OZCOTS 2016

Proceedings of the 9th Australian Conference on Teaching Statistics

OZCOTS

Editors:
Helen MacGillivray
Queensland University of Technology

Michael A. Martin
Australian National University

Brian Phillips
Swinburne University

Canberra
8-9 December, 2016

Satellite to ASC 2016 Australian Statistical Conference
Canberra, 5-9 December, 2016

SSAI
Statistical Society of Australia Inc.

ISBN: 978-0-9805950-2-4

Friday 9 December 2016 (continued)		
OZCOTS Program – Learning strategies for statistics using technology (Chairperson: John Harraway)		
1400 – 1418	Anthony Morphett – Applets to support reasoning about explained and unexplained variability	123
1418 – 1436	Sharon Gunn – To R or not to R – what should we be considering?	129
1436 – 1454	Alice Richardson – Australian statistics poster and project competitions	135
1454 – 1512	Dean Langan – Guidance for teaching R Programming to Non-Statisticians	141
1512 – 1530	Timothy Kyng – Big data, data science, computer science & statistics education	147
1530 - 1600	AFTERNOON BREAK	
OZCOTS Program - Statistics for researchers and workers (Chairperson: Michael Martin)		
1600 – 1618	John Harraway – Measuring learning within a large design research project	153
1618 – 1636	Peter Ellis – Statistical capability building in a government department - the New Zealand ministry of business, innovation and employment	159
1636 – 1654	Angie Wade – Development of a training program for non-statisticians	165
1654 – 1712	Ian Gordon – Were we at the same consultation?	171
1712 – 1730	Belinda Chiera – Teaching visualisation in the age of Big Data: Lessons from the past	172
1730	CONFERENCE CLOSE	

PREFACE

OZCOTS 2016

9th Australian Conference on Teaching Statistics

OZCOTS 2016 theme: Statistics Education in a Big Data Era

The teaching and learning of Statistics is now of more importance than ever to industry, government, business and indeed the whole of society. The roles of statistical understanding and statistical thinking are vital in all these areas across disciplines, increasingly driven by big data, evidence-based agendas, and technological advances which generate data as well as enabling more complex problem-solving, data visualisation and analysis. OZCOTS 2016, the 9th Australian Conference on Teaching Statistics, will consider many challenges of teaching Statistics for future statisticians, statistical users and consumers under the theme of "Statistics Education in a Big Data Era".

OZCOTS 2016 is again building on the success of the timing and format of OZCOTS 2008, 2010 and 2012 as a conjoint event with the (ASC) with an overlap on Thursday, 8 December.

The OZCOTS program includes keynote and contributed papers on issues across the statistical education spectrum of interest to the whole statistical profession and statistical educators. Topics addressing challenges in a big data era range across learning strategies, curricula and technology for teaching introductory and undergraduate statistics; resources and online learning; statistics learning for postgraduates, researchers and workers; and research for teaching statistics.

Helen MacGillivray, Michael Martin, Brian Phillips
Joint Editors

OZCOTS 2016 Conference Committee

Helen MacGillivray (joint chair, joint editor), Queensland University of Technology
Michael A. Martin (joint editor, ASC 2016 Program Committee), Australian National University
Brian Phillips (joint chair, joint editor), Swinburne University
Bronwen Whiting (ASC 2016 Local Organising Committee), Australian National University

OZCOTS 2016 Paper Refereeing Process

Papers referred to in the proceedings as refereed publications were reviewed and accepted as meeting the requisite standards by at least two referees selected from a panel of peers approved by the OZCOTS 2016 editors.

The review process was "double blind" - identification of both authors and referees was removed from all documentation during the reviewing process. The Conference Committee took the view that the review of papers would give conference participants and other readers confidence in the quality of the papers specified as "refereed" in the proceedings. The refereeing process also provided a mechanism for peer review and critique and so contributed to the overall quality of statistics education research and teaching. While the refereeing process essentially relied on subjective judgments, referees were asked to compare the paper being reviewed against the accepted norms for reporting of research. It was expected that each accepted paper would represent a significant contribution to advancement of statistics education and/or the research processes in statistical education. Authors verified that the refereed published papers for these proceedings were substantially different from papers that have been previously published elsewhere.

OZCOTS 2016 gratefully acknowledges the following referees for their assistance: Pip Arnold, James Baglin, Kaye Basford, Ayse Bilgin, Glenys Bishop, Stephanie Budgett, Michael Bulmer, Murray Cameron, Belinda Chiera, Sam Clifford, Kristen Gibbons, Ian Gordon, John Harraway, Paul Hewson, Peter Howley, Malgorzata Korolkiewicz, Kay Lipson, Peter Martin, Brad Payne, Peter Petocz, Robyn Pierce, Alice Richardson, Doug Stirling, Ross Taplin, Garth Tarr, Jane Watson, Neville Weber, Ian Westbrooke, Chris Wild, Therese Wilson, Richard Wilson.

OZCOTS 2016 – Keynote Speakers Biographies

Professor George W Cobb

George Cobb is emeritus Professor of Statistics at Mount Holyoke College, where he taught from 1974 through 2009. He received his undergraduate degree from Dartmouth College, a masters degree in biometry at the Medical College of Virginia, and a PhD in statistics from Harvard University. He was elected a Fellow of the ASA in 1993, served a term as vice president, and received a Founder's Award in 2006. In 2005 he received an award for lifetime achievement at the US Conference on Teaching Statistics. This current year, 2016, he was chosen Mosteller Statistician of the Year by the Boston Chapter of the ASA. His interests include statistics education, applications of experimental design, and applications of statistics to the law. He has written an elementary textbook on design of experiments, has helped write textbooks for first and second courses in statistics, and is at work on a "popular science" introduction to Markov Chain Monte Carlo.

George Cobb has spent his entire life changing his mind. His childhood ambitions, in succession, were to be a herpetologist, a center for the Green Bay Packers, a surgeon, a chemist, and, by the end of his senior year in high school, aspiring ever upward, a mathematician. It is no surprise, therefore, that when he graduated from Dartmouth College in 1968 he had been accepted into graduate programs in Russian literature. Like Mae West, he drifted, eventually earning a PhD in statistics from Harvard in 1974, but he maintains his love for mathematics as well as statistics and his conviction that Plato was right: mathematics is the easiest subject, and so every philosopher king should devote a decade to its study before presuming to tackle the harder subjects of politics and sociology. Cobb spent his entire teaching career at Mount Holyoke College, from 1974 until his retirement in 2009, dedicated to this conviction.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

OZCOTS 2016 – Keynote Speakers Biographies

Dr Kay Lipson

Kay is currently the Director of Strategy for Online Education Services (OES). Founded in 2010, OES is a public-private partnership between SEEK, who have matched over 150,000 students with their ideal course and Swinburne University of Technology, a pioneer in online education with over 20 years' experience.

The partnership's first endeavour, Swinburne Online, was launched in 2011 with 10 undergraduate courses in the fields of Business, Social Science and Communication. By 2015, Swinburne Online had expanded its offerings to include 21 undergraduate and postgraduate university courses, five TAFE courses and offerings for international students, with over 8,000 students currently enrolled.

Kay's discipline area is Statistics, and she has previously undertaken academic roles at the University of Melbourne, Monash University, and Swinburne University of Technology. Her most recent academic role in leadership was as Dean of the Faculty of Higher Education, Lilydale, at Swinburne University of Technology, from 2009 to 2011.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

Keynote 1

Ask not what data science can do for the Humanities. Ask rather, what the Humanities can do for data science.

George W. Cobb
Professor of Statistics Emeritus,
Mount Holyoke College, MA, USA

In a sense, all curriculum is ultimately local, which means that, not knowing your local environment, I can't aim higher than to hope to provoke some thinking that will continue after we all disperse. My planned provocations are based on three theses. In offering a paltry three, I recognize that I am 92 short of a Martin Luther. (Should data science therefore rate me at 3/95 = 3.71985%?)

My theses: (1) At its best teaching ought to confront students with essential tensions and overlooked context. (2) If data science is to be taught as a subject worthy of this ambitious goal, it must therefore involve students with some of those tensions and contexts. (3) If data science is to confront students with essential tensions and overlooked contexts, history and literature are both essential but too often overlooked.

I plan to base my argument on six examples: (1) four translations of the opening line of Pushkin's *Eugene Onegin*; (2) the ambiguous rhyme scheme of the Onegin stanza; (3) fog and mud in Dickens's *Bleak House*; (4) contrasting styles in Joyce's *Wake*, *Ulysses*, *Portrait*, and *Dubliners*; (5) Marx and the hanging of convicted witches in Salem of 1692; and finally -- to leave you with a concentrated mind – (6) some 1850s data analysis on the science of hanging.

For each set of examples, I suggest there is a role for data science, but also unresolvable tensions and a role for context. Each set of examples, viewed narrowly enough, offers a data analytic challenge. But for each set of examples I suggest that the larger challenge is to figure out how science can decide which of various narrow views are worth pursuing. To date history and literature do better than data science.

Keynote 2

Reimagining Online Learning

Kay Lipson
Director of Strategy for Online Education Services (OES) Melbourne

The online education industry continues to grow rapidly in Australia, with approximately 400,000 students across Australia currently undertaking university studies predominately online. Unlike the traditional undergraduate student who is studying full-time on campus, online students are often studying part-time whilst they juggle the demands of life, work and family. For these students, it is particularly important that learning activities are focused and meaningful and support services timely and accessible. This presentation looks at how learning analytics, the process of collecting, aggregating and analysying interactions with students, is informing the practice of online education today.

Keywords: online education, learning analytics

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

EXPLORING THE IMPACT OF VISUAL REPRESENTATIONS OF VARIATION ON INFORMAL STATISTICAL INFERENCE

BAGLIN, James and GRANT, Sally
School of Science, Mathematical Sciences,
RMIT University, Melbourne, Australia
james.baglin@rmit.edu.au

This experiment investigated how different methods of data visualisation can influence informal inference. Two hundred and five student participants with an introductory statistics background completed an online randomised survey. Participants were presented with a scenario that required them to make inferences about the difference between the life of two battery brands, based on a plot visualising the data. The plot type and the dataset visualised were manipulated based on the degree to which a plot depicted variability in the data (bar, box and dot) and whether or not the plot depicted data sampled from populations where each brands' battery life was equal or unequal. Both manipulations were found to significantly impact the informal inferences drawn by participants. The experiment also discovered an interaction between plot type and brands' battery life equality. Plots that depicted variability (box and dot) led to more correct inferences when the plots depicted data sampled from unequal populations, but fewer correct inferences when they displayed data sampled from identical populations. The findings build upon previous research by demonstrating that different methods of data visualisation can affect informal inference. These findings have important implications for statistics education and data visualisation.

INTRODUCTION

Statistical inference is the process of generalising beyond the sample data to make a statement about a population or a causal process (Rossman, 2008). This typically involves using a formal hypothesis test to probabilistically determine if sample data are consistent or inconsistent with assumptions made about a population. However, statistical inference can also be made informally using the same process of generalisations from the data to the population, but outside a formal probabilistic framework (Makar & Rubin, 2009). There are many situations where informal statistical inference arises, and as such, informal inference has received much attention in the statistics education literature (Paparistodemou & Meletiou-Mavrotheris, 2008; Pfannkuch, 2006; Rubin, Hammerman & Konold, 2006). Many of these studies have focused on the use of data visualisations to assist with making informal inferences.

Common data visualisations used in statistics include bar charts, histograms, box plots, dot plots, line graphs, error bar plots, and scatter plots, to name a few. These visual methods, which are capable of representing counts, percentages, proportions, central tendency, variation and even the raw data itself, are often relied upon to make informal inferences. A person's visual sense and knowledge of statistics are utilised to examine a data display to determine if the situation depicted by the sample is indicative of an effect that is likely to exist in the population. This happens frequently during exploratory data analysis and also in many contexts where time restrictions or inadequate knowledge prevent formal processes (e.g. media, workplaces, advertising etc.).

Visualisation is seen as the most intuitive medium through which to introduce the challenging concept of statistical inference (Makar & Rubin, 2009). Previous literature has mostly utilised qualitative studies to determine how best to teach inference to school students using visualisation before they learn formal tests (Makar and Rubin, 2009; Paparistodemou & Meletiou-Mavrotheris, 2008; Pfannkuch, 2006; Rubin, Hammerman & Konold, 2006). There have also been numerous experimental studies that have investigated how differences in one kind of data visualisation method influence interpretation (Best, Smith & Stubbs, 2008; Carpenter & Shah, 1998; Cleveland, Diaconis & McGill, 1982). For example, Cleveland, Diaconis and McGill (1982) manipulated the scale used to present a scatter plot showing the correlation between two continuous variables to 109 participants. Plots where the data took up a smaller area of the graph were judged by participants to exhibit stronger correlations than plots of the same data where the data took up a greater space on the plot. Experiments such as this have shown that the choices made about the way in which data are visualised can have a direct impact on interpretations.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

However, with the exception of Shah, Mayer and Hegarty (1999) very few investigations have considered how different types of data visualisations of the same data impact informal inference.

Shah, Mayer and Hegarty (1999) conducted a series of experiments comparing how undergraduate students read graphs of historical data presented as either a bar or line chart. In these experiments, participants provided open-end interpretations. One of their key findings was that line charts prompted more interpretations and comparisons based on time when compared to the same data depicted as bars charts. Shah et al. demonstrated that the choice of data visualisation method can have an impact on the interpretations made. However, they did not compare different methods of visualising variability of quantitative variables, nor, specifically, how different data visualisations influence informal inference, or generalising beyond the sample.

Therefore, this experiment aimed to explore how representations of variability in common data visualisations of quantitative data influence informal inferences. Specifically, this experiment aimed to address the following exploratory research question: Do different visualisations, which depict variability to various extents, facilitate more accurate informal inferential reasoning in detecting differences between a quantitative variable sampled from two populations?

METHOD

Following ethics approval and permission from relevant subject coordinators, students from three Australian universities were invited to partake in a 5 – 10 minute, anonymous, online survey. Advertisements to recruit participants were posted on statistics subjects' learning management systems. The final participant count included 205 (120, 58.5% female; 83, 40.5% male; 2, 1%, prefer not to say) students, with 233 having commenced the survey (88% completion). Most students were aged from 18-25 (*N* = 125, 61%). The majority of respondents came from university A (58.1%), with 36% coming from university B and 5.9% from university C. There were 59.5% of the sample that identified as students from natural and physical sciences and health science backgrounds. Most students self-rated their level of statistics knowledge at the introductory level (59.5%), with 38% at the intermediate level and 2.4% at an advanced level.

Participants clicked through the initial link to the anonymous online survey, which was created and hosted using Qualtrics (www.qualtrics.com). The online survey included three sections: i) a participant information statement (PIS), ii) the stimulus, which included the presentation of the scenario comparing battery brand life and an experimentally manipulated plot, and iii) participant demographics (see previous paragraph).

Participants read the online PIS, and consent was implied if they clicked through to the stimulus section. In the stimulus section, each participant was presented with the following scenario:

> *You are trying to decide between buying two different battery brands, A or B, which have a similar cost. You come across the following plot which compares the life of a random sample of 20 batteries from each brand. Battery life was measured as how many digital photographs a camera was able to take before the battery was considered dead. Using the plot below, answer the following questions.*

Participants were also randomly assigned, according to a 3 by 2 between-subjects factorial design, to view one of six different plots (Figure 1). The independent variables were plot type (three levels: bar, box or dot plot) and brand equality (two levels: equality and inequality). The three different plots were chosen due to the difference in the degree to which they represented variability in the data. The bar plot, with error bars representing ± 1standard deviation (*SD*), was a poor approach to visually summarising quantitative data, but was selected because bar plots appear frequently in scientific literature to represent means. Box plots present minimum and maximum values and quartiles and dot plots represent binned raw data. The second factor, brand equality, and its associated levels, equality and inequality, were chosen in order to avoid confusion with the common terms used to refer to effect size and difference between means when reporting the results of the experiment. Battery brand equality meant that the population mean battery life (pictures taken) for the two brands was equal, $\mu_A - \mu_B = 0$, whilst inequality meant that the brands' mean battery life was unequal, $\mu_A - \mu_B \neq 0$. The following population distributions were used to randomly generate two datasets: Equality: $x_{A,B} \sim N(69.78,\ 39.44)$; inequality: $x_A \sim N(69.78,\ 39.44)$ and $x_B \sim N(76.07,\ 39.44)$. A and B referred to battery brand and x was a normally distributed random variable, pictures taken (rounded to zero decimal places), with a population mean, μ, and variance, σ^2, $N(\mu, \sigma^2)$. A small sample size of N = 20 for each battery brand was selected, which equated to an estimated statistical power of 87% (two-tailed, two-sample t-test with α = 0.05) for the inequality condition. The scenario and variables were based loosely on the battery brand comparison tests reported by Dunn (2013). The inequality condition corresponded to a population standardised mean difference, or Cohen's $d = (69.78 - 76.07)/\sqrt{39.44} = -1.00$.

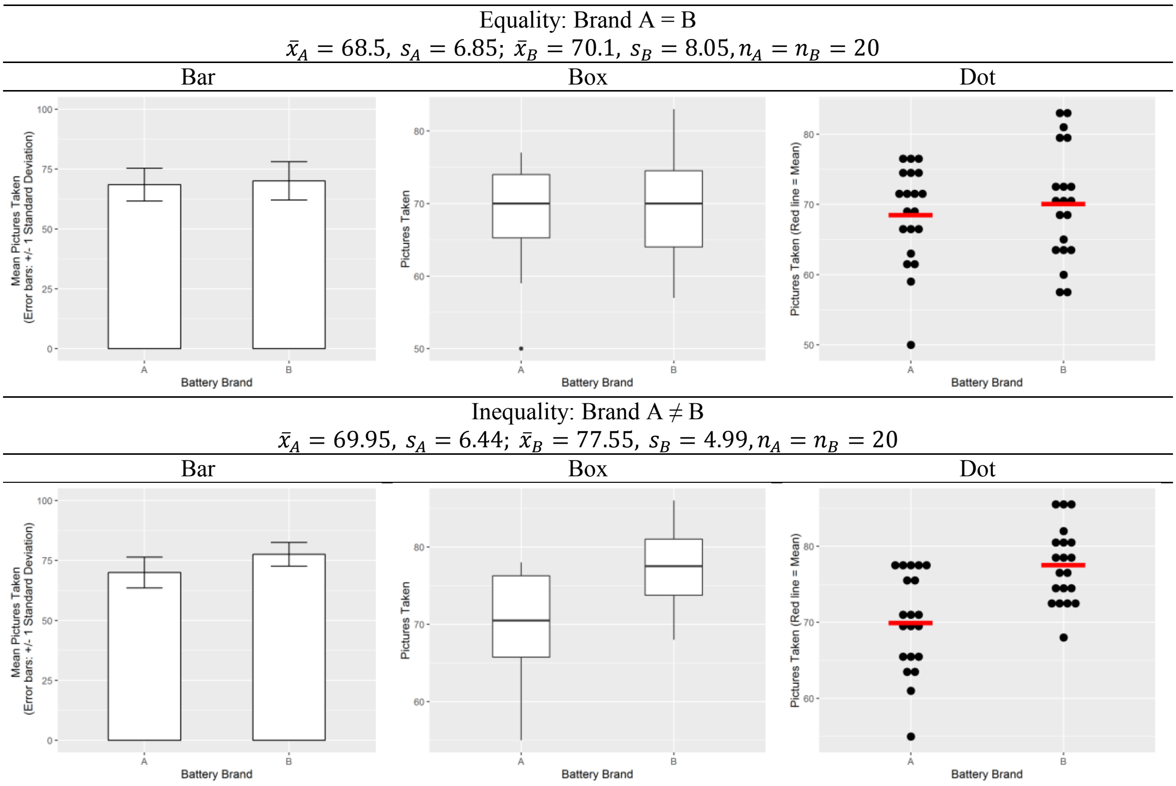


Figure 1. The six plots randomly assigned according to the 3 by 2 factorial design.

