semester unit at an Australian university (Reaburn, 2011). This study took place over four teaching semesters, each with a new cohort of students. The majority of these students studied computing, biomedical science, aquaculture, or sports science and were not required to have previous experience of calculus. The participants in this research were volunteers and there was considerable variation in the proportion of students who agreed to let their data be used for this research. In the first semester there were 12 volunteers out of a possible 20 students; in the second semester 23 out of a possible 26; in the third semester 6 out of a possible 27; and in the last semester 12 out of a possible 26. The researcher was the lecturer and tutor, and owing to ethical considerations the reasons for this variation could not be investigated.

Teaching and Learning Activities

Introduction to the concept of randomness.

Before the concept of randomness was introduced, the students were shown two grids (Figure 1). Their instructions stated that these grids indicated the number of buttons that had landed in each square after 50 buttons had been dropped onto the grids. They were asked to state if the two grids represented a real situation, or if one or both had been made up. Using the random number generator in *Microsoft Excel* they repeatedly simulated this scenario. Before the simulation, some of the students thought the second grid may have been made up because of the apparent clustering. After the simulation, all of the students suggested that both of the grids could be real.

3	0	3	4	1
3	3	3	3	5
1	1	3	0	3
2	1	2	1	2
0	2	2	1	1

2	1	3	0	3
1	3	1	2	1
5	3	1	2	1
2	2	2	1	2
3	1	1	6	1

Figure 1. The two grids representing the number of buttons that landed in each square.

After this exercise there was a lecture about the long-term frequencies of repetitive, stochastic processes using examples such as tossing a coin. It was emphasised that, whereas each outcome cannot be predicted, it is expected that there will be structure in the long term: for example, the proportion of a particular outcome. In addition, it was also pointed out that short-term clusters might appear, one illustration being the button simulation.

Introduction to random sampling.

The idea of random sampling was introduced in a formal lecture. In this lecture, random sampling was defined as a process by which every member of a population is equally likely to be selected, and where the selection of one member does not affect the probability of any other member being selected. It was also pointed out that if every possible sample of fixed size is equally likely to be selected, then this process is also considered to have the characteristics of random sampling, a point reinforced in the notes for the unit. The notes for the unit emphasised, by the use of examples, that random sampling does not guarantee that a sample is representative. During the computer session for this week the students used the random number generator in *Microsoft Excel* to take random samples (n=10) from a population (N=100). By this means they could observe that it was not uncommon to select two members of the population who were listed near to each other in the original list.

The principles of random sampling were reinforced with a scenario in a mock test used for revision. In this scenario it was proposed that a psychologist had advertised for volunteers. Because he had had no control over who volunteered he claimed that random sampling had been used. This provided the students with an opportunity to revise the characteristics of random sampling as they made judgments on the claim.

Introduction to random allocation.

The concept of random allocation was introduced in a formal lecture that included some of the principles of experimental design and double-blind clinical trials. The first reason given for random allocation was that if there are any unknown confounding factors present, they were likely to be evenly distributed among the treatments (Sedgwick, 2012). Secondly, by setting up a process where all possible allocations are equally likely, it is possible to carry out the probabilistic calculations which are made on this assumption.

After this lecture the students used the random number generator in *Microsoft Excel* to allocate members of a group between a control and treatment group. In the second of their four written formal assessments they were also required to define random allocation and explain how this could be carried out. Further consolidation was provided in the written feedback from their markers. Random allocation was also used by the students in computer simulations that introduced them to the two-sample *t*-test and the chi-squared test for independence (Reaburn, 2014).

Assessment of student understanding.

Student understanding of random sampling and random allocation were assessed by: (1) items in a written questionnaire which were filled out only by the volunteers, and (2) questions in an open-book test that was taken by all the students. The questionnaires were administered by a member of staff who was not involved in the teaching or assessment of the unit; after collection they were withheld from the researcher until the official unit results had been released. The test was administered by the lecturer/researcher but at the time of administration it was not known which of the students were participants. This information was given to the researcher after the official unit results had been released. After the data on the relevant questions for the research was collected all identifying information was removed.

Relevant question from the test.

"Random sampling is used to ensure representative samples are obtained."
TRUE/FALSE? Give a brief justification of your answer.

Relevant questions from the questionnaire

A. Give an example of something that happens in a random way. Explain why you think this is an example of randomness.

B. A scenario was given where a consumer organisation was testing the weights of a brand of breakfast cereal. The students were then given a series of sample means and asked to make estimates of how likely each would be, given the stated weight on the box. The last question for this scenario was:

   If you didn't use random sampling, how would this affect your previous answers to this question?

C. In this scenario a university was studying the effects of sleep on test scores. Using random allocation half the participants stayed up all night before a test, and the other half went to bed by 11:30pm. Another university undertook the same study but let the participants choose which group they went into. The question was:

   Which group of researchers was correct? Explain your reasoning.