Figure 1 also reports the descriptive statistics for the randomly generated data for both the equality and inequality conditions. A two-sample t-test comparing the mean pictures taken between brands A and B for the equality condition was not statistically significant, $t(df = 37.09) = -0.68$, p = 0.50, 95% CI (-6.38, 3.18), Cohen's d = -0.21. A two-sample t-test for the inequality condition was statistically significant, $t(df = 35.76) = -4.17$, $p < .001$, 95% CI (-11.30, -3.90), Cohen's d = -1.32. From a formal statistical inference perspective, the randomly generated data met the requirements for the experimental conditions.

Participants were randomly allocated to view one of the six plots by the randomisation feature built into Qualtrics. Once participants viewed their randomly allocated plot, they were required to respond to the following two questions using forced-choice response options: Question 1) *Based on the plot presented, do you think there is sufficient evidence to prefer one brand over the other?* Options: A) *Yes, the plot suggests that Brand A would tend to have better battery life.* B) *Yes, the plot suggests that Brand B would tend to have better battery life.* C) *No, the plot suggests there is no real advantage to choosing one brand over the other.* Question 2) *Based on the plot presented, how different would you expect the typical battery life for the two brands to be using the following scale.* Question 2 was rated on a scale ranging from 1) *No difference* to 7) *An extremely large difference*. Once the participants had submitted their responses, they completed the demographic section.

RESULTS

Question 1. In order to determine if plot type and brand equality had a statistically significant effect on informal inference for question one, responses were recoded into a binary "correct" or "incorrect" variable. For the equality condition, participants who inferred there was no apparent difference shown in the plot between the two brands were coded "correct", and those who selected brand A or brand B were coded as "incorrect". For the inequality condition, those who selected brand B got the question "correct", and those answering otherwise were "incorrect". Figure 2 reports the percentage of correct and incorrect responses for Question 1.

Figure 2. The percentage of participants and 95% *CIs* (Wilson method) for selecting the correct inference based on whether the plots depicted equality or inequality and the type of plot viewed. *N* = number of randomised participants in each 3 x 2 factorial condition.

Binary logistic regression was conducted, using correct responses as the dependent variable, and plot type and brand equality as the independent variables. The levels for plot type were bar, box and dot, with the reference category being dot, and the levels for brand equality were equality and inequality, with inequality used as the reference category. The overall model was statistically significant, $\chi 2(5) = 73.298$, $p < .001$. There was a statistically significant main effect for both brand equality, Wald ($df = 1$) = 26.45, $p < .001$ and plot type, Wald ($df = 2$) = 10.84, $p = .004$, and a significant interaction for plot type by brand equality, Wald ($df = 2$) = 13.89, $p = .001$. The results indicated there was a significant decrease in correct inferences for the equality condition compared to inequality, *OR* = 0.02, 95% *CI* (0.01, 0.09). This suggested that participants may have had a bias towards inferring population inequality.

The main effect for plot type should be interpreted within the context of the statistically significant interaction. When the participants were presented with plots depicting inequality, the likelihood of a correct inference significantly decreased for bar plots, *OR* = 0.19, 95% *CI* (.05, .734), when compared to the reference category of dot plots. However, for the equality condition, the likelihood of a correct inference was significantly higher for bar plots, *OR* = 22.01, 95% *CI* (3.70, 131.1), when compared to the reference category of dot plots. In terms of the main effect, there was no significant difference in correct inference for box pots when compared to dot plots.

Question 2: In order to test the effect of plot type and brand equality on participants' mean ratings of the difference between battery life, a 3 by 2, between subjects, factorial ANOVA was performed using the same factors and levels as the logistic regression model. The Levene's test indicated that the assumption of homogeneity of variance was not violated, $F(5, 199) = 2.04$, $p = .075$. The factorial ANOVA revealed a statistically significant main effect for plot type, $F(2,199) = 21.21$, $p < .001$, $\eta_p^2 = .176$, and brand equality, $F(1,199) = 24.65$, $p < .001$, $\eta_p^2 = .11$, but no statistically significant interaction effect, $F(2,199) = 2.47$, $p = .087$, $\eta_p^2 = .024$. Ratings of the difference between brands were significantly higher on average when battery brand life inequality was depicted.

In order to interpret the statistically significant main effect for plot type, Bonferroni post-hoc pairwise comparisons based on the estimated marginal means from the factorial ANOVA were reported. These revealed significant differences in participants' ratings between box and bar plots ($p < .001$), and dot and bar plots ($p < .001$), but not between box and dot plots ($p = .066$). Bar plots were rated significantly lower on average than box and dot plots.

Figure 3. Mean ratings and 95% *CI*s for the difference between battery brand life for plot type and equality. Also displayed are the descriptive statistics for each condition.

DISCUSSION

The aim of this experiment was to explore how different data visualisations of the same data influence informal inference. This was examined by manipulating the degree to which a data visualisation depicted variability comparing two populations and whether or not the data were sampled from populations that had different means. In terms of choosing between brands, participants were most accurate in the brand inequality condition when presented with either a box (94.59%) or dot plot (91.18%). Participants might be drawing upon the additional visual representations of variation over the bar chart (65.7% correct) to support their decision to choose one brand over the other. This could be due the bar chart's crude display of variation (error bars showing ± 1 *SD*) or could be due to the different scale of the y-axis which was anchored at 0.

When the plots depicted data sampled from identical distributions, correct inferences were significantly lower at 46.67%, 42.86% and 17.65%, for the bar, box and dot plot, respectively. The high proportion of participants incorrectly inferring the presence of a brand effect in the equality condition suggested a bias towards rejecting the null. This effect interacted with the type of data visualisation. Observers of the dot plot in the brand equality condition were correct less than half as often as other visualisations, but almost as equally accurate as box plots in the brand inequality condition. The high level of variation depicted by a dot plot's presentation of binned raw values may have exacerbated a bias toward rejecting the null. Comparatively, the information poor bar charts showed a reverse trend, being the most accurate in the brand equality condition and least accurate in the brand inequality condition. The bias towards rejecting the null might suggest that students need to develop better visual awareness of the inherent variability in samples taken from identical distributions, especially when sample sizes are small.

Participants' ratings of the degree of difference in battery life central tendency demonstrated a similar trend. As one would expect, participants' mean ratings of difference were significantly higher across plots in the inequality condition when compared the equality condition. In addition, when participants viewed box or dot plots, regardless of brand equality or inequality, participants rated the perceived difference significantly higher than bar charts. These findings suggested that the bar charts, which were anchored at 0 and displayed only errors bars (± 1 *SD*) as an indication of variability, might cause participants to underestimate effects when they are present in the population, but lead to more accurate inference when the null hypothesis is true.

This experiment had a number of limitations. The stimulus plots were based on a single random sample for the inequality and equality conditions. Using randomly generated plots for each participant based on the underlying population distribution might impact the findings. The degree to which the sample can be considered representative was limited by the fact that participants were recruited from only three Australian universities and participation was voluntary (response bias). The closed-end questions, while efficient, might have missed unexpected insights and subtleties in students' inferences. Lastly, while common in the scientific literature, the use and anchoring of the bar chart at a value of 0, might explain the large difference in the inference reached when compared to box and dot plots.

Further research must address the major limitations of this experiment and explore more diverse data visualisations and inferential scenarios involving multiple variables. Future studies should also explore these effects outside the university student population. This preliminary experiment supports the notion that decisions about how to visualise data can have a significant influence on informal inference. Careful consideration must be given when visualising data and statistics educators must continue to optimise the use of data visualisation used to teach informal inference.

ACKNOWLEDGEMENTS

This experiment obtained institutional ethics approval from RMIT University on the 13th of July 2015 (Project No. BSEHAPP 34-15). The authors must thank the course coordinators and students who made this experiment possible.

REFERENCES

Best, L. A., Smith, L. D., & Stubbs, D. A. (2008). Detection of sample differences from dot plot displays. In *Diagrammatic Representation and Inference* (pp. 293-307). Springer Berlin Heidelberg. DOI: 10.1007/978-3-540-87730-1_27

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4(2), 75. DOI: 10.1037/1076-898X.4.2.75

Cleveland, W. S., Diaconis, P., & McGill, R. (1982). Variables on scatterplots look more highly correlated when the scales are increased. *Science*, 216(4550), 1138-1141. DOI: 10.1126/science.216.4550.1138

Dunn, P. K. (2013). Comparing the lifetimes of two brands of batteries. *Journal of Statistics Education*, 21(1). Retrieved from http://www.amstat.org/publications/jse/v21n1/dunn.pdf

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82-105. Retrieved from https://www.stat.auckland.ac.nz/~iase/serj/SERJ8(1).pdf

Paparistodemou, E., & Meletiou-Mavrotheris, M. (2008). Developing young students' informal inference skills in data analysis. *Statistics Education Research Journal*, 7(2), 83-106. Retrieved from https://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Paparistodemou.pdf

Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal*, 5(2), 27-45. Retrieved from https://www.stat.auckland.ac.nz/~iase/serj/SERJ5(2).pdf

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5-19. Retrieved from https://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2).pdf

Rubin, A., Hammerman, J., & Konold, C. (2006, July). Exploring informal inference with interactive visualization software. In *Proceedings of the Seventh International Conference on Teaching Statistics*. Retrieved from https://www.stat.auckland.ac.nz/~iase/publications/17/2D3_RUBI.pdf

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signalling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology*, 91(4), 690. DOI: 10.1037/0022-0663.91.4.690

THE IMPACT OF CURRICULUM CHANGE ON THE TEACHING AND LEARNING OF TIME SERIES AT YEAR 13

PASSMORE, Rachel and PFANNKUCH, Maxine
University of Auckland,
New Zealand
r.passmore@auckland.ac.nz

The secondary school statistics curriculum in New Zealand has experienced substantial change. The catalyst for these changes was a desire to improve students' statistical reasoning and to narrow the gap between the statistics taught and the practices and thinking of professional statisticians. Anecdotal evidence suggested the quality of Year 13 student work in time series had improved. This research sought to test whether these claims could be supported. Exemplars of student work before and after the curriculum change were analysed. To obtain a more holistic perspective of the curriculum change, teachers were also surveyed and/or interviewed. To analyse the exemplars a framework was developed based on the student data and a synthesis of established frameworks from the literature concerned with levels and development of mathematical reasoning, dimensions of statistical reasoning and interpretation of data and data displays. Analysis of exemplars against the framework provided strong evidence that after the curriculum change higher levels of reasoning were observed for time series. A major facilitator of this change was the availability of free data visualisation software, which liberated teaching and assessment from a focus on procedures to one of data interpretation and interrogation.

INTRODUCTION

The topic of time series was first introduced into the Year 13 New Zealand school curriculum in 1996. Since then it has been assessed under three different systems. The most recent and arguably the largest change was implemented in 2013. Anecdotal evidence from teachers indicated that the latest time series Achievement Standard had resulted in a huge improvement in the quality of work and level of reasoning demonstrated by students. Hence, the motivation for this research was to determine whether this anecdotal evidence could be confirmed or refuted when investigated using robust research methods. In order to pursue this research question, student work produced before and after the curriculum change would need to be assessed against some form of independently produced framework in order to provide an unbiased analysis of the level of reasoning found. An extensive search of the literature failed to discover the existence of any such framework. As a result a framework was developed as part of this research. In order to provide a more holistic view of the curriculum change, some teachers were also surveyed and interviewed in order to gain greater insight into how the change had been implemented and received by students and teachers. The development of *iNZight* (https://www.stat.auckland.ac.nz/-wild/iNZight/) also played an important role in this particular curriculum change. The new standard required full integration of context into the teaching, learning and assessment of time series through use of real data sets (http://new.censusatschool.org.nz/resource/time-series-data-sets-2012/) and the inclusion of a research component.

DEVELOPMENT OF FRAMEWORK

The development of statistical reasoning can be examined through a number of different theoretical lenses. The frameworks considered had slightly different foci but elements from each of them were utilised to establish a synthesised framework for the types and levels of statistical reasoning with time series. It is this synthesised framework that was used to determine the level of statistical reasoning in exemplars of student work from before and after the curriculum change. Similar synthesised frameworks have been developed in other areas of statistics education research, in particular, Mooney, Langrall and Hertel (2014), who developed a synthesised framework for probabilistic thinking. The frameworks considered for this research were divided into three categories

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

1. Frameworks developed for analysis of levels of reasoning in assessment tasks
2. Frameworks characterising the development of mathematical thinking
3. Frameworks characterising the development and dimensions of statistical thinking

A full review of the frameworks in these categories is outside the scope of this paper (see Passmore, 2016) but some key components from each are summarised in Table 1.

Table 1 Key components of theoretical frameworks used to develop synthesised framework

Frame-work name	*Primary focus of framework*	*Authors*	*Cognitive levels*	*Key ideas used in synthesised framework*
De Lange's Pyramid	Assessment analysis	Verhage & De Lange (1997)	Levels of reasoning – low, middle, higher Context – context free, camouflage & authentic Complexity – simple, middle, complex	Context is crucial Higher levels of thinking not accessible without integration of context.
SOLO taxonomy	Assessment analysis Mathematical reasoning	Collis & Biggs (1982)	Five modes: Sensori – Motor, ikonic, concrete-symbolic, formal & post-formal Within each mode: Pre-structural, uni-structural, multi-structural, relational & Extended abstract	Students in yr 13 at formal level but will be drawing on ikonic and concrete-symbolic modes.
Pirie-Kieren	Mathematical reasoning	Pirie & Kieren (1994)	Primitive knowing, Image making, Image having, Property noticing Formalising, Observing, Structuring Inventising	Property noticing is an important stage of reasoning in time series analysis.
Statistical thinking in empirical enquiry	Statistical reasoning	Wild & Pfannkuch (1999)	Investigative cycle, Interrogative cycle Types of thinking – transnumeration, consideration of variation, reasoning from models and integration of statistical and contextual knowledge, Dispositions	All four dimensions are present in the analysis of time series.
	Graphicacy	Konold et al. (2015)	Data as a pointer Data as a case value Data as a classifier Data as an aggregate	To interpret time series students must be at the level of viewing data as an aggregate.
	Graphicacy	Curcio (1987) Shaughnessy (2007)	Reading the graph Reading within the graph Reading beyond the graph Reading behind the graph	Higher order thinking skills are evident at the 3rd level. Proficiency at all four levels required.

Apart from the statistical thinking framework, all of the other frameworks are hierarchical in nature which tends to suggest a sequential progression in the development of student reasoning from one level to the next. Cycles were features of several frameworks but the mathematical cycle of the Pirie-Kieren framework, (Pirie and Kieren, 1994) and the statistical cycles (defined by Wild and Pfannkuch, 1999), are somewhat different in nature. A cycle in the former framework was regarded as a *mechanism* by which a student could develop higher levels of reasoning. The statistical cycle did not characterise a level of statistical reasoning but was rather a *mechanism* for empirical enquiry. Repetition of the statistical cycles may, but do not of themselves, generate higher levels of statistical thinking. The other hierarchical frameworks stress that progression is not necessarily linear but may involve revisiting lower levels of reasoning in order to progress to a higher level of reasoning. This type of development is termed *folding back* in the Pirie-Kieren framework. One fairly consistent theme across these hierarchical frameworks concerns higher order thinking skills. For students to demonstrate higher order thinking skills they need to utilise their mathematical or statistical knowledge in some unfamiliar way. Lower order thinking skills across the frameworks typically included performing calculations according to previously learned algorithms.

Table 2 shows an abbreviated version of the synthesised framework that was developed including examples of how each level of reasoning might be demonstrated by a student. These overall levels of reasoning were intended to be hierarchical in nature, but within each level a number of sub-codes (not shown) were developed which represented various ways in which a level of reasoning could be demonstrated. These sub-codes were not hierarchical. Initially 21 sub-codes were formulated but during coding these were extended to 28 to cover unanticipated student responses (see Passmore, 2016 for further details).

Table 2 Synthesised framework of the types and levels of reasoning with time series, abbreviated version

Level	*Type of Reasoning*	*Characteristics of Level*
1	Vertical Reasoning	Ability to read and understand scale, units and basic graph construction. Context may be present but not critical to task.
2	Horizontal Reasoning	Ability to look across data not just individual values, e.g. to examine trend. Context may be present but not critical to task.
3	Procedural Reasoning	Use of mathematical procedures to calculate properties of data set or to examine data features. Context may be present but not critical to task.
4	Extended Procedural Reasoning	Use of a mathematical model to predict beyond domain of current data set or to interpolate within given domain. Context may be present but not critical to task.
5	Interpretive Reasoning	Ability to interpret implications and characteristics of a mathematical model. Transnumeration. Context must be present and integrated with interpretation.
6	Interrogative Reasoning	Additional information is sought to assist in interpretation and to confirm/refute initial findings. Own analysis is critiqued and further questions posed. Integration of contextual and statistical analysis is essential.

METHODS

Ideally to analyse the levels of reasoning demonstrated by students before and after the curriculum change, the same or similar group of students would be taught then assessed under the old and the new achievement standards. Educationally and ethically this was not an option. Alternatively, but still in an ideal world, a representative sample of student work produced before and after the curriculum change could be analysed. Access to such examples of student work proved impossible to obtain and a compromise solution was sought. Exemplars of student work posted online by the New Zealand Qualifications Authority (NZQA) as a resource for teachers and freely available were used as a proxy for the ideal representative samples. A total of 30 student exemplars were examined, 15 from before the curriculum change and 15 from after. The exemplars contained the achievement levels Achieved, Merit and Excellence. Four exemplars that did not achieve the standard were also available but were not included in the analysis. Data from teachers were also obtained but these teachers had no connection to the students who created the exemplars. A total of 18 teachers answered a survey sent to all Year 13 statistics teachers at 8 different schools. Five lead statistics teachers from five of these schools were then interviewed. Both quantitative and qualitative data were collected. This mixed methods approach was adopted in order to gain a more holistic perspective of the teaching of time series in New Zealand secondary schools rather than solely an analysis of student responses to assessments.

RESULTS

Student exemplars

Each of the 30 student exemplars was coded according to the synthesised framework. Every calculation, comment and equation was categorised. The most striking difference in the student exemplars was the substantial increase in the quantity of literary statements made. This resulted in an entirely different style of response when compared with the calculation dominated student responses completed for the old standard. 336 literary statements were made under the new standard compared to only 81 under the old standard. Given this disparity in the frequency of literary statements, these new types of statement were regarded as a separate but new component. Four new codes were created to code such statements but these codes were *not included* in any analysis of results since their inclusion skewed comparison of the original levels of reasoning. A shift to a higher overall level of reasoning was demonstrated in the exemplars produced after the curriculum change. This shift was observed at all levels of achievement but was particularly noticeable at the highest achievement level, Excellence. Table 3 summarises the mean levels of reasoning in the student exemplars by achievement level. No examples of the lowest level of reasoning, vertical reasoning, were found in the student exemplars that received an Achieved or higher grade.