D. In this scenario a plant geneticist (Fred) has a paper rejected because he did not use random allocation between the control and treatment, even though he had controlled for temperature, light and "everything else". He has a discussion with a statistician who shows him how to randomly allocate treatments. When this allocation is carried out the geneticist gets the same allocation he had previously used and is very annoyed by this. The statistician tells him the problem was not the layout itself, but how the layout was determined. The question was:

   Can you explain to Fred why randomisation was so important? See if you can provide an argument for randomisation that will overcome Fred's problem.

Analysis

The entire test and questionnaire were coded using the Rasch Partial Credit Model (Masters, 1982). Using this model, questions are coded according to their level of understanding. For example, an answer that is correct but gives no explanation would be given a code of "1"; and answer with some explanation would be given a code of "2" and if more understanding is demonstrated a code of "3" would be given. Incorrect answers receive a code of "0". In addition, Rasch Analysis gives independent assessments of the difficulty of the items and students' ability. It was found, however, that the test items did not fall on a unidimensional scale and were therefore unsuitable for Rasch analysis.

Results

Questionnaires were received for Semesters 1, 2 and 4 of the research. For reasons beyond the researcher's control, the questionnaire was not administered in the third semester. The tests were administered for all four semesters, but unfortunately not all the responses for the test question reported here were available for analysis. A total of 31 questionnaires were received and 26 tests analysed for this paper.

Question from Test

For this question there were 26 responses available for analysis. The answers demonstrated that, whereas approximately two-thirds of the participants could demonstrate knowledge that random sampling did not ensure representative samples, they could not necessarily explain why. Approximately half of these stated that as each possible sample was equally likely to be selected, it was possible that samples could be obtained that were not representative of the population overall. The explanations that the other participants in this group gave included that random sampling ensured there would be "no patterns" and that random sampling precludes bias. Answers that demonstrated less understanding included: that as the size of the sample increases it will be more representative and that with random sampling we get the probability of an event. Of the participants who answered that the statement was true, four gave no explanation, two stated there would be no bias, and one defined Random Sampling but then stated that the samples from each category would be selected. There were two answers that were idiosyncratic (did not directly answer the question or could not be understood) and one did not answer.

Questions from the Questionnaire

The level of difficulty (in logits) for each item is reported in Table 1 below. Analysis of the Rasch-Thurstone thresholds, which compares the difficulty scores of the items with the ability of the participants, indicated that all participants had found Questions B, C and D difficult; Table 1 indicates that Question C was found to be the most difficult.

Question A.

Whereas approximately three-quarters of the participants gave suitable examples, the others gave answers that did not reflect the work throughout the semester, such as being struck by lightning or a car breaking down. All except two of the participants who gave suitable examples (coins, die or similar) included an appropriate explanation (e.g., cannot determine individual events; long term pattern). One participant did not answer.

Table 1

Levels of difficulty of each item

Item	Level of difficulty
Question A	0.8
Question B	1.61
Question C	1.83
Question D	1.35

Question B.
Thirteen of the participants gave explanations for why random sampling was required in this context: that without random sampling the sample could be biased in some way and invalidate the results, or that boxes manufactured at the same time could have similar errors. A further six participants stated that the results could be “biased”, “skewed”, or that not every sample would be equally likely, without any further explanation. Other reasons given by participants included that the results would be unrepresentative of the population or the results would not fit a Normal Distribution. Nine of the participants either did not answer or gave answers that were idiosyncratic.

Question C.
Overall, the participants could not explain their reasoning for this question. Seventeen of the participants stated that letting the students in the experiment self-select would bias the results without further explanation. Six of the participants stated that the university using random allocation was correct or that self-selection would lead to non-random allocation but could give an explanation beyond this. Six participants stated that it would be better for the students to self-select.

Question D.
Whereas approximately three-quarters of the participants gave legitimate reasons for their answers to this question, overall there was little explanation. The reasons given included that random allocation “removes bias”, that each layout needs to equally likely, and that random allocation was needed to make the statistical calculations legitimate but no further clarification was added. Similarly to Question C, five participants stated that random allocation “should be used” as the only explanation. One participant stated that it was “only chance” that made it “happen that way”, and 19% did not answer or gave answers that were idiosyncratic.

Discussion and Conclusion

It is apparent that many of the participants realised that in experiments random allocation is important for reducing the chance of bias and that all possible allocations should be equally likely, but in general they could not explain why. Similarly, many of the participants knew that random sampling results in all members of a population having the same chance of being selected but could not explain the importance of this. Only three of the participants mentioned the issue of the validity of the calculations used in any analysis. It was also of interest that there were six participants who were prepared to agree with the statement that random sampling leads to representative samples in a formal assessment, even though this was an open-book test and the information was available in the unit notes.

When asked to nominate a random event, eight of the participants gave examples that did not relate to the content of the unit, even though the questionnaire was administered in the lecture theatre where all their statistics lectures were held and were therefore in their normal learning environment. Despite being consistently exposed to ideas of randomness in inferential statistics these were not used in their answers.

This paper gives further evidence of the difficulties students have understanding the strange, non-deterministic world of inferential statistics. In this world, students, often without strong mathematical backgrounds, are required to cope with ideas of randomness (where often intuition is incorrect), probabilistic reasoning, distributions of sample statistics, variable identification and model building (Mustafa, 1996). It is possible that it is the difficulty of the subject that led several students to revert to the everyday world when asked to give examples of random events.

Another explanation of this reversion may be the quantity of material that students are required to manage when learning to work in this discipline for the first time. Learners may not retain new information if it is too much for their short-term memory, especially if the information is given in a didactic way (Krause, Bochner, & Duchesne, 2007), and this is how much of the information involving randomness was presented in this research.

This research was part of a wider Action Research project the main aims of which were to investigate students' understanding of Confidence Intervals and *P*-values (Reaburn, 2011; Reaburn, 2014). As this project progressed, the level of didactic teaching decreased as the use of other strategies increased with a concurrent increase in understanding of these topics. These strategies included: simulation before teaching the formal concepts (such as the Central Limit Theorem) using a predict-test-revise format (Lane & Peres, 2006); encouraging the students to write about their understanding to assist in the refining of their ideas (Ozgun-Koca, 1998); and relating the material to examples that could be easily understood (Shaughnessy & Chance, 2005). These strategies, however, were used in a limited fashion for the topics involving randomness.

How could the teaching of these topics have been improved? One strategy could have been to relate the equal likelihood criterion of random sampling and random allocation to the classical view of probability – that probability is the ratio of the outcomes of interest to the total number of outcomes, but only if all the outcomes are equally likely. It would be expected that students would easily understand the invalidity of such calculations if a die, for example, was weighted.

In addition, simulation in a spreadsheet such as *Microsoft Excel* could have been used to demonstrate the effect of random allocation in distributing confounding factors. If a column is created of 'subjects' to be distributed between two treatments, and a number of these marked in some way (for example, by colour) to indicate those with a confounding factor, random numbers can be used to mix up the subjects between the two groups and the distribution of the subjects with the confounding factors observed.

The discipline of inferential statistics is based on probabilistic reasoning. The validity calculations and conclusions, therefore, depend on probabilistic processes being used to select samples and to allocate treatments. This paper adds to our knowledge of the difficulty students have in adapting their ideas to ideas of randomness in inferential statistics, and in explaining their knowledge. It further illustrates the importance of educators constantly reviewing students' understanding, and to search for non-didactic methods for presenting the material.

References

Batanero, C., Green, D., & Serrano, L. (1998). Randomness, its meanings and educational implications. *International Journal of Mathematics Education in Science and Technology, 29*(1), 113-123.

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6*(2), 25-58.

Kaplan, J., Fisher, D., & Rogness, N. (2009). Lexical ambiguity in statistics: What do students know about the words association, average, confidence, random and spread? Journal of Statistics Education, *17*(3), www.amstat.org/publications/jse/v17n3/kaplan.html

Krause, K., Bochner, S., & Duchesne, S. (2007). *Educational psychology for learning and teaching*. South Melbourne, VIC: Thomson.

Lane, D., & Peres, S. (2006). Interactive simulations in the teaching of statistics: Promise and pitfalls. In B. Phillips (Ed.), *Developing a statistically literate society* (Proceedings of the 7th International Conference on Teaching Statistics, Salvador, Brazil) [CDRom]. Voorburg, The Netherlands: International Statistics Institute.

Lecoutre, M., Rovira, K., Lecoutre, B., & Poitevineau, J. (2006). People's intuitions about randomness and probability An empirical study. *Statistics Education Research Journal, 5(1*), 20-35.

Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.

Mustafa, R. (1996). The challenges of teaching statistics to non-specialists. *Journal of Statistics Education, 4*(1). Retrieved from http://www.amstat.org/publications/jse

Ozgun-Koca, S. (1998, October-November). Students’ use of representations in mathematics education. Paper presented as the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Raleigh, NC.

Reaburn, R. (2011). *Students’ understanding of statistical inference: Implications for teaching*. Doctoral Thesis. Australia: University of Tasmania.

Reaburn, R. (2014). Introductory statistics course tertiary students' understanding of *P*-values. Statistics *Education Research Journal, 13* (1), 53-65.

Sedgwick, P. (2012). Why randomise in clinical trials? *British Medical Journal, 345*. doi: https://doi.org/10.1136/bmj.e5584

Shaughnessy, J., & Chance, B. (2005). *Statistical questions from the classroom*. Reston, VA: NCTM.