Table 3 Comparison of mean levels of reasoning in student exemplars between old and new standards

		Number of codes at each level of reasoning					
	No. Exemplars	2	3	4	5	6	*Mean*
		Old Achievement Standard					
Achieved	4	1	15	0	0	0	*2.94*
Merit	6	9	34	30	8	0	*3.46*
Excellence	5	3	25	16	3	0	*3.40*
		New Achievement Standard					
Achieved	6	35	28	27	16	0	*3.23*
Merit	6	29	20	38	34	0	*3.64*
Excellence	3	28	18	26	29	22	*3.99*

The differentiation between levels of reasoning by achievement levels was greater in the new standard, particularly between Merit and Excellence. Under the old standard the two higher levels of achievement had similar mean levels of reasoning. The highest level of reasoning, interrogative reasoning was only observed in exemplars completed for the new standard. Figure 1 shows that the levels of reasoning found in exemplars from the new standard demonstrated greater diversity in levels of reasoning than exemplars from the old standard.

Another change observed was an increase in cases of horizontal reasoning, level 2, in exemplars from the new standard. In the light of the overall increase in level of reasoning this presents an apparent paradox. This can be explained using De Lange's pyramid one of the assessment frameworks used to create the synthesised framework. Lower level reproductive reasoning assessment tasks, in this framework, range from simple to complex; horizontal reasoning fits into this category. The ability to perform procedural type tasks, from simple to complex, lay the foundation for access to higher levels of reasoning. A time series whose overall trend is clearly linear represents a relatively simple task to describe. Under the old standard, such a scenario was common, with teachers often using simulated data sets that would create a well-behaved trend. However, the statistics curriculum as a whole in New Zealand now encourages teachers to use real data sets in all aspects of their teaching and assessment. Confronting students with these real data sets has meant that the task of describing the overall trend has become much more complex, although the level of reasoning required is unchanged.

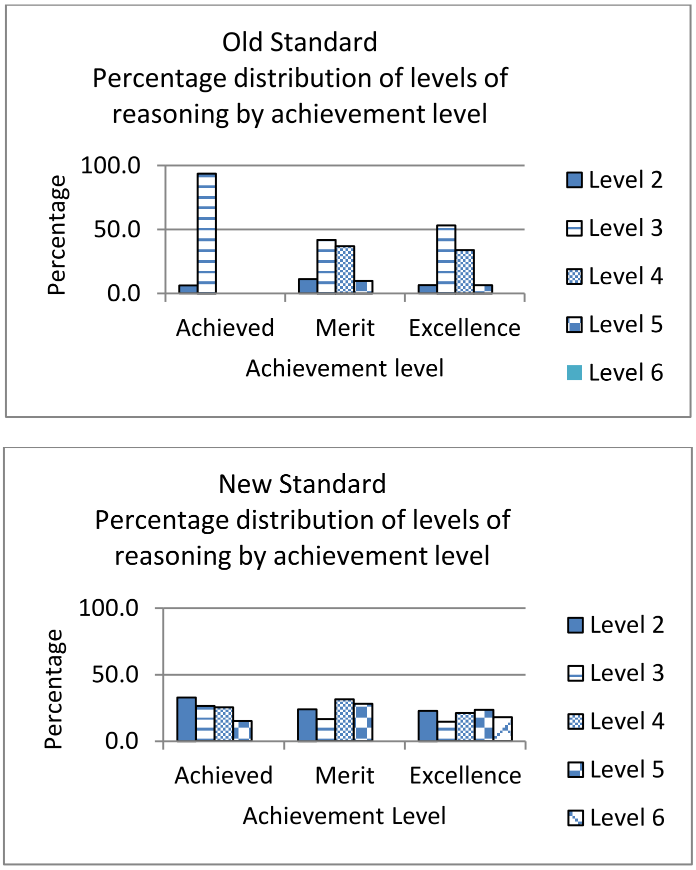


Figure 1 Comparison of reasoning levels across achievement levels and by standard

Teacher perceptions

Teachers reported some substantial changes in the management of the assessment of this standard. All but one school had shifted from a task completed under exam conditions in class to one completed in a mixture of class and own time or wholly in own time. Under the old standard most teachers had used EXCEL to complete repetitive calculations, but all teachers surveyed had now switched to using *iNZight*, although some still used EXCEL as well. The majority of teachers surveyed reported that the quality of student work under the new standard was either much or slightly improved. Only one teacher thought the quality was lower. Many teachers commented on the greater connection their students enjoyed with the context of the data under the new standard.
One teacher noticed her students

"*get a more in-depth knowledge rather than a knowledge that is gone as soon as the test is over*"

However, some teachers felt the black-box approach of *iNZight* which automated the calculations required by the old standard had been detrimental to student understanding.

"*Calculation of trends, seasonal effects etc. helped student understanding better*"

Use of *iNZight* does not preclude the inclusion of some *by-hand* activities such as calculation of trend or smoothed data values. Indeed there is strong research evidence to suggest that such activities conducted before the introduction of new technology is beneficial to students (e.g. Arnold, Pfannkuch, Wild, Regan & Budgett, 2011). Teachers reported few technological issues or problems with access to computers. However, they did report a lack of useful text-book resources, increased marking work load, lack of moderation feedback and difficulties in sourcing good time series data sets.

CONCLUSION

The research methods employed were not those envisaged as ideal, thus any reader of this research should be cognisant of its limitations. The student exemplars used were not a randomly selected representative sample of student work and neither were the teachers randomly selected. The exemplars were provided by NZQA and the teachers were selected from schools which had participated in past professional development events; their views may differ markedly from those of teachers who do not attend such events. The quality of the survey data from teachers may also have been affected by questionnaire fatigue and the interviews by interviewer bias. The research also suffered from non-response bias with only 62% of the questionnaires returned completed. The findings of this research should consequently be treated with caution.

Analysis of the student exemplars revealed strong evidence that after the curriculum change students were exhibiting higher levels of reasoning at all levels of achievement. One of the major facilitators of this change was the availability of free data visualisation software, which liberated teaching and assessment time from a focus on procedures to one of data interpretation and interrogation. Some teachers voiced concern about the absence of procedural calculations, suggesting that this had adversely affected student understanding. Although this is a legitimate concern it can be resolved easily by a change in pedagogical practice; including some by-hand calculation activities before analysing time series using technology. Differences in the style and management of the assessment were also instrumental in effecting this change. The style change involved an increase in literary statements. Student responses changed from a presentation of a series of calculations with some commentary to a more formal report style with an introduction followed by sections on results, analysis and conclusions. After the curriculum change most schools surveyed had moved to a project-style assessment, which allowed students time to pursue research of similar time series or read relevant articles. This additional time could provide the potential for students to fully integrate their statistical and contextual knowledge.

The curriculum change which is the focus of this research is only one of many introduced in the New Zealand statistics curriculum since 2010; further research is recommended to investigate whether similar changes in the learning outcomes of students have occurred in other new Achievement Standards.

REFERENCES

Arnold, P., Pfannkuch, M., Wild, C., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: From hands on to movies. *Journal of Statistics Education, 19*(2), 1-32.

Collis, K., & Biggs, J. (1982). *Evaluating the quality of learning - the SOLO taxonomy*. New York, NY: Academic Press.

Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*(5), 382-393.

Konold, C., Higgins, T., Russell, S., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305-325.

Mooney, E. S., Langrall, C. W., & Hertel, J. T. (2014). A practitional perspective on probabilistic thinking models and frameworks. In E. Chernoff, & B. Sriraman (Eds.), *Probabilistic thinking* (pp. 495-507). Dordrecht, The Netherlands: Springer.

Passmore, R. (2016), *The impact of curriculum change on the teaching and learning of time series*, Auckland, NZ: University of Auckland. Retrieved from http://hdl.handle.net/2292/29102

Pirie, S., & Kieren, T. (1994). Growth in mathematical understanding: How can we characterise it and how can we represent it? *Educational Studies in Mathematics, 26*(2-3), 165-190.

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

Verhage, H., & De Lange, J. (1997). Mathematics education and assessment. *Pythagoras*, 14-20.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-265.

TEACHERS' PERCEPTIONS OF TEACHING STATISTICS IN SCHOOLS USING THE ISLANDS

HUYNH, Minh[1], BAGLIN, James[2], HART, Claire[2], MacGILLIVRAY, Helen[3], BULMER, Michael[4], DUNN, Peter K.[5] and MARSHMAN, Margaret[5]
[1]Swinburne University of Technology, [2]RMIT University of Technology, [3]Queensland University of Technology, [4]University of Queensland, [5] University of the Sunshine Coast
mhuynh@swin.edu.au

In 2014, the Australian Maths and Science Partnership Program (coordinated by the Australian Government Department of Education) funded the national Islands in Schools Project. The purpose of this project was to improve junior secondary school students' engagement with statistics and to provide maths and science teachers with innovative, classroom-ready resources for teaching statistics through data investigations. The learning activities developed were based upon the Islands, a free virtual human population. Professional learning resources to assist teacher preparation for using the resources were also provided. A total of 12 partnered secondary schools from across Queensland, Victoria and Western Australia participated. This led to 73 maths and science teachers and in excess of 1,900 students from years 7 to 12 participating in the project. This paper reports the findings from feedback provided by 26 teachers involved in the implementation of the Islands-based activities. The findings provide valuable insight into the teachers' experience of using the resources in their classrooms and the major challenges they faced with engaging students in statistics. The findings are discussed within the context of refining the resources, the challenge of teaching statistics through data investigations in schools, and improving teacher professional learning for statistics education.

INTRODUCTION

According to the Australian Industry Group (2013), the current supply of science and mathematics graduates necessary to keep Australia economically competitive does not meet the demand set by government, industry and education sectors. In 2015, a report by the Department of Organisation for Economic Co-operation and Development (OECD, 2015) indicated that only 0.46% of all Australian graduates completed with a mathematical or statistical background, in comparison to the OECD global average of 1.19%. Prieto and Dugar (2016) claim this is due to fewer students electing to study intermediate or advanced mathematics in senior secondary school (Kennedy, Lyons & Quinn, 2014).

Therefore the challenge to engage students in mathematics and statistics is of critical importance in middle school. This is increasingly important for statistics, which is needed across all disciplines, for general statistical literacy, and its increasing roles in the current explosion of analytics and big data. This importance is reflected in the emphasis on statistics and statistical thinking in school curricula around the world, including the Australian curriculum F-10. In statistics, both educational researchers and statisticians have long claimed that the teaching of statistics must reflect the practice of statistics (see Gibbons and MacGillivray, 2014, for extensive references) and that student attitudes drive engagement in learning and determine later learning and use of skills (Ramirez et al., 2012). Hence if students are not positively engaged during middle school statistics classes, they will be unlikely to develop constructive attitudes towards statistics, including reluctance to pursue further statistics study, no matter how attractive the career prospects.

The Office of the Chief Scientist report (2012) raised concerns around the teaching of mathematics and science in schools and recommended that schools and universities develop innovative approaches that bring the "practice" and relevancy of mathematics and science into the classroom to improve student interest. Such recommendations are in particular alignment with the long-time advocacy of experts in statistics education that learning through the full process of *data investigations* not only aligns with real world statistical practice but also provides the ideal vehicle for developing statistical thinking, knowledge and skills as well as satisfying engagement for students and teachers (MacGillivray & Pereira-Mendoza, 2011). However, practical implementation of this advocacy requires statistically and pedagogically sound teacher preparation and support and classroom-ready resources. This is especially so for the middle school curriculum and diverse student and teacher populations.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

This gave rise to the *Islands in Schools Project*, which was funded by the Australian Maths and Science Partnership Program (coordinated by the Australian Government Department of Education), to improve junior secondary school students' engagement with statistics and to provide teachers with innovative, classroom-ready resources for teaching statistics through "real" data investigations. Presently, the types of data collection activities that teachers can employ are often restricted by ethical and practical constraints.

The current project proposed to utilise the *Islands* software (Bulmer and Haladyn, 2011), in conjunction with innovative, classroom-ready resources, to teach statistics through data investigations. The *Islands* (http://islands.smp.uq.edu.au) is a free, online, human population simulator that provides students with a virtual environment to explore data investigation tasks. The program can be navigated by clicking between the 27 villages, across three unique islands, that are home to approximately 34,000 virtual residents. Simulation models govern the population's births, deaths, health, social lives and residency. Each resident has their own personal story and genetic code that are linked to their appearance and responses to over 200 tasks governed by realistic models. These tasks include a range of surveys, blood tests, physiological measures, mental tasks and exercise. By utilising the *Islands*, teachers and students can propose questions, design and carry out data collection, use their newly-acquired statistical knowledge to explore data and comment on issues, thus allowing for more effective teaching of the data investigation process and, therefore, of statistics.

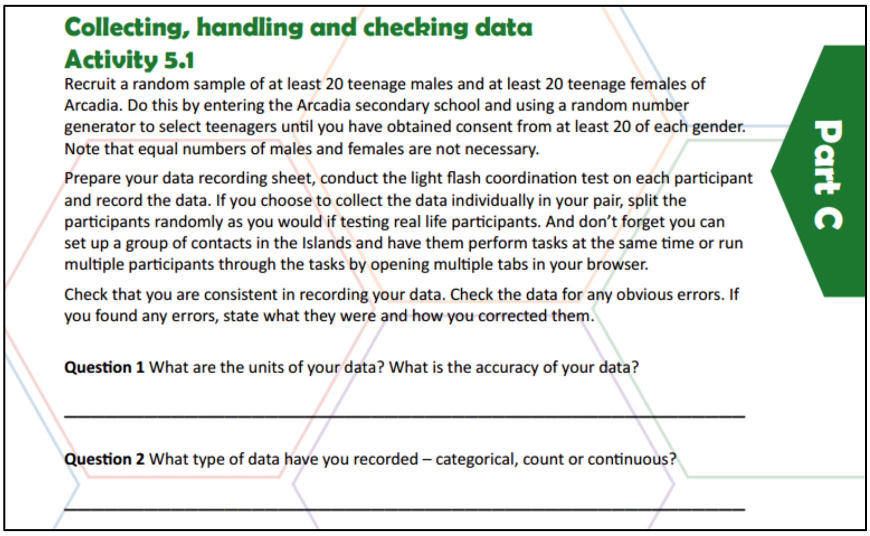
In addition to classroom-ready resources, there is an urgent and strong need to provide support for the educators who teach statistics, and examine how their practices are influenced by the Australian Curriculum. Even though statistics is a dedicated discipline, in Australia and internationally it cannot be an independent subject in the primary and junior school curriculum, and is usually taught as part of mathematics so that all students have a common grounding. In addition, the collection and interpretation of data features strongly within science. Naturally, this has substantial implications for the pre and in-service training of all secondary school teachers. In recent works, researchers have suggested that many teachers feel that their students experience difficulty in learning statistics, and that they themselves are not adequately trained to help their students overcome these challenges (Batanero, Burrill and Reading, 2011). Furthermore, teachers reported difficulty when implementing an experimental approach to teaching probability or through statistical investigations (Stohl, 2005). This is because few teachers have prior experience using statistical investigations to conduct experiments or simulations.

According to Batanero and Diaz (2010), many prospective secondary teachers, even those who have majors in mathematics, have not received specific preparation in designing data collections or experiments, analysing data from real applications, or adequately using statistical software. Experiencing data investigations firsthand provides the best professional development for not only the teaching of statistics through data investigations, but also for developing and consolidating statistical understanding, so that they may better teach statistics in their classrooms. As such, one of the primary aims of the *Islands in Schools project* was to provide school-level mathematics and science teachers with professional development training material relating to the use of the *Islands* within the classroom. Hence, as well as developing a suite of classroom-ready resources, including model assessment questions and marking rubrics, aligned with the Australian Curriculum for years 7-10, the project team conducted workshops for teachers, developed online professional development materials, and surveyed teachers and students in the participating schools. This paper reports findings on how the *Islands* resources influenced teachers' confidence and competence towards using data investigations in teaching statistics within the Australian Curriculum.

METHOD

Twelve partnered secondary schools were recruited from across Queensland, Victoria and Western Australia. This led to 73 maths and science teachers and in excess of 1,900 students from years 7 to 11 participating in the project. In most instances, the teachers arranged for intensive face-to-face training to be delivered by members of the project team. These sessions typically included two parts. Initially, the project team member would outline the project, providing background information pertaining to: (1) Statistics in the Australian context, (2) Statistics in the Australian Curriculum, (3) the statistical data investigation process and, (4) teaching data investigations. In the second part of the training, the teachers were taught how to use *The Islands* and the project resources. Exemplars of how to complete selected *Islands* activities were provided. In the event that teachers could not attend the workshops, or the project team was restricted due to reach, teachers were provided with online learning resources to guide preparation. These resources included professional learning material such as detailed readings about teaching statistics within the Australian Curriculum, self-guided tutorials for using the *Islands* and exemplar datasets and answers to assist teachers' self-preparation.

After completing the training workshops, the teachers were encouraged to access and view the *Teacher Resources* content from the website. The *Teacher Resources* section contained a series of activities designed for students to engage in data investigations on the *Islands*. In total, there were ten curriculum-aligned, classroom-ready activities for Years 7 – 10. Each activity included a detailed teacher summary, learning objectives, curriculum alignment outline, student activity sheets, and suggested marking rubric. The activities can be freely accessed (without a sign in) via the *Islands in Schools project* website (http://www.islandsinschools.com.au/). The classroom activities contextualised statistics by having students engage in learning through virtual data investigations (initial questions, issues and planning, collecting, handling and checking data, exploring, interpreting data in context). Figure 1 provides a screenshot example which outlines part of one of the Island activities. After examining all of the *Islands* activities, the teachers were able to select the task(s) they felt were most appropriate for their respective classes, and save / print them directly from the website.


Collecting, handling and checking data
Activity 5.1

Recruit a random sample of at least 20 teenage males and at least 20 teenage females of Arcadia. Do this by entering the Arcadia secondary school and using a random number generator to select teenagers until you have obtained consent from at least 20 of each gender. Note that equal numbers of males and females are not necessary.

Prepare your data recording sheet, conduct the light flash coordination test on each participant and record the data. If you choose to collect the data individually in your pair, split the participants randomly as you would if testing real life participants. And don't forget you can set up a group of contacts in the Islands and have them perform tasks at the same time or run multiple participants through the tasks by opening multiple tabs in your browser.

Check that you are consistent in recording your data. Check the data for any obvious errors. If you found any errors, state what they were and how you corrected them.

Question 1 What are the units of your data? What is the accuracy of your data?

Question 2 What type of data have you recorded – categorical, count or continuous?

Part C

Figure 1: Screenshot of part of one of the *Islands* activity worksheets

Following teacher preparation, each school partner and their respective teachers implemented the *Islands*-based activities in their classrooms. The number of activities varied by school, year level, and the time available; all teachers implemented at least one activity in their class. Following a school's implementation, an online, anonymous questionnaire was sent to teachers to complete. The questionnaire contained 12 items focusing on how teachers felt regarding professional development material, in conjunction with 11 items examining how teachers perceived the Island activities operated within their classrooms. Example items include: *I felt confident after completing the professional development that the Islands in Schools Project activities would meet the students' curriculum needs*, and, *I am confident that I can implement the Islands in Schools Project's activities effectively in future classes*.

In addition, there were four open-ended questions that asked teachers to provide feedback relating to their experience of the professional development and implementation of the *Islands* activities in class. Teachers were asked to share both positive feedback and areas they perceived to be in need of improvement. The questions were as follows: 1) *Please share some positive feedback or experiences of the professional development*. 2) *Please share some feedback on the areas of the professional development that are in need of improvement. This feedback will help inform future development*. 3) *Please share some positive feedback or experiences of the classroom implementation of the Islands in Schools Project's activities.* 4) *Please share some feedback on the areas of the activities that are in need of improvement. This feedback will help inform future development.*

RESULTS AND DISCUSSION

Overall, the responses showed a mostly neutral–positive perception towards the professional development training. The highest rated items, based on a five-point scale ranging from 1) *Strongly disagree* to 5) *strongly agree*, indicated that the *Islands* activities met the students' curriculum needs (*Mean [M]* = 3.96, *Median* [*Md*] = 4, *Agree* = 82.6%), that the *Islands* activities were more engaging than what they would typically use or plan to use for teaching statistics (*M* = 3.96, *Md* = 4, *Agree* = 87.0%), and that the professional development helped them develop new approaches to teaching statistics through data investigations (*M* = 4.00, *Md* = 4, *Agree* = 78.3%). The lowest-rated items related to the professional development taking too long to complete (*M* = 3.04, *Md* = 3, *Agree* = 39.1%), improving their understanding of teaching statistics through data investigations (*M* = 3.35, *Md* =3, *Agree* = 39.1%), and improving their confidence in teaching statistics through data investigations (*M* = 3.39, *Md* =3, *Agree* = 43.5%). However, these lowest-rated items revealed more neutral responses (*Neither Agree nor Disagree*), with averages close to ratings of 3, rather than negative. This suggests that teachers might require much more intensive and directed instruction to improve their general knowledge of teaching statistics.

Similarly, responses pertaining to the *Islands* activities were also mostly neutral–positive. The highest rated items related to teachers believing that the activities were more engaging than what they typically use or would plan to use for teaching statistics (*M* = 4.13, *Md* = 4, *Agree* = 83.3%) and feeling confident that they can implement the activities effectively in future classes (*M* = 4.17, *Md* = 4, *Agree* = 87.5%). The lowest rated items related to the *Islands* activities taking too long to implement in class (*M* = 3.21, *Md* = 3, *Agree* = 45.8%), which was a positive result given the negative wording of the statement, improving teachers' confidence in teaching statistics through data investigations (*M* = 3.38, *Md* = 3, *Agree* = 41.7%), and improving teachers' understanding of teaching statistics through data investigations (*M* = 3.54, *Md* = 4, *Agree* = 54.2%).

To dig deeper into the factors influencing teacher ratings, the qualitative data were analysed using a six-step inductive thematic analysis method described by Braun and Clarke (2006). The six steps are: (1) data familiarisation, (2) initial code generation, (3) theme searching, (4) theme revision, (5) theme definition and naming, and (6) reporting. The themes identified from the thematic analysis (see Figure 2) are reported and discussed below.

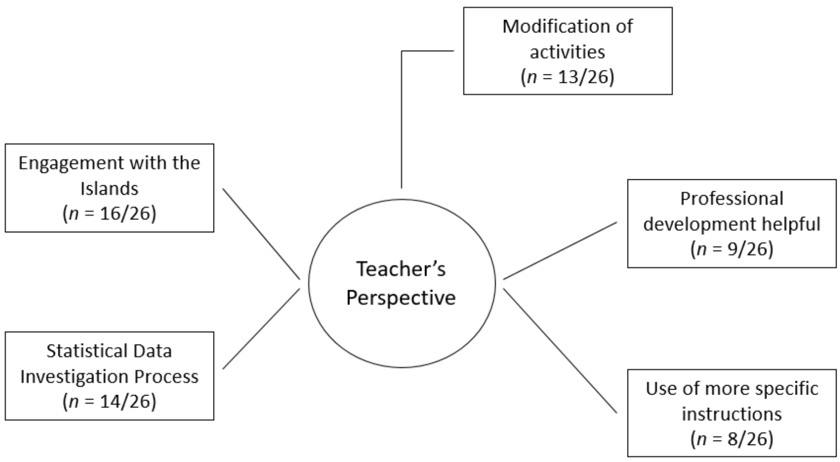


Figure 2: Thematic map of the teachers' responses

1) Statistical Data Investigation Process: There were 14 responses that positively discussed the ability of the *Islands in Schools Project* to effectively teach the statistical data investigation process to teachers and students. Teachers referred to specific stages in the process (i.e. initial questions, experimental design, data collection and analysis, and interpretation of data in a broader context) and understanding the concept of sample variation. This feedback applied to both the professional development and the implementation of *Islands* activities and highlights the potential of *Islands* learning resources to engage teachers and students in all stages of the statistical data investigation process. For example:

> *"The most important message… shared during the PD was that we usually only perform one part of the process of statistics (calculating/analysing data) and that it's almost meaningless without being involved in designing an experiment and getting the data yourself."*

2) Use of more specific instructions: Despite the provision of the online resources, there were 8 responses indicating that instructions provided throughout the project could have been more specific. This included those instructions provided to teachers (e.g. using coordinator tools, setting up classes) and those included in *The Islands* activities (e.g. students experiencing difficulty with the technology, students struggling to understand the task). For example:

> *I would have liked more explanation of the settings and setting up of class groups and knowledge of those tests that really should be done in school. I would like more control over the setting up of classes.*

One teacher suggested the use of a video tutorial for teachers, "*Maybe a video tutorial of how to get a move around the islands and deal with common problems would be useful".*

3) Engagement with the *Islands*: This theme related to positive feedback regarding the level of engagement with the *Islands* learning resources. This was the most evident theme, being expressed by a total of 16 teachers. Teachers described the *Islands* as '*interesting'*, '*fun'*, '*engaging'*, and '*so realistic*', and also commented that students '*really enjoyed the work'*, were '*engaged and enthusiastic*', and that '*even the students who are easily distracted in class were loving working on the modules*'. For example, one teacher commented, *"The student engagement was phenomenal. No pupil was off task during our computer lessons."* Teachers referred to specific aspects of the *Islands in Schools Project* that they found engaging. This included data collection using the *Islands,* interacting with the *Islanders,* learning the history of the *Islanders,* the variety of tasks available, the gamified nature of the *Islands,* and the realistic challenge of obtaining consent from *Islander* participants.

4) Professional development/resources helpful: There were 9 responses that described the professional development/resources as being helpful to teachers. Teachers described the professional development as '*informative*', '*easy to locate*', '*useful'*, and '*easy to follow'.* An example response was:

> *I found the professional development easy to follow and I learnt quite a lot about the Australian Curriculum and what the expectations were for skills and concepts to be taught to the different year levels. I thought the slide shows were particularly useful.*

While it was clear that some teachers reported increased knowledge of teaching statistics within the Australian curriculum, it was not clear the degree to which this knowledge extended to teaching statistics through data investigations. The open-ended questions may have failed to probe deep enough into teacher's prior pedagogical knowledge and confidence.

5) Modification of activities: Teachers discussed modifications that could be made to the *Islands* activities to help improve the quality of the learning experience for students. Several of these 13 responses pertained to reducing the time required for data collection (e.g. '*clunky and slow'*, '*very time consuming'*). These teachers discussed the issue of *Islanders* denying consent; however, this is reflective of the realistic nature of the *Islands* (i.e. consent must be obtained from *Islanders* prior to them participating in research). Interestingly, teachers appeared to differ in terms of whether the activities require further depth or whether they are currently too difficult or easy for students; this is most likely to be reflect prior experience in different states – for example, Queensland teachers are very accustomed to investigative projects:

I like the way in which students collected their data but the tasks in which they needed to display and comment on their data was ok for beginners but lacked the extra depth of skill that real analysis (or even more mature analysis) could investigate.

We need to be told that there could some flexibility with the program as some of the activities and questions were above some of our students' capability.

CONCLUSION

One of the aims of the *Islands in Schools Project* was to improve teachers' knowledge and confidence with teaching statistics through data investigations. Feedback from teachers indicated that the project provided valuable resources and training that enhanced their knowledge and ability to use technology to engage junior secondary schools' students in learning statistics through data investigations. Teachers perceived their students to have engaged well with the *Islands*-based activities and agreed the resources were an improvement on previous material used to cover the statistics curriculum. However, the findings also suggest that without further education and professional development opportunities for teachers, such resources might be limited in their ability to impact students (Batanero, Burrill and Reading, 2011). By the end of the project, teachers still rated themselves as lacking general knowledge and confidence in teaching statistics. The feedback provided numerous suggestions for improvements to the project's resources and highlighted the major challenges that teachers are likely to face during implementation. The teacher insights have proven invaluable for planning future efforts that will address the professional development needs of Australian teachers for teaching statistics across the curriculum.

REFERENCES

Australian Industry Group. (2013). *Lifting our science, technology, engineering and maths (STEM) skills*. Australian Industry Group. Retrieved from http://www.aigroup.com.au/policy/reports

Batanero, C., & Díaz, C. (2010). Training teachers to teach statistics: What can we learn from research? Statistique et enseignement, 1(1), 5-20. Retrieved from http://statistiqueetenseignement.fr/ojs/

Batanero, C., Burrill, G., & Reading, C. (2011). Teaching statistics in school mathematics - Challenges for teaching and teacher education: A Joint ICMI/IASE Study (pp. 407-418). Dordrecht, The Netherlands: Springer.

Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. Qualitative Research in Psychology, 3 (2). pp. 77-101. ISSN 1478-0887 Available from: http://eprints.uwe.ac.uk/11735

Bulmer, M., & Haladyn, J. K. (2011). Life on an Island: A simulated population to support student projects in statistics. *Technology Innovations in Statistics Education*, *5*.

Gibbons, K. and MacGillivray, H.L. (2014) *Education for a workplace statistician*. In MacGillivray, H.L., Martin, M. and Phillips, B. (2014) (eds) *Topics from Australian Conferences on Teaching Statistics: OZCOTS 2008-2012*, (pp267-294) Springer Science+Business Media, LLC, New York.

Kennedy, J., Lyons, T & Quinn, F. (2014). The continuing decline of science and mathematics enrolments in Australian high schools. *Teaching Science*, *60*, 34-46.

MacGillivray, H., & Pereira-Mendoza, L. (2011). Teaching statistical thinking through investigative projects. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICM/IASE Study: The 18th ICMI Study* (pp. 109–120). New York, NY: Springer Science+Business Media B. V.

Office of the Chief Scientist. (2012). *Mathematics, engineering & science in the national interest*. Canberra, ACT.

Organisation for Economic Co-operation and Development (OECD). (2015). OECD Stat Extracts. Retrieved May 1, 2015, from http://stats.oecd.org/

Prieto, E., & Dugar, N. (2016). An enquiry into the influences of mathematics on students' choice of STEM careers. *International Journal of Science and Mathematics Education*, 1-20

Ramirez, C., Schau, C., & Emmioğlu, E. (2012). The importance of attituds in statistics education. *Statistics Education Research Journal*, *11*(2), 57–71.

Stohl, H. (2005). Probability in teacher education and development. In G. Jones (Ed.). *Exploring probability in schools: Challenges for teaching and learning* (pp. 345-366). New York: Springer

A REPORT COMPARING METHODS USED TO TEACH REGRESSION ANALYSIS TO NON-STATISTICIANS

LEE, Sophie and WADE, Angie
Centre for Applied Statistics Courses,
Institute of Child Health, UCL, London, UK
sophie.a.lee@ucl.ac.uk

This paper compares 3 methods of teaching introductory regression analysis to non-statisticians from diverse backgrounds. The Centre for Applied Statistics Courses at UCL has developed regression courses that are structured differently and use software to varying degrees. 'Logistic Regression' is taught on a single day in a traditional classroom, is completely theory-based and involves no software. 'Regressions with R' runs over two shorter days and allows delegates to apply the theory being taught in the course to datasets using R software. This course is held in a computer room where everyone has their own computer to apply the theory just taught. There is complete integration between the theoretical sessions and the immediate practical applications interspersed. 'Introduction to Regression' combines the two ideas, with one short day teaching the theory of regression, followed by a second day interpreting SPSS output from an example application. The dataset is provided for delegates to go away and practice replication if they wish to, with whatever package they may prefer. Feedback and questionnaire results from course delegates to each of the 3 types of course will be used to investigate which method appears more effective in providing them with useful tools for their own research.

INTRODUCTION

The Centre for Applied Statistics Courses (CASC), based in UCL's Institute of Child Health, comprises a small team that organise and teach a range of short-courses to non-statisticians. The audience for these courses is diverse, with delegates coming from a wide range of backgrounds in terms of discipline, profession, and seniority. CASC currently delivers 16 different short-courses, and delegates attending these courses are asked to complete a short feedback questionnaire to assess how suitable the course was to their needs, and how future courses could be improved.

The understanding of, and ability to build, regression models is a cornerstone for any quantitative researcher. Hence it is important that individuals whose research conclusions are based on the analysis of a collation of sampled data learn about the technique. Included in the programme of courses we deliver are a range of introductory regression courses; these courses were developed at different times and utilise slightly different approaches with respect to the breakdown of classroom-based theory and software application. This paper compares feedback provided by students who have attended one of three CASC introductory regression analysis courses: 'Introduction to Regression', 'Introduction to Logistic Regression' and 'Regressions with R'. Feedback from these courses is presented in the context of andragogic research in order to discuss the optimum method to teach regression analysis to non-statisticians, so that they may best comprehend the theory underpinning regression analysis, retain this information, and appropriately apply it in real-world contexts.

THE COURSES

'Introduction to Regression Analysis' was the first introductory course developed and delivered in 1998, before CASC was formally founded, in response to a large number of enquiries from researchers within the Institute of Child Health regarding the application of regression analysis. When this course was first designed, there was only a single computer lab within the institute which was often not available for teaching in its entirety and very few delegates would own a laptop, making hands-on practice of the methods infeasible. To overcome the lack of resources, the notes contained output and interpretation of an example analysis which had been carried out using the statistical package SPSS. Additional to the lack of resources, there was a belief that learning away from the computer screen may be more effective as software operation may detract from the course aims of instilling understanding of the technique and application.

The basic format of the 'Introduction to Regression Analysis' has not changed since its inception, although the content has been updated. It consists of two short days (7.5 teaching hours in

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

total) with the first day devoted to the theory of linear regression, including interpretation of model coefficients, assessing goodness-of-fit, prediction and validation. During the second day of the course, an example dataset is introduced and output presented from an analysis carried out using SPSS. Group discussion is used throughout day two to aid delegates' understanding of the methods taught on day one. The example dataset introduced during the course is made available for delegates after the course so that they are able to apply the taught methods using the software of their choice. There is no hands on demonstration or application of any software, even by the presenter.

'Introduction to Logistic Regression' was developed in 2005 as an additional one day (5 hours) course to complement the existing 'Introduction to Regression Analysis' course. Although this was 7 years after the first course was produced, computer facilities for teaching large groups were still somewhat limited, so hands-on practice of the models on the day was again discounted as an option. The Logistic Regression course also included SPSS outputs, but in contrast to the earlier course, these were interspersed in the notes and discussion of them integrated with the formal training of the theory. Delegates are taught how logistic regression can be applied to explain the variation in the probability of an outcome occurring and how to interpret and utilise model diagnostics produced by the computer software. The course also contains a supplementary section that explains how this method can be extended to deal with ordinal and nominal outcomes.

'Regressions with R' is the newest of the three regression courses currently offered by CASC, and was developed in 2014. By the time 'Regressions with R' was first run, teaching facilities with access to computers and ICT support was readily available making a software-based course feasible. These resources were utilised to create a teaching environment different to the other introductory courses, allowing delegates to not only learn the theory of the methods but also practice using software to apply them to real data. The course is taught over the same timescale as the 'Introduction to Regression Analysis' (two short days) in a computer room in which each delegate has their own computer and access to R software. The theory that is covered in the course is supplemented with examples of R code and output, which aim to explain key ideas and demonstrate how these methods can be applied to real-world problems. Linear, logistic, ordinal logistic, poisson, negative binomial, Cox proportional hazard and multilevel regression models are covered. Importantly, the course also includes instruction regarding interpretation of models, model diagnostics, and good practice when fitting regression models. Each type of regression is introduced using an example dataset on which the teacher demonstrates analyses using R software. Delegates are then asked to complete practical exercises, based on the theory just covered. Hence the computer interaction is interspersed throughout the course, with alternation between taught theory and hands-on practice.

COURSE FEEDBACK

Since 2010, following each course that is run by CASC, all delegates who attend are sent an online feedback request. The feedback form includes multiple questions regarding content, presentation, usefulness and the organisation of the course. For each question, delegates are asked to respond on a 5-point likert scale from (1) *not at all* to (5) *very much.* Respondents are also asked several open-ended questions to give them the opportunity to explain their scores and/or make more detailed suggestions about how they believe the course could be improved. These feedback forms are intended to inform future courses and give indications of when courses should be changed or are missing anything of importance. We present here selected quotes from all courses on which electronic feedback was available together with a more detailed quantitative assessment of those within the current year (2016).

Introduction to Regression Analysis

This flagship course has run 27 times since 2010 and the feedback was typically completed by approximately half of the delegates that attend this course. Comments frequently referred to the absence of practical application involved in the course. Some delegates suggest that the instructor could/should carry out the analysis as a live demonstration during the class:

"…*would be good to have time to practice example live on SPSS and then trouble shoot with trainers*"

In contrast, others implied that hands-on practical sessions would be preferable, where each delegate has access to their own individual computer:

"*A stats course must involve a practice component running analysis in a statistical package. The practice stuff involved only theoretical questions, print outs and slides".*
"I would have liked to use SPSS simultaneously when doing the regression analysis rather than follow handouts".

Although most groups contained at least one comment regarding the lack of practical application within the course, the majority of comments were positive and the use of example outputs, even without hands-on demonstration of the software, was appreciated:

"*…the course was extremely informative. The SPSS examples were very useful and I was impressed with the depth that the course when into in relation to SPSS outputs - I have not experienced this on other courses*".

In 2016 the course has so far run 4 times with a total of 90 people attending of whom 59 (66%) provided feedback. The responses received were overwhelmingly positive: 45 (76%) of the delegates that responded to feedback rated the course overall 'very good' (5), 11 (19%) rated it 'good' (4), the remaining 3 (5%) rated the course 'average' (3).

<u>Introduction to Logistic Regression</u>

Since 2010 this course has been run 24 times with most feedback expressing satisfaction regarding the level at which the course was pitched, and the way in which it was structured. However, some delegates suggested that more practical sessions should be included to help consolidate their knowledge. Since 2015, the course has received an increase in the number of suggestions for the inclusion of a software element to the course. One delegate reported that he/she felt that "*it was a bit too abstract*", while another suggested changing the format of the course, and felt that running it "*as a 2-day course with some practice activities in SPSS would be better*".

However, the most recent feedback from Logistic Regression in July 2016 contained two individuals who felt that some parts of the course were irrelevant as they were based on SPSS:

"*I use STATA rather than SPSS so the SPSS parts weren't so relevant to me. Perhaps some written examples of how to run/interpret STATA outputs would have been helpful*".

One of these delegates had brought example STATA output to the course to interpret, and subsequently struggled to apply the interpretation of SPSS outputs to the STATA output:

"*Even with the STATA output in front of me, I wasn't sure exactly how this related to the SPSS outputs*".

The 'Introduction to Logistic regression' course has run three times so far in 2016 with 38 delegates attending in total; of those 38, 22 (58%) responded to feedback requests. Of the 22 who provided feedback, 12 (55%) rated the course overall as 'very good' (5), 7 (32%) rated the course 'good' (4) and the remaining 3 (13%) rated the course 'average' (3).

<u>Regressions with R</u>

This course, first run in 2014, has elicited electronic feedback for all of the 11 occasions on which it has been presented. While 'Regressions with R' incorporates the software exposure that delegates felt was missing from the other two regression courses, some delegates reported that they found it difficult to concentrate on both the theory being taught and the R code being explained at the same time:

"*I found myself trying too hard to keep up with typing the code and missing the real point*".

The feedback for this course suggested that combining software exercises and theory within the same session may not be the most effective structure for a course of this type; the implication of feedback being that some delegates struggle to understand the theory whilst also trying to understand the software application, and vice versa. Some delegates suggested spending one day of the course solely teaching the theory, and the second day applying this theory using computer programs.

"Thought [the teacher] went quite quickly through the R code, (which was fine if very familiar with R) but I am not so familiar. I feel a day could have been dedicated to the theory and a day dedicated to the exercises"

This course has a pre-requisite stated on the website that delegates must have a working knowledge and ability in R in order to attend this course; however, the feedback received suggests that this criteria is not always met.

"Many participants seemed to have pretty advanced knowledge of R and could keep up, but for others (like me!) it was difficult to keep up."

Introducing a test of some sort, or clearer guidelines regarding who this course is appropriate for, may improve the feedback for this course and reduce the number of delegates who feel overwhelmed by the combination of theory and software application.

This course ran twice in 2016 with 33 attendees in total; of which, 21 responded to a request for feedback. Of the 21 respondents, 9 (43%) rated the course overall 'very good' (5), 10 (48%) rated the course 'good' (4) and the remaining 2 (9%) rated the course 'average' (3). Although this feedback was still overwhelmingly positive, the 'Regressions with R' course received fewer 'very good' (5) responses than the other two courses despite containing the software exposure that delegates felt were missing from the other courses; 75% of delegates that attended 'Introduction to Regression Analysis' in 2016 rated the course 'very good' (5) overall whereas just 43% of delegates attending 'Regressions with R' 'very good' meaning almost a third less (32.1%) rated the 'Regressions with R' course 5.

DISCUSSION

Active learning is a method of teaching which encourages students to actively take part in the lesson and gain hands-on experience of the subject matter being taught; active learning is "*anything that involves students in doing things and thinking about the things they are doing*" (Bonwell & Eison, 1991, p. 2). Since they require motivation, and are said to learn better when they feel the subject matter is important to them, active learning may be an effective tool for adult learners, as it can be used to demonstrate how the theory is applied in practice. There are many ways in which active learning can be incorporated into courses, such as use of practical exercises and discussion sessions within teaching. Problem based learning (PBL) is one such method of active learning, which provides delegates with hands-on experience of analysis. PBL requires the teacher to introduce a problem to students as well as guidelines about how the problem should be approached; the teacher acts as a facilitator whilst students take the lead and aim to solve the problem either as a group or alone (Hmelo-Silver, 2004). PBL encourages delegates to critically think and engage with the subject area.

All three regression analysis courses offered by CASC contain elements of active learning, but adopt different methods with regards to hands-on use of statistical software. We have here described these courses, the differences between them with regards to software usage, and contrasted student feedback. Through this initial look at our data, we have identified some potentially useful insights to inform our future direction.

Although all three courses received favourable feedback, the response to software components (or lack thereof) was rather mixed. It appeared that some delegates prefer to learn the theory behind the methods before moving on to discussing how those methods can be applied, whilst others commented that separating theory and application made it more difficult to retain the information taught in theory sessions without practical applications. Since the 'Introduction to

Regression Analysis' and 'Introduction to Logistic Regression' courses were first designed, computers have become substantially more accessible and powerful; many delegates now have access to one or more statistical packages and will have used them in their day-to-day work. Based on the number of comments noting the lack of software element in both courses, it is clear that delegates feel software exposure is something that they find necessary to any contemporary statistical courses. Additionally, the increased exposure to and use of computer based applications, does mean that usage in the classroom is less novel, more expected, and less likely to involve unnecessary disruption of the flow by unrelated computer concerns.

Another problem faced when combining a course with software use, is the choice of package. Dealing with delegates from diverse backgrounds, some may be familiar with more complex packages such as R, while others may have used a simpler package such as SPSS; similarly, some may have extensive experience, while other may only ever have used a statistical software package briefly or not at all. The risk of incorporating a day spent on either R or SPSS to 'Introduction to Regression Analysis' or 'Logistic Regression' is that some delegates may feel the course is no longer suitable to their needs.

The overall view from the feedback received is that, although there is no perfect method to teach regression analysis, a software component is desirable. This is a common theme across all three courses as reflected by their course evaluation comments. However, there is a danger, if these two components run concurrently (as they currently do in 'Regressions with R'), that delegates will focus too much on understanding the intricacies of the software program rather the methods being demonstrated.

Based on the evidence from course feedback, I feel that the effectiveness of methods for teaching regression analysis to non-statistician could be improved in one of two ways. Firstly, the current method used to teach Introduction to Regression Analysis could be extended to allow delegates access to computers in order for them to carry out analyses on the second day, with the help of the teacher. Alternatively, the method currently used to teach Regressions with R could be adopted, but with stricter guidelines regarding the level of software knowledge required for the course, and a greater allowance of time is given for explanations of theory. Both of these methods would provide exposure to software for delegates, would demonstrate the usefulness of methods taught in a real-world setting, and would include a PBL component to challenge and motivate adult learners.

REFERENCES

Bonwell, C. C., & Eison, J. A. (1991). Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports. ERIC Clearinghouse on Higher Education, The George Washington University, One Dupont Circle, Suite 630, Washington, DC 20036-1183.

Rubenson, K. (2011). Adult learning and education. (Ed.). Academic Press.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? Educational psychology review, 16(3), 235-266.

A PILOT EXPERIMENTAL STUDY COMPARING DIFFERENT ACTIVE LEARNING STRATEGIES FOR UNDERSTANDING SAMPLING

TIRLEA, Laura[1], BAGLIN, James [2], HUYNH, Minh[1], and ELPHINSTONE, Bradley [1]
[1] Swinburne University of Technology
[2] RMIT University of Technology
lauratirlea@swin.edu.au

This paper reports the preliminary results of an experiment evaluating the effect of two different active learning strategies for learning about the process of sampling. The experiment compared an interactive classroom exercise for taking a sample from a population, to an online simulation method, which had students engage in sampling using an online virtual world. A total of fifteen participants were randomly allocated to one of the two learning strategies where they completed a one-hour sampling lesson. Prior to randomisation, all participants completed a short quiz assessing their current understanding of sampling, in conjunction with rating their attitudes towards statistics. One week after completing their respective sampling task, the students completed a follow up quiz and questionnaire. Overall, the results of this pilot study indicated that both groups had improved their knowledge of statistical sampling after one week follow up, however, there was no evidence of a difference between the groups. This paper summarises the results of this pilot study and will utilise the findings to design the next phase of this research.

INTRODUCTION

The ability to understand and apply statistical concepts is becoming increasingly essential for people living in data-driven societies (Garfield & Ben-Zvi, 2007). However, statistical concepts are abstract and difficult to understand, thus statistical *misconceptions* can occur frequently among students, researchers and even teachers (Iten, 2015). For example, concepts pertaining to samples and sampling procedures are often misconstrued within the classroom (Ben-Zvi et al., 2015). Specifically, there is evidence to suggest that students commonly believe that a sample is only "good" if the sample size represents a large proportion of the target population (Smith, 2004). In addition, Huck (2008) suggested that many students believe Simple Random Sampling (SRS) procedures will *always* result in a representative sample, irrespective of the nature of the population. Such misconceptions result in the unfortunate outcome of impeding students' learning and limiting their enthusiasm and engagement towards statistics (Huck, 2008). As such, it is imperative that statistical misconceptions within the classrooms are addressed, and new strategies to overcome them are investigated.

Over the past couple of decades, educational researchers have been advocating for the use of active-learning within the classrooms. Whilst there is no universally accepted definition, active-learning is generally described as any instructional method that engages students in the learning process (Prince, 2004). A review of the literature has extensive empirical support for active learning across many disciplines (Michael, 2006). In statistics education, misconceptions regarding sampling are much more easily undone through active-learning processes, rather than traditional passive lectures. For instance, actively engaging students in collecting their own data permits them to construct their own understanding of what it means to take a representative sample. In present times, the integration of information technology (IT) has substantially changed the content and pedagogy of learning and teaching statistics (Moore, 1997), and has made the data collection process much more practical.

IT mitigates the time and burden of handling cumbersome and tedious calculations, enabling students to have more time to adequately explore, analyse and interpret their data (Chan & Ismail, 2012). According to Rubin (2007), complex concepts within statistics are made more accessible owing to the role that IT has when learning. For example, numerous studies have demonstrated that the usage of spreadsheets and simulation programs have diminished misconceptions regarding correlation and distributions (Alias, 2009; Liu, 2010). However, very little research has focused on correcting misconceptions towards statistical sampling (Ben-Zvi et al., 2015), which is unfortunate because understanding how samples vary is crucial in order to make reasoned data-based decisions.

Therefore, the current study experimented with using the *Islands* software (Bulmer and Haladyn, 2011) to teach students about statistical sampling. The *Islands* (http://islands.smp.uq.edu.au)

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

is a free, online, human population simulator that provides students with a virtual environment to explore and experiment. The program can be navigated by clicking between the 27 villages, across three unique Islands, that are home to approximately 34,000 virtual residents (see Figure 1). Simulation models govern the population’s births, deaths, health, social lives and residency. Each resident has their own personal story and genetic code that are linked to their appearance and responses to over 200 tasks modelled on real studies. This expansive environment provides students with incredible opportunities to learn about sampling as they explore and collect their own “representative” samples.

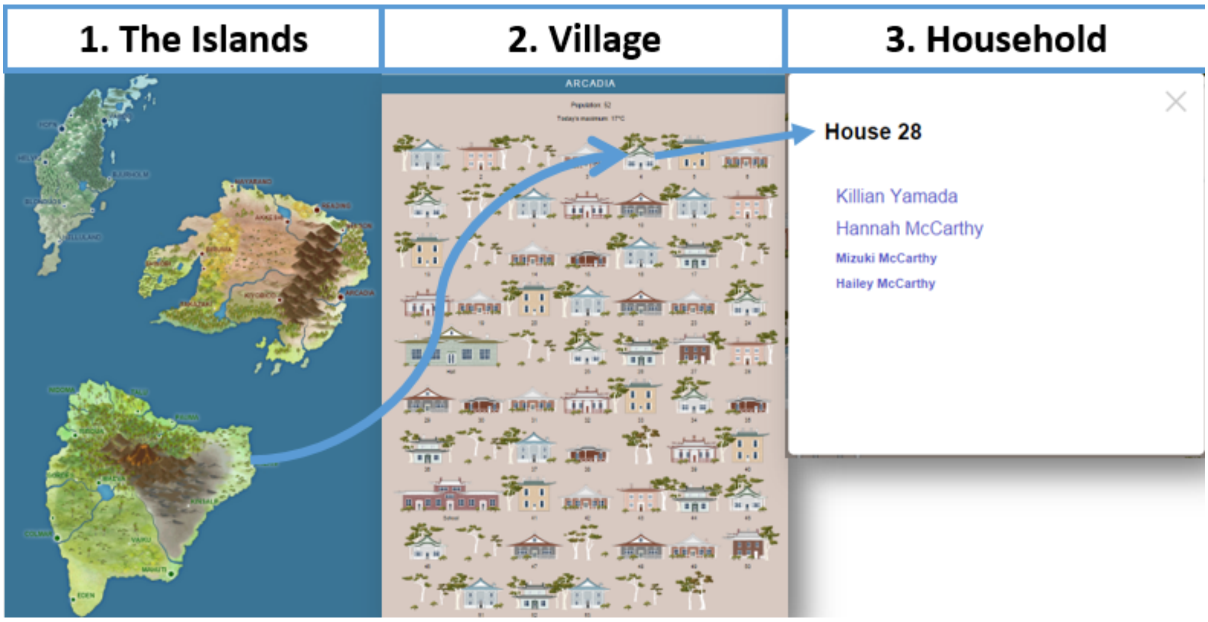


Figure 1: Screenshot demonstrating how to navigate the *Islands*

As such, the aim of this study was to investigate if the *Islands* program provides a suitable environment for learning about statistical sampling. Specifically, the researchers were interested in determining if misconceptions pertaining to sampling persisted after using the *Islands*. In addition, the researchers were also interested in comparing the effectiveness of an *Islands*-based activity to a non-IT (traditional classroom) activity for learning about sampling procedures.

METHOD

Participants: The sample consisted of 18 participants currently enrolled in an introductory statistics unit. Three students had dropped out from the study leaving a final sample of 15 (two students chose not to attend the session and one had dropped out of the degree). There were 11 female and 4 male student participants. All students were full time 1st year students and one was an international student.

Measures: Sampling misconception was measured by using short online quizzes (pre / post) pertaining to themes identified from Huck (2008): *Statistical Misconceptions*. These items were modified to relate specifically to sampling misconceptions from a research perspective. The quiz contained four common misconception questions with which students would select a yes or no response. These items were:

1. Using Simple Random Sampling will always select the sample evenly from the population
2. If a truly random process is used, then the sample will be just like the population, but smaller
3. The size of the sample should be proportionate to the size of the population
4. The larger the sample, the greater the accuracy

The quizzes also contained 10 multiple choice items which tested the students' abilities to select the correct type of sampling procedure based upon provided scenarios. The scores on these items were totalled to derive a total quiz score, used during the analysis phase.

Procedure: Following ethics approval and permission from relevant unit conveners, prospective students were made aware of the study by online announcements made on their units' respective learning management system (LMS). Involvement in the study was completely voluntary, with participants completing the study tasks outside of their regular scheduled class times. Participants were randomly allocated to one of the active learning sessions (*Islands*-based versus *Control*) prior to attending. They were also sent the pre-quiz via email and informed to complete all questions before their session. As a precaution, hard copies of the quizzes were made ready on the actual testing day in the event that the participants did not complete the online version.

Participants randomised into the *Islands* intervention group (n = 8) were required to use the *Islands* program to collect a "random" sample of thirty *Islanders*. Brief web-based instructions on how to do so were provided to the participants at the start of the session. On the other hand, participants randomised into the Control group (n = 7) were shown a large A1 sized map of a drawn fictitious town. The town contained 65 houses of varying sizes, and each of these houses contained a different number of jellybeans (to represent people), placed on top. The number of people / jellybeans for each house varied from 0 to 4, with a total population of 213. To keep things simple, the 65 houses all belonged to the same street and were numbered in sequential order. Using an online random number generator, the researcher demonstrated to these participants a number of ways in which jellybeans could be drawn from this town. Afterwards, students would replicate the process and collect their own samples / jellybeans. Whilst completing their tasks, participants (from both groups) were free to ask the researcher / demonstrator questions if they encountered any difficulties.

One week after the sessions were completed, the participants were sent the post-quiz via email. The post-quiz questions were identical to the pre-quiz questions, so to avoid practice effects, students were not informed if their answers on the pre-quiz were correct or incorrect. Subsequently, any changes in their responses would be due to the classroom interventions.

RESULTS AND DISCUSSION

The results of this study were restricted to a descriptive analysis due to the small sample size. Descriptive statistics comparing the two groups are shown in Table 1. The median was the preferred measure of central tendency due to the susceptibility of the mean to outliers in small samples. Bar charts representing the percentage of incorrect responses between before and after testing periods for the four common sampling misconception questions were also included (see Figure 2). The maximum score for the quiz was 10 and the maximum score for the misconception questions was 4.

Table 1: Descriptive statistics between groups

	Islands Group			Control Group		
	Median	*Mean*	*SD*	*Median*	*Mean*	*SD*
Quiz Score						
Pre	3	2.75	1.28	3	2.57	1.21
Post	7	6.75	1.39	7	7.14	1.68

For the multiple choice quiz component, the percentage of correct responses for the *Islands* group on the pre-test (29%) was substantially lower than that of the post-test (68%). Similarly, there was a substantial change for the Control group between pre (29%) and post (71%) quiz scores. The median score for both groups at the pre-test was equal to 3, and the median post-test score for both groups was equal to 7.

For the first sampling misconception question (Using SRS will always select the sample evenly from the population), both the *Islands* and Control groups performed relatively poorly on the pre-test (25% and 29% respectively). However, both groups performed much better for this question on the post-test (63% and 57% respectively). Whilst there is no clear indication of a difference between the two groups for this item, Figure 2 clearly indicates substantial improvement from pre to post scores. These improvements on the post-tests are not surprising given that participants from both groups used a variety of sampling techniques (i.e. cluster sampling) to complete their activities. Of great pleasure to the research team, a number of students enquired further about other types of sampling techniques, and in what situations would they be useful.

For the second misconception question (if a truly random process is used, then the sample will be just like the population, but smaller), similar outcomes to the first item resulted. The pre-test scores for both the *Islands* and Control groups (38% and 29% respectively) were substantially smaller to their post-test counterparts (75% and 71% respectively). For the *Islands* group, the participants were requested to used cluster sampling to draw a random sample of size thirty from the same town. Afterwards, the students were told to collect height data from their respective samples. When comparing their results (the height data) the students realized there was a lot of variability between the samples. The students noted that whilst some of the samples were similar to the population data (shown by the researcher afterwards), others were quite different. This explained the improvement on the item for the post-test. For the Control group, the students were told to identify the jellybeans by colour (five different colours in total). For this group, only one of the student's sample was similar in colour proportion to the overall population. As such, these students came to similar conclusions as the *Islands* group.

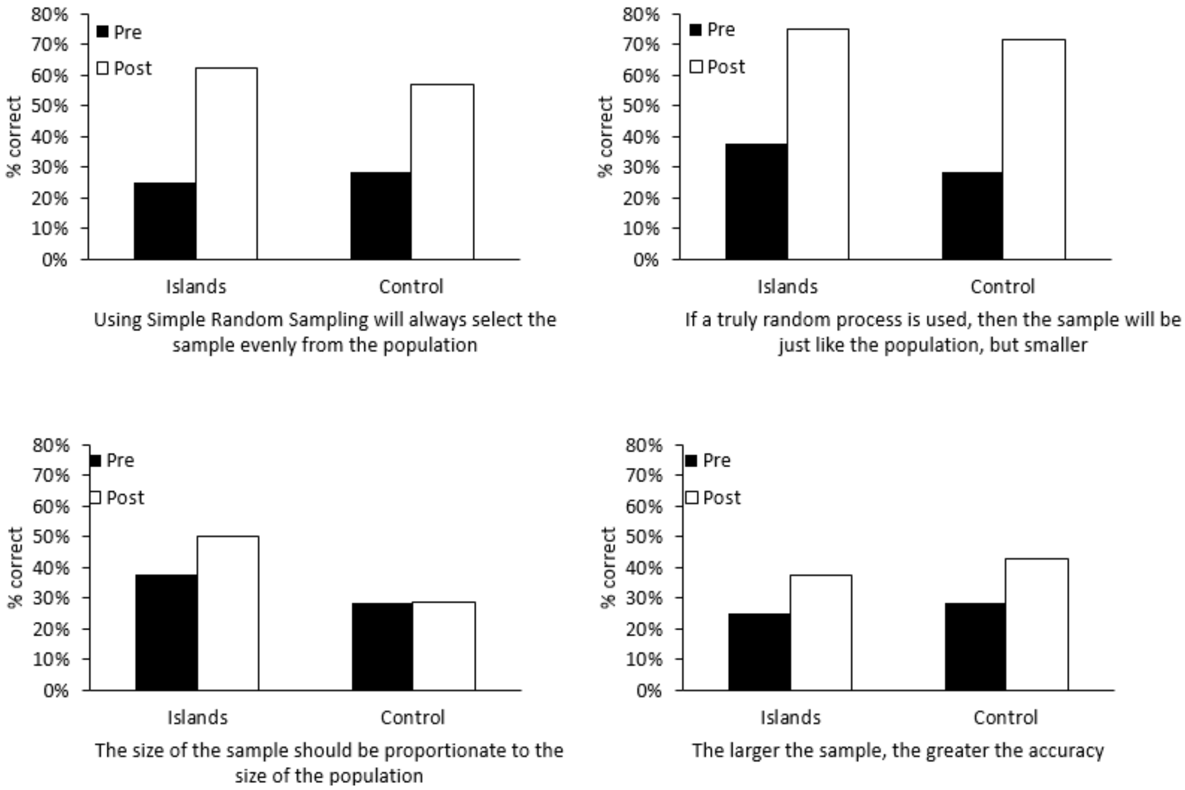


Figure 2: Bar charts comparing the % of correct responses for the Islands and Control groups

For the third (the size of the sample should be proportionate to the size of the population) and fourth (the larger the sample, the greater the accuracy) common misconception items, the change in pre and post scores was not as apparent as the first two items, for either group. To properly address these items, students would have required to have some knowledge of the standard error (*SE*) formula, which shows that when N is much larger than n, the ratio of n to N does not influence the *SE* to great extent. For this particular unit, the students had not yet covered the topics of sampling error during their regular scheduled classes, which would explain the pre-test results. Similarly, the tasks required from both groups during the testing period did not cover the concepts of the *SE* either, leading most of the participants to make the incorrect assumption regarding these items.

LIMITATIONS AND CONCLUSION

This experiment had a number of limitations to be noted. Firstly, despite the researcher's best efforts to maximise participation (e.g. LMS posts, emails, lecture visits), students overall did not appear interested, and the resulting sample was small. The study itself attempted to avoid monetary incentives to encourage participation, however perhaps another form of incentive (e.g. raffle prize) should be considered for future studies. Alternatively, including these activities into the regular scheduled classes could also be a possibility for follow up studies. Future studies should also consider the potential for a third, non-active-learning group (such as a traditional lecture / tutorial). The present study initially hypothesised that IT based learning strategies (such as the *Islands*) will be superior to other activities, such as the task completed by the Control group, and subsequently did not include a traditional learning group. However, given these findings, it would appear that active-learning tasks, irrespective of technological platform (or lack thereof), is successful, at least in the short term, of correcting common misconceptions towards sampling.

Overall, this pilot study suggests, at the very least, that the two active-learning strategies were comparable for overcoming certain sampling misconceptions. Furthermore, participants from both groups demonstrated clear improvements in their ability to differentiate between different types of sampling procedures. We will utilize the results of this investigation to carry out future studies examining more students and the possibility of running the *Islands* activity during a regular class.

REFERENCES

Alias, M. (2009). Integrating technology into classroom instructions for reduced misconceptions in statistics. *International Electronic Journal of Mathematics Education*, *4*(2), 77-91.

Ben-Zvi, D., Bakker, A., & Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics*, *88*(3), 291-303.

Bulmer, M., & Haladyn, J. K. (2011). Life on an Island: A simulated population to support student projects in statistics. *Technology Innovations in Statistics Education*, *5*. Retrieved from http://escholarship.org/uc/item/2q0740hv

Chan, S.W., & Ismail, Z. (2012). The Role of Information Technology in Developing Students' Statistical Reasoning. *Procedia-Social and Behavioral Sciences*, 46, 3660-3664.

Garfield, J. & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372-396.

Huck, S.W. (2008). *Statistical misconceptions*. New York: Psychology Press, Taylor & Francis.

Iten, G. (2015). *Impact of visual simulations in statistics: The Role of interactive visualizations in improving statistical knowledge*. Springer: Switzerland

Liu, T.-C. (2010). Developing simulation-based computer assisted learning to correct students' statistical misconceptions based on cognitive conflict theory, using "correlation" as an example. *Educational Technology & Society*, *13* (*2*), 180-192.

Michael, J. (2006). Where's the evidence that active learning works? *Advances in Physiology Education*, *30*(*4*), 159-167

Moore, D.S. (1997). *Statistics: Concepts and Controversies (4th Ed.).* New York: Freeman.

Prince, M. (2004). Does active learning work? A review of the research. *J Engr. Education*, *93*(*3*), 223-231.

Rubin, A. (2007). Much has changed; little has changed; revisiting the role of technology in statistics education. *Technology Innovations in Statistics Education*, *1*(1). Retrieved on 4 October, 2011, from http://escholarship.org/uc/item/833239sw

Smith, M. H. (2004). A sample/population size activity: Is it the sample size or the sample as a fraction of the population that matters? *Journal of Statistics Education*, *12*(*2*).

TEN SIMPLE RULES FOR LEARNING THE LANGUAGE OF STATISTICS

RICHARDSON, Alice M.[1], DUNN, Peter K.[2]
CAREY, Michael D.[2], MCDONALD, Christine[3]
[1]Australian National University, Acton ACT
[2]University of the Sunshine Coast, Sippy Downs, QLD
[3]University of Southern Queensland, Toowoomba QLD
PDunn2@usc.edu.au

In this paper we propose ten simple 'rules' for guiding students' learning of the language of statistics. Learning any new subject brings with it the requirement to learn the language associated with that subject. Students also bring with them varying understandings about the relationship between statistics and mathematics. Many students expect the formality and precision of mathematics to transfer to statistics, and are baffled to discover this is not the case.

The first four rules will guide instructors and learners around the landscape of tricky terms, from general English to the English of mathematics, statistics and other disciplines. The remaining six rules will establish some signposts along the way to assisting students to overcome the challenges of the language of statistics. We acknowledge that there is no single route to enforce here, and that management of expectations, embracing ambiguity in terminology, and reinforcement of new language through writing and speaking all have a role to play in teaching and learning the language of statistics.

INTRODUCTION

The need for technical terms to communicate complex concepts is evident even in subject areas not traditionally associated with language, such as mathematics and statistics. Indeed, "communication is at the heart of statistics" (Rangecroft, 2002; p. 34). Further supporting this, the American Statistical Association's Guidelines for Assessment and Instruction in Statistics Education (GAISE) committee made six recommendations (Aliaga et al., 2010) for introductory statistics courses, one of which is to emphasise statistical literacy, which they define as "understanding the basic language of statistics… and fundamental ideas of statistics" (Aliaga et al., 2010; p. 14). Teaching statistics, however, often focuses on quantitative aspects, which is often reflected in the number and the types of textbook exercises. While these quantitative aspects are important, students may not completely understand the concepts in these problems if they do not understand the language surrounding them at a deeper conceptual level. Furthermore, some students may be able to obtain correct answers for these quantitative questions, but not understand what these results mean, or be able to communicate these results.

In this paper we follow the "ten simple rules" tradition continued by Kass et al. (2016) and propose ten rules for learning the language of statistics. These rules fall into two groups. The first four rules guide instructors around the landscape of tricky terms in statistics. The final six establish signposts along the way to assisting students to overcome the challenges of the language of statistics.

THE LANDSCAPE OF TRICKY TERMS

One reason why the language of statistics is hard to learn is that it uses terms from different sources (Dunn et al. 2016). Terms may come from general English (GE); these are terms used in everyday language, but some have specific meanings in statistics. As with any discipline, statistics also uses terms unique to the discipline, which we call Statistical English (SE). In addition, statistics is usually studied by students in different disciplines, but sometimes terms from those other disciplines (discipline English, DE) have different meanings in statistics. Finally, terms may come from mathematical English (ME), but confusingly some of these may have slightly different meanings in statistics. Figure 1 shows the relationships discussed in graphical form.

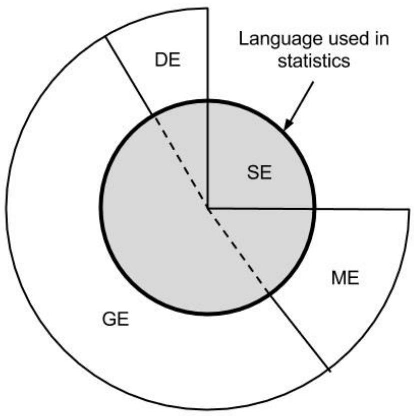


Figure 1: The language used in statistics is a combination of general English (GE), statistical English (SE), discipline English (DE), and mathematical English (ME).

Rule 1: Teach students terms they "know" but have new meanings.

Learning statistics involves learning how some terms from GE are given new meaning; these terms are also known as "lexically ambiguous". This includes terms such as "significant", "random" and "power". Kaplan et al. (2011) focused on "spread" as a particularly slippery statistical term and suggest it should be avoided. More recently, Kaplan et al. (2014) have addressed "random", though this time they do not go as far as proposing it be deleted from statistical discourse. A cause for confusion is that students need to learn which GE terms have specialised meanings. For example, "significant" has a special meaning in SE, but "substantial" does not. Lavy and Mashiach-Eizenberg (2009) encourage instructors to explicitly discuss both the GE and SE meaning of such terms, so "that students be presented with both their formal and informal definitions" of terms (p. 7) and be able to compare and contrast these definitions.

Phrases built from GE words are a second level of complication to learning the language of statistics. Some of these phrases comprise words that do not help students to understand the concept; for example (Dunn et al. 2016), "standard error" is not "standard" in any GE sense, and is not an "error" (in the sense of a "mistake").

Rule 2: Introduce students to the meaning of terms that are completely new.

Statistical English forms a smaller but recognisable part of learning statistics. "ANOVA" and "boxplot" are two examples of terms that would not be known outside of statistical English. For many students, Greek symbols will be new also (such as μ and β), and for some students even mathematical symbols such as $\pm$ will be new. In some sense these terms and symbols are the easiest to teach since there are no prior meanings for students to unlearn.

Perhaps the hardest group of such terms with which to engage are the statistical English terms with no universal definition. Examples include "lurking variables" and "confounding variables" (for example, see Dunn et al. (2016) and Flanagan-Hyde (2005)). Other examples include "degrees of freedom" and even "discrete" and "continuous" variables.

If there is no commonly accepted definition, it is likely to be fruitless to expect students to be able to learn it as if there is. However, a balance between being prescriptive on the one hand and being simplistic on the other hand needs to be reached. For example, give the term a specific context relevant to the level of the students and the topic under discussion so that at least one version of the definition can take shape, while warning students that they will come across other versions in other contexts (for example, "degrees of freedom" can be taught multiple times in a number of different contexts). It is clear, then, that a difference exists between knowing the definition of a term, and understanding the concept that the term represents.

Rule 3: Clarify terms that may have different meanings in discipline and statistical English

Statistics is usually taught to students who come from other disciplines. In some cases, terms used in statistics and their home discipline may be similar but with different meanings. Because these students are likely to study more courses in, and be more passionate about, their home discipline, statistics instructors must ensure students and discipline instructors are aware of these potential ambiguities. One complication is that the statistics instructor may not be aware of these potential overlaps.

One example is the word "sample". In statistics, a sample is a set of observations drawn from a population. In biomedicine a sample is a single specimen (of blood, urine, etc.) rather than a set of observations. Statistics students have been heard to incorrectly state "I have taken 40 samples" rather than "I have taken one sample of 40 observations". Other examples include the terms "blocking" and "regression" as well as symbols such as ρ (for density) and β (for beta-radiation) (Dunn et al. 2016).

Rule 4: Clarify terms that may have different meanings in mathematics and statistics

The final component of the language landscape in these 'rules' comprises terms taken from mathematical English (ME) and given a subtle twist in SE. One example is the word "estimate" (Dunn et al. 2016) which, in statistics, means a numerical approximation of a population quantity by a sample quantity. However, in ME it means to make a sensible guess ("estimate the results of this calculation"). Even the innocuous "linear" fits into this category. In regression, for example, the model $y = a + bx + cx^2$ can be fitted as a "linear model" (that is, the equation is "linear in the parameters"). However, in a mathematics class, the model is called non-linear (specifically, a "quadratic").

The symbol ± is used differently in mathematics and statistics also. In mathematics, it designated two possible solutions to (for example) a quadratic equation. When used in the context of a confidence interval in statistics, is designates a *range* of possible values.

When students are alerted to these differences, the shape of the landscape of statistical language becomes clearer. The same approach is advocated in the teaching of foreign languages where the concepts of a new term, a similar term and an identical term help shape the foreign language learning in terms of what learners already know from their first language.

SIGNPOSTS AS GUIDES FOR OVERCOMING CHALLENGES

Having established the landscape of statistical terms, we now discuss some basic ideas for helping students to learn their way around it. There are many ideas for how to teach students statistical terms, and many lack supporting evidence. Consequently, these remaining six rules are broad pathways rather than narrow maxims, and so we have designated them as signposts along the journey of learning the language of statistics.

Rule 5: Manage synonyms sensibly

For some statistical concepts, many synonyms exist. For example, in regression, the phrases "explanatory variable", "independent variable", "regressor", "predictor" (among others) are synonymous. Synonyms may occur in multiple places in one course, such as across resources (textbook, readings, journal articles that are read, etc.) and between personnel (tutors from different discipline areas).

One solution will not suit every class here. Give every synonym, and students are likely to become overwhelmed. However, avoiding mentioning synonyms may lead to even good students becoming confused if they are reading a different text or a journal article which uses a different term. A middle path that should cater for most situations would be to make students *aware* that the statistical language is inconsistent and offer some common alternatives. This can be handled across the range of materials in a course as the weeks go by.

Rule 6: Embrace ambiguity

One of the most fundamental concepts in statistics, the "mean", is actually a classic example of an ambiguity. While students may be taught about *the* mean, the mean can be modified in numerous ways: "arithmetic mean", "geometric mean", and "sample mean" for example. If a student sees, for example, "arithmetic mean" mentioned in a research article, they may not know if this is the "mean" they were taught or not. Another example is "correlation", which may refer to either a Pearson or Spearman correlation.

These are two interesting cases with different issues. With "mean", the modifiers attached are technical terms that attempt to clarify the statistic. On the other hand, the descriptors attached to "correlation" are names honouring the discoverer of the relevant statistic and offer no clue as to the construction or use of the statistic.

An interesting example is the ambiguity attached to "significance": this can refer to statistical, practical or clinical significance (Thompson 2002). Though all three are related, often results are presented using just the term "significant" or "significance".

Instead of taking a deficit view of the ambiguity of statistical terms, one approach to overcoming it is to embrace it, highlight it, and discuss it, because this creates the opportunities to grasp teachable moments and explicitly clarify the ambiguity. Ambiguities can be used as a scaffold to discuss how some terms have different meanings in GE and SE, or how some concepts have many terms associated with them. Many introductory statistics classes act as service courses with students from programs as diverse as Sports Studies, Commerce and Engineering sitting in the same lectures. They will be bringing their understanding of the DE in their discipline to the statistics classroom and can help students from other disciplines to construct their own knowledge of ambiguous terms.

Rule 7: Avoid ambiguity

On the other hand, there are ways to avoid *unnecessary* ambiguity and be quite prescriptive, modifying potentially ambiguous terms thereby giving students firm ground to stand on when it matters. For example, instructors can use the phrase *statistical* significance whenever referring to the results of hypothesis tests, and otherwise shun the word "significant" in favour of, say, "important" or "substantial".

Another pair of terms about which it would be wise if instructors made up their minds is "correlation" and "association". "Correlation" has a fairly well-agreed specific SE definition, and has been discussed already in this paper. "Association" is a looser GE term that students are prone to use interchangeably with "correlation", a habit which could easily lead to confusion (Kaplan, et al. 2010).

Rule 8: Introduce language recursively through authentic contexts

The reason for incorrect vocabulary learning can be attributed to a number of factors, but one of the most common causes is learning terms and expressions without a context provided (Nassaji, 2003). To maximise and consolidate the learning of new vocabulary, provide opportunities for students to "recycle" vocabulary and to make lexical inferences to generate meanings for unknown terms they encounter in context (Deschambault, 2012).

Don't assume students know a word after defining it once: the word and its meaning need to be reinforced through many activities. Richardson et al. (2013a) found that tutors already had a sense of which terms students would stumble over, because the tutors are immersed in the language of statistics, have met tricky terms many times and have developed a deep sense of the meaning of a particular term. They can then use activities that expose students to tricky terms using two approaches: context-free exercises ("define significant"; Kaplan et al., 2009) or contextual ("What does 'significant' mean in this journal extract?"; Richardson et al., 2013b). A semester's worth of activities can then expose students to appropriate use of a term. Richardson et al. (2013a) found, this type of reinforcement led to substantial gains in understanding after one semester.

Rule 9: Use verbal interaction

Verbal interaction is also important. Instructors can model correct pronunciation of terms which helps develop the sense of community in a class, with a shared way of speaking. Sound bites demonstrating correct pronunciation have migrated from the cassette decks to the websites of language courses. Why not in statistics websites too? For example, the Wikipedia pages for most common statistical terms do not include any audio or video where students would hear the terms being said aloud.

For some terms, students may say the term incorrectly (such as "confident interval" and "outliner"), and it may simply be that students have misheard what an instructor or tutor is saying.

Other terms, such as "Kolmogorov" and "heteroscedasticity", can be tricky for native English speakers.

Activities that foster conversation amongst students are useful too (see Garfield 1993). Explanations of collaborative learning appear in Willis & Willis (2008), and include techniques such as the jigsaw (Tran & Lewis, 2012), Think-Pair-Share (Rudolph, Lamine, Joyce, Vignolles & Consiglio, 2014), Read-Ask-Tell (Remsburg, Harris & Batzli, 2014) and game-based activities (Lesser et al. 2013).

Rule 10: Find your place on the implicit/explicit teaching continuum

Although this paper has been all about 'rules', we acknowledge that there is no single solution to the challenges of teaching and learning the language of statistics. In particular, there is a continuum with use of explicit methods on one end, and implicit methods on the other.

Instructors may help students acquire vocabulary using implicit learning or explicit teaching approaches (Carlisle, Fleming, & Gudbrandsen, 2000). *Implicit learning* places an emphasis on comprehending meaning rather than learning new terms, and so often uses authentic tasks such as discussing a research article. *Explicit teaching* focuses on teaching terminology explicitly, by (for example) matching terms to definitions.

Rumsey (2002) supports the idea of implicit teaching, suggesting that instructors should not concern themselves too greatly about specific terms and their specific meaning in some situations: "… terms, like 'precision,' 'accuracy,' 'reliability,' 'bias,' and 'consistency' all sound the same to students. In my opinion, splitting hairs about these terms will only create confusion and frustration. My advice is to choose the most important ideas, and stick to them." (Rumsey, 2002, Section 3.5).

On the other hand Kaplan et al. (2011) used an explicit teaching approach, arguing in favour of using specific statistical terms where possible (and sometimes it is not possible); for example, using "dispersion" rather than the more general and potentially ambiguous "spread".

CONCLUSION

In this paper we have introduced ten simple 'rules' for the teaching and learning of the language of statistics, consisting of four aspects of the landscape of tricky terms, and six signposts for navigating the landscape successfully. Just like its parent, English, the language of statistics is rich and draws from a variety of sources, including general, mathematical, statistical and discipline-specific English. We believe these rules offer statistics educators a way to welcome learners into the community of statistics speakers and enhance statistical literacy across the board.

ACKNOWLEDGMENTS

This research was partly supported by a University of the Sunshine Coast Open Learning and Teaching Grant (OLTGP2012/04).

REFERENCES

Aliaga, M., Cobb, G.,Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, P., & Witmer, J. (2010). *Guidelines for assessment and instruction in statistics education: College report*. Technical report, American Statistical Association.

Carlisle, J.F., Fleming, J.E., & Gudbrandsen, B. (2000). Incidental word learning in science classes. *Contemporary Educational Psychology, 25*: 184–211.

Tran, V.D. & Lewis, R. (2012). The effects of jigsaw learning on students' attitudes in a Vietnamese higher education classroom. *International Journal of Higher Education, 1*(2), 9-20.

Deschambault, R. (2012). Thinking-aloud as talking-in-interaction: reinterpreting how L2 lexical inferencing gets done. *Language Learning, 62*(1), 266-301.

Dunn, P.K., Carey, M.D., Richardson, A.M., & McDonald, C. (2016). Learning the language of Statistics: challenges and teaching approaches. *Statistics Education Research Journal 15(1)*.

Flanagan-Hyde, P. (2005). Confound it! I can't keep these variables straight! *STATS, 43,* 21–23.

Garfield, J. (1993). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education, 1*(1).

Kaplan, J. J., Fisher, D. G., & Rogness, N. T. (2009). Lexical ambiguity in statistics: what do students know about the words association, average, confidence, random and spread? *Journal of Statistics Education*, *17*(3).

Kaplan, J. J., Fisher, D. G., & Rogness, N. T. (2010). Lexical ambiguity in statistics: how students use and define the words: association, average, confidence, random and spread. *Journal of Statistics Education Education*, *18*(2).

Kaplan, J. J., Rogness, N. T., & Fisher, D. G. (2011). Lexical ambiguity in statistics: making a case against spread. *Teaching Statistics*, *34*(2), 56–60.

Kaplan, J. J., Rogness, N. T., & Fisher, D. G. (2014). Exploiting lexical ambiguity to help students understand the meaning of random. *Statistical Education Research Journal*, *13*(1), 9–24.

Kass, R.E., Caffo, B.S., Davidian, M., Meng, X.-L., Yu, B. & Reid, N. (2016). Ten simple rules for effective statistical practice. PLOS One Computational Biology 12, e1004961.

Lavy, I., & Mashiach-Eizenberg, M. (2009). The interplay between spoken language and informal definitions of statistical concepts. *Journal of Statistics Education*, *17*(1), 1–9.

Lesser, L. M., et al. (2013). Using fun in the statistics classroom: An exploratory study of college instructors' hesitations and motivations. *Journal of Statistics Education*, *21*(1), 1–33.

Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *Tesol Quarterly*, 645-670.

Rangecroft, M. (2002). The language of statistics. *Teaching Statistics*, *24*(2), 34–37.

Remsburg, A. J., Harris, M. A., & Batzli, J. M. (2014). Statistics across the curriculum using an iterative, interactive approach in an inquiry-based lab sequence. *Journal of College Science Teaching*, *44*(2), 72.

Richardson, A. M., Dunn, P. K., & Hutchins, R. (2013a). Identification and definition of lexically ambiguous words in statistics by tutors and students. *International Journal of Mathematical Education in Science and Technology*, *44*(7), 1007–1019.

Richardson, A. M., Dunn, P. K., & Hutchins, R. (2013b). The impact of tutor, extract and word on the correct definition of lexically ambiguous words in Statistics. In: *ACSME Proceedings: Students in transition - the learners' journey.* Canberra, Australia, 185–192.

Rudolph, A. L., Lamine, B., Joyce, M., Vignolles, H., & Consiglio, D. (2014). Introduction of interactive learning into French university physics classrooms. *Physical Review Special Topics-Physics Education Research*, *10*(1), 010103.

Rumsey. D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, *10*(3).

Thompson, B. (2002). "Statistical", "practical," and "clinical": How many kinds of significance do counselors need to consider?. *Journal of Counseling & Development*, *80*(1), 64–71.

Willis, D., & Willis, J. (2008). *Doing task-based teaching*. Oxford: Oxford University Press.

WHOM SHOULD (STILL) ATTEND LECTURES AND TUTORIALS?

KWAK Do Won, SHERWOOD Carl, and TANG Kam Ki
The University of Queensland, Australia.
c.sherwood@uq.edu.au

Lecture and tutorial attendance in large first year courses are typically high at the start of the semester and decline as the semester progresses. With most (if not all) lecture and tutorial material available to students online nowadays, the benefits for students attending live lectures and live tutorials is questionable. This study investigates the impacts of class attendance on the learning outcomes in a large first year introductory statistics course. We find that both lecture and tutorial attendance improves students' assessment scores. The effects of class attendance remain positive and statistically significant, but reduce substantially after correcting for the biases. Furthermore, quantile regression results indicate that the effects from attendance decrease as the performance of students improves. The effects of attendance are largest for male, domestic students, and in particular for students at the lower end of the score distribution.

INTRODUCTION

The number of students attending live lectures and tutorials, in a higher education setting, are observed to typically decline across a semester. The relationship between class attendance and student performance in general has previously been investigated by others and suggests a positive relationship exists. However, with many learning resources now available to students online, we explore if low attendances today at live lectures and tutorials have any significant impacts on learning outcomes. Using recent data, we investigate the impact of attendance and student learning outcomes in a large, first year undergraduate introductory statistics course.

LITERATURE AND FRAMEWORK

The effectiveness of students attending live classes has continued to be questioned after the key empirical research undertaken by Romar (1993). A consensus has emerged in the literature that student attendance at live classes is positively correlated with assessment performance. As a result, various interventions have explored making lecture attendance compulsory in an attempt to improve both student attendance and assessment performance (Marburger, 2006; Teixeira, 2016). However, most previous research has looked at student attendance as an input, with performance being an output, and far less attention has been devoted to investigating the possibility of reverse causality, namely that performance is what determines student attendance.

As highlighted by Dobkin, Gil, & Marion (2010, p. 566), previous studies fail to "disentangling the causal effect of class attendance from unobserved factors correlated with both attendance and class performance" such as ability, motivation, and effort. For example, Romar (1993) reported a positive and significant relationship between student attendance and performance, using OLS regression analysis that he conducted, but admitted this may have simply reflected the impact of omitted variables rather than the true effect being investigated.

Stanca (2006) recognised how most early investigations used cross-sectional data and thereby suffered from omitted variable biases. By using panel data instead, he was able to control for time invariant unobservable individual characteristics such as ability, motivation, and effort (sources of self-selection bias) and reported a positive relationship between attendance and performance. On the contrary, Arulampalam, Naylor, & Smith (2012) addressed the endogeneity of class attendance using the random assignment of class timeslots as an instrument. They reported a significant association between attendance and performance, though like Stanca (2006), the affects became less significant after controlling for individual effects. A more recent study by Latif & Miles (2013) focused on an introductory statistics course and reached the same conclusion as the others.

Though recent research efforts have sought to introduce additional controls to account for self-selection biases and address the endogeneity of class attendance, few efforts have been made to identify groups of students who benefit most from attending live lectures and tutorials. In other words, it is unclear as to what the profile is for the student who will benefit most in today's learning environment from attending live lectures and tutorials.

METHOD

This study aims to answer the following research questions:

- Are falling student attendance levels during the semester at live lectures and tutorials cause for concern in today's learning environment, or is it really nothing to worry about?
- Can we find evidence today for a positive and statistically significant relationship between live lecture and tutorial attendance and student performance after controlling for self-selection biases?
- What is the student profile for those who benefit the most from attending live lectures and tutorials?

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

The Setting

This study occurred in the School of Economics at the University of Queensland (UQ) during semester 1 and 2 of 2015, in a first year introductory statistics course. It involved a heterogeneous combined group of 834 students from the two semesters, mostly from economics, commerce, and business management programs, with the rest from engineering, law, arts, and science programs. Most students were first years, aged between 17 and 20, along with a relatively small number of mature aged students, with 30% being international students. In both semesters, the same lecturer, learning materials, assessment tasks, and examination materials were used. All lectures were recorded and available to students at the end of the same day as the lectures. Tutorials were not recorded but the solutions were made available online at the end of each tutorial week. All students had access to all the online materials, regardless if they attended the live lectures or live tutorials.

Course Design and Assessment

The course offered one live lecture each week over 13 weeks of the semester, varying from 1 to 2 hours in length. Lectures covered six broad topics, namely descriptive statistics, probability, the normal and sampling distributions, confidence intervals, hypothesis testing, and simple linear regression. Each topic was presented across two lectures, resulting in 12 different lectures. The last lecture was used for revision. Tutorials, each of 80-minute duration and with a maximum of 25 students, were offered each week to solve problems relating to key concepts covered in each lecture. Assessment consisted of a mid semester exam at the end of week 6 (worth 25%) and a final exam at the end of the course (worth 55%). The remaining 20% was from the completion of six fortnightly online quizzes during the semester. Each quiz was worth 4%, with the best five quizzes only contributing to the final assessment mark.

Attendance Data Collection

Paper rolls were circulated during the first hour of each weekly lecture to record student attendance. A clean copy of the roll was printed each week and included only the student's name and lecture number. The rolls moved from the back of the lecture theatre to the front in the same manner each week, and once collected at the front, all students asked if they had seen the roll (to ensure attendance records were as accurate as possible). Tutorial attendance was also recorded each week by tutors using a paper roll, with attendance data entered into a spreadsheet by tutors and sent to the course administrator approximately every four weeks. Tutors were sent reminder emails regularly during the semesters to ensure they maintained the accuracy of their rolls.

Analysis

Our dataset is a pooled panel data using the results of seven test scores for each student. These test scores include a mid semester exam, final exam, and the first five online quizzes (the sixth quiz was not used in our analysis as it is not attempted by a majority of students). In the estimations, the variables that vary for both tests and students are the outcome variable and the attendance data. In our analysis, we first regress the standardised test score against live lecture attendance and live tutorial attendance, then test fixed effects (TFEs) and individual characteristics using ordinary least square (OLS) estimation. In this specification, we account for individual level differences (being the main source of sample selection bias in this estimation) by using observed student characteristics such as age, gender, birth country, degree program, number of courses completed among others. We then redo the estimations by adding individual fixed effects (IFEs) to better control for individual omitted variables (the main source of sample selection bias in this estimation) that may be correlated with class attendance. Lastly, we estimate a quantile-regression to examine the impact of live lecture and tutorial attendance on students' exam scores.

FINDINGS

With or without IFEs, attending live lectures and live tutorials both have statistically significant effects on test scores at the 5% or higher (i.e. 1%) level. However, the effects are much smaller when controlling IFEs, verifying that unobserved individual characteristics are correlated with class attendance. With IFEs, attending one more live lectures is found to increase the test score by 0.033 standard deviation, and attending one more live tutorial by 0.014 standard deviation. The mean and standard deviation of the final exam (out of 100) is 68.03 and 24.13 respectively for all students in the data set. Thus, attending one more lecture increases the final exam score by 0.80 points, and attending one more tutorial increases the final exam score by 0.34 points. This implies that for a student attending 50% of the 13 lectures and 12 tutorials available during a semester, it would increase their final exam score by 5.2 points from attending the lectures, and 2.0 points from attending the tutorials (compared to a student attending no lectures and no tutorials).

The differential effects across various groups from attending live lectures and tutorials, after controlling for student fixed effects, are presented in Table 1. Firstly, consider the impact of lecture attendance. It can be seen that lecture attendance positively impacts males much more so than females (coefficients for male being 0.504 compared to 0.285 for female), while native English speaking students show more than twice the impact than non-native speakers from attending lectures (coefficients of 0.52 compared to 0.216 respectively). For groups using age, degree program, country of origin, and the number of courses completed, the differences between coefficients are much smaller relative to their standard errors. The impacts of students attending live tutorials is presented in the middle of Table 1. Overall, this reveals small differences across all the groups regarding tutorial attendance when taking into account the standard errors, with the majority of the coefficients having low or no statistical significance. The regression coefficients, for live lecture and tutorial attendance data shown in Table 1, are presented graphically in Figure 1 and Figure 2.

Table 1: OLS Regression coefficients for live lecture and tutorial attendance and student characteristics (controlling for student fixed effects).

	Gender		Age		Program		English speaking		Country origin		Course completion	
	male	female	<19	>=19	bm major	non-bm major	native	non-native	foreign	domestic	>=4	<4
Lecture attendance	0.504***	0.285**	0.373***	0.465***	0.384***	0.482***	0.520***	0.216*	0.406***	0.460***	0.466***	0.363***
	(0.127)	(0.122)	(0.140)	(0.121)	(0.116)	(0.144)	(0.117)	(0.129)	(0.152)	(0.108)	(0.151)	(0.115)
Tutorial attendance	0.136	0.120	0.0913	0.205*	0.100	0.224*	0.195**	0.0354	-0.00176	0.202**	0.0925	0.188*
	(0.110)	(0.131)	(0.125)	(0.115)	(0.112)	(0.127)	(0.0982)	(0.157)	(0.169)	(0.0940)	(0.138)	(0.108)
N	1862	1491	1505	1848	1778	1575	2345	1008	763	2590	1267	2086
R-sq	0.022	0.021	0.017	0.015	0.024	0.022	0.021	0.028	0.030	0.016	0.021	0.020

Note: Robust standard errors are in the parentheses. *, ** and *** denote significance at 10%, 5% and 1% respectively. All regressions include test and individual fixed effects. "bm" stands for "business management".

Using the OLS estimate of attendance on test score, we find the average effect for the coefficient for live lecture attendance to be 0.434*** (0.089) and live tutorial attendance to be 0.168** (0.085), with the standard errors in brackets. In order to obtain richer information about specific individual student performances across the distribution, regarding their attendance at live lectures and tutorials, we conducted a quantile regression analysis. The results are presented in Table 2.

Table 2: OLS Regression coefficients at various percentiles.

	0.1	0.25	0.5	0.75	0.9
lecture attendance	1.30*** (0.16)	1.05*** (0.09)	0.66*** (0.06)	0.29*** (0.04)	0.06 (0.05)
tutorial attendance	0.95*** (0.16)	0.83*** (0.12)	0.47*** (0.06)	0.26*** (0.05)	0.10*** (0.06)

Note: Robust standard errors that are clustered at student are in the parentheses. *, ** and *** denote significance at 10%, 5% and 1% respectively. The number of observations for each quantile is 3,353.

The quantile regression analysis reveals that the lower performing students at the 0.1 percentile benefit the most from attending live lectures and tutorials (coefficients 1.30 and 0.95 respectively). Students at the middle of the distribution gain approximately half the benefit as those at the lower end of the distribution (coefficients 0.66 and 0.47 respectively). For the best performing students at the 0.9 percentile, the effects from attending both live lectures and tutorials are effectively the same and have almost no impact on a student's test score. The data in Table 2 is presented graphically in Figure 3. This visually demonstrates the effect of attendance at live lectures and tutorials on students' performances at different locations in the score distribution.

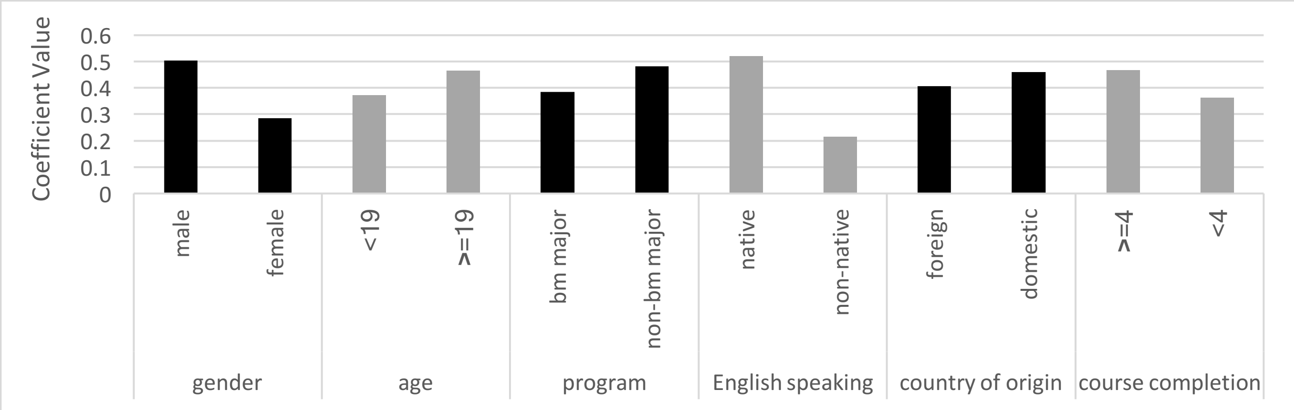


Figure 1: Coefficients for live lecture attendance from regressions.

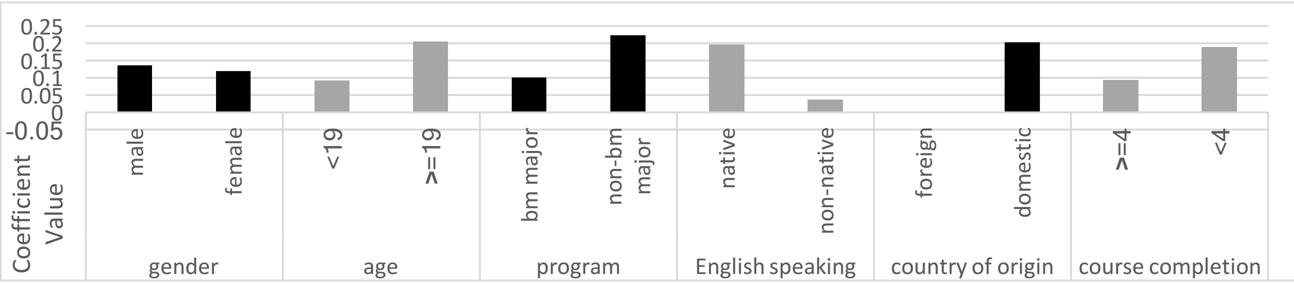


Figure 2: Coefficients for live tutorial attendance from regressions.

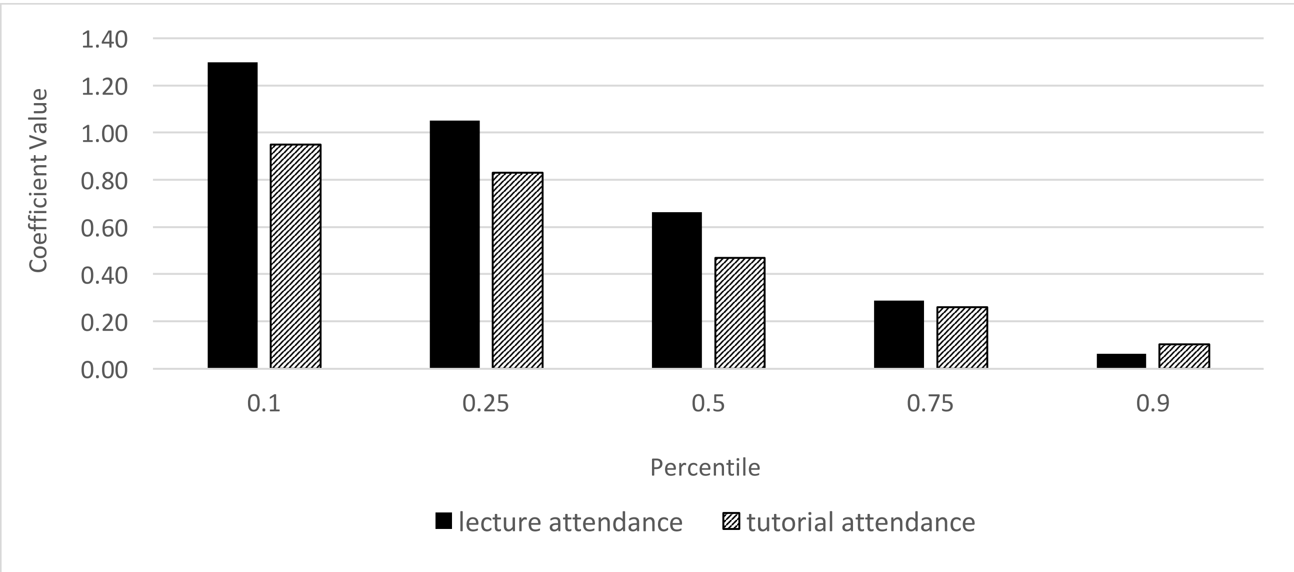


Figure 3: Regressions coefficients at various percentiles.

DISCUSSION

Investigating the impact of student attendance at live lectures and tutorials on learning outcomes is not new. However, in an environment where increasingly learning materials and activities are being made available online, it is timely to investigate the impacts of attendance on student performance using the most recent data. We investigated three research questions and offer answers to these questions in light of our analysis.

The first question wished to investigate whether falling student attendance at live lectures and tutorials across a semester should be cause for concern, or whether it is really nothing to worry about.

Using data from an introductory statistics course, the answer clearly depends on specific individual student characteristics. Our analysis reveals the top 10% of students are not greatly impacted by attending live lectures and tutorials. However, if it is the students in the bottom 50% of the score distribution who are missing live lectures and tutorials, then there is cause for concern. Students in the bottom half of the distribution do clearly benefit through their attendance. So we can conclude that we should not worry about the average attendance of all students, but rather that of the lower performing students. This reflects the concerns of Marburger (2006) who sought to address this issue by making attendance compulsory at live lectures for the lower performing students.

The second question wished to investigate if evidence could be found in today's learning environment for a positive and statistically significant relationship between both live lecture and tutorial attendance with student performance (after controlling for self-selection bias). As reported by various authors since Romar (1993), we do find a positive relationship between student attendance and their performance in today's learning environment, though the effects are reduced when controlling for self-selection bias. We do so by analysing regression coefficients using a fixed effects estimator to account for individual student characteristics. So despite the introduction of newer technologies and increasing access to learning materials online, we find evidence that student performance can be improved by attending live lectures and tutorials.

However, we find that the effects of attendance across the score distribution on student performances are not uniform. This relates to our third research question. Namely, in today's learning environment, whom should attend lectures and tutorials? Through a quantile regression analysis, evidence clearly indicates that students at the bottom of the score distribution benefit most from attending live lectures and tutorials, with the impact of attendance at lectures being larger than tutorials. Indeed, the positive impacts of attending both live lectures and tutorials are clear for the lower performing half of all students. This suggests that live lectures and tutorials in the higher education setting remain an important component of the student learning experience and continues to provide a valuable source of learning for students. It is only the top 10% of students who essentially gain little benefit, in terms of test score outcomes, from attending live lectures and tutorials.

In conclusion, using the most recent data from a first year introductory statistics course, we find that live lecture and tutorial attendance continues to have positive impacts on students' performances in today's learning environment. In particular, for our study, we find evidence that the positive effects from attendance are strongest for male, domestic, lower performing students. In addition, these positive impacts are found to be beneficial to varying degrees for all students across the score distribution, except for the very top performers.

REFERENCES

Arulampalam, W., Naylor, R.A., & Smith, J. (2012). Am I missing something? The effects of absence from class on student performance, *Economics of Education Review*, 31(4), 363-375.

Dobkin, C., Gil, R., & Marion, J. (2010). Skipping class in college and exam performance: Evidence from a regression discontinuity classroom experiment, *Economics of Education Review*, 29(4), 566-575.

Latif, E. & Miles, S. (2013). Class Attendance and Academic Performance: A Panel Data Analysis, *Economic Papers,* 32(4), 470-476.

Marburger, D. R. (2006). Does Mandatory Attendance Improve Student Performance?, *The Journal of Economic Education*, 37(2), 148-155.

Romar, D. (1993). Do Students Go to Class? Should They?, *Journal of Economic Perspectives*, 7(3), 167-174.

Stanca, L. (2006). The Effects of Attendance on Academic Performance: Panel Data Evidence for Introductory Microeconomics, *The Journal of Economic Education*, 37(3), 251-266.

Teixeira, A. A. C. (2016). The impact of class absenteeism on undergraduates' academic performance: evidence from an elite Economics school in Portugal, *Innovations in Education and Teaching International*, 53(2), 230-242.

DIGITAL ASSESSMENT SUBMISSION AND FEEDBACK – THE REWARDS AND ISSUES

PRVAN, Tania[1] and ASCIONE, Judith[2]
[1]Department of Statistics, Macquarie University, Sydney NSW, Australia
[2]Faculty of Education, Science, Technology and Mathematics,
University of Canberra, ACT, Australia
tania.prvan@mq.edu.au

We all know the scenario: students submit a hard copy of a worked assignment, the assignments get allocated and marked - and are often left languishing for pick up. Students don't see their feedback. A variation is for students to submit assignments to a Learning Management System (LMS), for example Moodle, and receive perhaps rubric-based marks and a general statement of feedback. In the case of an uncollected hard copy or with LMS feedback, highly individualized, contextual feedback is lost. We will share our experiences in overcoming the barriers to digital management of effective feedback and what issues remain. This topic is important to universities hoping to implement fully online assessment, personalised feedback, better records management and improved student satisfaction.

Since 2013, in a third-year Macquarie University statistics course all assignments and other assessment have been marked on an iPad, and returned in our Moodle-based LMS. In 2012 and 2013 in a large second-year University of Canberra course all assignments were submitted and returned in Moodle. Tutors did not have to mark electronically but if they chose not to they wore the cost of printing and collation and scanning back into pdf format for returning to the students in Moodle.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

Teaching and Learning Elementary Statistical Concepts through Self-Identified Problems

ABEYGUNAWARDANA, Rushan
Department of Statistics, Faculty of Science,
University of Colombo, Sri Lanka
rab_abey@stat.cmb.ac.lk

Teaching basic concepts of statistics to fresher is a real challenge to any teacher. For the students, they should be able to learning these concepts easily and firmly. However, some students are just to trying to memorise these concepts by targeting the exams. They do not try to understand the real meaning. Homework/tutorials/case-studies are the traditional formative methods use to assess the student's knowledge. In these traditional methods, every students have to solve the same problem given by teacher. This may leads to the problem of plagiarism. Here, it is propose to use a different approach for homework assignments known as self-identified problems. At the end of the each lesion, student should apply the knowledge gain in the class to a real-life applications/problems. They have freedom to obtained/developed questions by themselves. Students can refer different secondary sources (newspapers, books or internet) for this purpose. The length of homework is restricted to one page to maintain the consistency of marking. This method was applied to the science fresher who have similar background knowledge. The marks were compared with the marks of traditional assignments. It is found that the performance of the students were improved with the continuous self-identified problems. They motivated to find applications from there area of interest.

INTRODUCTION

Statistics is a really practical subject, which use in almost all discipline. Statistics has applications in every areas. Educational experts have proposed to use student-centred methods in teaching statistics which has good active student participation. However, it is real challenge to any teacher to increase students' motivation and encouraging them to learn basic statistical concepts (Gomez, 2014). Students can show their statistical literacy in different ways. They should be able to critically evaluate and interpret statistical results derived from the collected data. Communicating statistical information which were obtained by analysing data are also important.

Teaching elementary statistical concepts for the first year undergraduates is a real challenge. First year undergraduates may face difficulties to understand the real meaning of basic statistics. Students try to memorize statistical methods to do their exams but not try to understand the real meaning of it. Conners, *et.al.* (1998) discussed unique challenges faced by statistics teachers in their teaching. Teacher has to prepare teaching materials interestingly and should deliver the statistical concepts in the class room in interesting ways. Students should get motivation to learn further on the statistical concepts in more detail. If the learning materials are uninteresting then students do not get self-motivation to study. On the other hand statistics is mainly based on numbers and arithmetic operations.

Most of the students may have psychological fear to dealing with numbers which is known as maths anxiety. Since the statistics is mainly based on numbers, statistics teacher should be able to handle student's maths anxiety very carefully (Petocz, and Reid, 2008). If a teacher use complex mathematical expressions at the early stage of courses, the students may not get noble motivation to study statistics. Therefore the teacher should be very careful when delivering elementary statistics to fresher. Teacher can indirectly force and motivate students to learn statistics without using direct teaching methods. Teacher can change her teaching styles where the students can memorise the statistical concepts properly (Weldon, 2008). Furthermore, Weldon (2008) stated that it is better to experience early and learn logic later. On the other hand most of the universities using English as their teaching language where it is not the student's native language. Therefore language may be another barrier to get proper understanding of the statistical concepts (Richardson and Zhang, 2008).

In statistical reasoning, it is important to connect the statistical concepts such as variability, uncertainty, and probability for an appropriate interpretations about a summary of a certain data set. Furthermore, it is necessary to combine statistical knowledge to represent and summarize the data correctly. This will leads student's ability of understand, interpret, and explain the statistical results. (Garfield, 2002). In order to build-up competencies of statistical reasoning Campos, Wodewotzi,

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

Jacobini and Ferreira (2014) suggested several possibilities such as; working with real data and relate it to the theories learnt in the classroom, encourage students to interpret, explain, criticize, justify, and evaluate the results. Furthermore, they suggested to use problematize teaching, debates/dialogues among students and between them and the teacher to deliver basic statistical concepts. Teachers can use current and important social and political issues related to the statistical reasoning (Gordon and Nicholas, 2008). To improve the critical thinking ability statistics teachers need to encourage the student to improve their daily critical thinking skills and problem solving skills (Ennis, 1985). However, Reyneke and Fletcher (2014) stated that traditional teaching methods such as pre-reading of the textbook, pre-class homework assignments, weekly tutorials classes and an innovative tutoring systems may not successful in teaching for first year students. Therefore there is need to identify good and appropriate statistical teaching method to fresher.

PROPOSED METHODOLOGY

Homework, assignment and class room test are traditional method use to evaluate students' performance. In the traditional homework method all the students have to solve the same problem given by the teacher. This may leads to the problem of plagiarism. Since the questions were given by the teacher, they are definitely answerable. Sometimes the teacher can give complicated questions which may be difficult to understand and not interesting for students. Furthermore student's interesting may be different form the each other. Some may interesting in the field of astronomy while another may interested in the field of computer science. In the field of sports students' may interested in many different sports such as cricket, football, baseball, golf or athletics. So, it may be useful, if the student have freedom to select questions for their homework from the area of interest. However, it is not easy task to the teacher to give many different questions according to the student's interests. This study consider the effect of such a situation.

Physical science undergraduates who are in the first semester of their first year was considered as the target group for this study. These students were entered to the university based on the performances of their advanced level examination in mathematics stream. The advanced level curriculum focused on applications of mathematics and not given much attention for statistics. The selected group of students do not have sound theoretical knowledge on statistics as they are following the elementary statistics course as a main subject in the first time. Under this elementary statistics course fresher are learning data collection methods, Types of variables, levels of measurements, graphical/numerical representation of data and elementary probability. Total number of students in the class is 120. The total number of lecture hours of the course is thirty hours and they are schedule in two hours per week.

Here it is proposed to use homework assignments with *self-identified problems*. In this method, students were asked to submit homework in every week. They have to find some real life applications/problems related to the knowledge gain in the classroom. Students have to find a solution to these real life problem or question which is found by themselves. They were supposed to apply the knowledge gain in the classroom to solve the problem. Here, student can develop/obtain a question related to their area of interest. These type of homework questions will improve the student's logical thinking ability and calculation skills. Selecting a question which is tally with the knowledge gained is a real challenge to a student. Rather than solving the problem given by the teacher, students have freedom to find a problem according to their area of interest.

The teacher can motivate, encourage and help students to find a real life scenario according to their own interest. The question should be answerable with their knowledge. Students have to continuously working with the concepts they learnt in the classroom. They cannot postponed to understand the concepts till the end of the semester. This formative continuous assessment method will increase their active participation in the class room. Every student can find different problems with their interest. Students can refer newspapers, books, and internet or create their own problems. So they can directly related the concepts discussed in the class room to their field of interesting. They do not depend on the problems or examples given by the teacher.

EVALUATION CRITERIA

Marking assignment is a time consuming and tired-full task for any teacher. To maintain the consistency of marking one person evaluated all homework assignments. Since students can solve any question it is not easy task to evaluate these assignments. To overcome the problems in evaluation process students were instructed as follows: the length of the homework assignment is one page and they should provide the complete reference. If they produced lengthy answers then marks were reduced. So they have to "answer to the point". To check the accuracy of the question the complete reference is requested. Students were informed that they are getting marks for both question and answers. The both question and answer should be clear enough to read and understand to the examiner. It was found that students were given interesting real-life scenario. Since, students have to refer many sources and many questions to select one problem they will automatically enhance the ability of statistical thinking. This will indirectly encouraged the student to read statistics books. After few weeks it was noted that the students come up with self-developed questions. Students had good discussion with the teacher to enhance their idea and to apply statistical methods to solve the problem. By reference many sources to find a problem students can enhance their knowledge. Teachers are also getting a chance to enhance their knowledge and applications through many different questions submitted by the students.

DEALING WITH PLAGIRISM

Plagiarism is one main problem in many homework assignments. In the traditional assignment methods students are getting same questions to solve. So there is a high chance to copy answers from the other students and difficult to check the plagiarism. In the proposed method students are not getting same set of questions. Therefore the occurrence of plagiarism can be reduced using the proposed method. Students have freedom to find their own questions for the evaluation. It can be assume that the chance of selecting the same problem by two different students is less. Furthermore it can be assume that, although the same question is selected by two or more students, chance of giving the same answer is less. Here the same examiner doing marking the similar answers and similar questions can be identified. The next issue is students may get both questions and answer form the websites, books and other sources. To overcome this problem student were asked to give complete reference of question. If they were taken the question from a website, they have to give complete web address; if it is from a book they should provide all necessary information including page numbers. Examiner can check the truthfulness of the question though the reference, at least for randomly selected or suspected questions. Students were not allowed to submit the sample questions given in the websites or books. If the student did not provide the complete reference or the examiner were not able to find the source then the zero marks were given.

RESULTS

The lectures of elementary statistics course considered in this study were schedule in fifteen weeks and two hour per week. Students were asked to submit ten homework assignments. It was found that the student's performances were gradually increased from the first homework to the last homework. At the begging most of the students were developed simple real life applications while at the end of the semester they were selected more complicated situations. Students used different sources to obtain the real life applications according to their interesting. Mainly they used web sites, textbooks, newspapers/published projects, and self-developed questions. The percentages of the different sources used for the first, fifth and tenth homework assignments are given in the Table 1.When compared to the first homework in the tenth homework students used more self-developed questions.

Table 1. Summary of sources for first, fifth and last homework assignments categories

Homework	Homework number	Websites	Textbook	Published projects	Self-developed questions	Total (%)
First	HW-01	66.7	16.7	5.6	11.1	100
Fifth	HW-05	48.5	24.3	2.9	24.3	100
Tenth	HW-10	44.4	18.5	7.4	29.6	100

The final exam results were compared with the progress of submission of the homework assignments. In 2015 traditional homework assignments were given to the students who followed the same elementary statistics course. Three class room tests and two take-home assignments were given by the teacher and all the students answered to the same set of questions. But in 2016 the new homework assignments method was applied. Pearson correlation coefficients were calculated between the assignments/homework marks, final paper marks and final marks. The obtained results were given in Table 2 and can be compared in 2015 and 2016.

Table 2. Correlation between marks in 2015 and 2016

Year	Type of evaluation	Assignments Marks	Final Paper Marks	Final Marks
2015	Assignments Marks	1		
	Final Paper Marks	0.27	1	
	Final Marks	0.45	0.98	1
Year	Type of evaluation	Homework Marks	Final Paper Marks	Final Marks
2016	Homework Marks	1		
	Final Paper Marks	0.47	1	
	Final Marks	0.71	0.78	1

According to the Table 2, it can be seen that the correlation between homework marks and final marks is higher in 2016 (0.71) than the correlation in 2015 (0.45). Correlation between final paper marks and homework marks is also higher in 2016 (0.47) than the correlation in 2015 (0.27). That indicates when the homework marks increasing both final paper marks and final marks were increasing in 2016 than 2015. However, the correlation between final marks and final paper marks is high in 2015. This may due in 2015 students tried to improve their final marks through the final test. Here in 2015 and 2016 two group of first year students were studied. However, it is reasonable to assume that that their initial knowledge is same and no any other influencing factors for their results.

CONCLUSION

Teaching elementary statistics is a real challenge to teacher and learning statistics is a challenge to the fresher. Good motivation is required to improve the statistical thinking. Here it is proposed to apply the statistical concepts learnt in the class into real world problems. Students have freedom to select any issue according to their interest. This can be used to avoid the issue of just answering questions given by the teacher. Traditionally, teachers are giving homework questions to directly apply the knowledge gain in the class. This will help to improve student’s calculation skills. But every student are not interesting to study the applications given by the teacher. If students can identify a real life problem according to their interest and apply the statistical knowledge gain in the classroom then they will get self-motivation. Based on the summary measures derived on marks of the homework assignments and final marks it was found that the new method have clear advantages

REFERENCES

Campos, C. R., Jacobini, O. R., Wodewotzki, M.L.L., and Ferreira, D. H. L. (2014). Environmental interface in teaching statistics. *Proceedings of the Ninth International Conference on Teaching Statistics.*

Chance, B. L. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment. *Journal of Statistics Education*, Vol. 10, No.3.

Conners, F.A., Mccown, S. M., and Ewoldsen, B. R. (1998). Unique challenges in teaching undergraduate statistics, *Teaching of Psychology*, Vol. 25, No.1, pp 40-42.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership,* Vol. 43, No. 2, pp 44–48

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education,* Vol.10, No.3.

Gomez. R. (2014). Teaching innovations in introductory statistics courses, *Proceedings of the Ninth International Conference on Teaching Statistics*.

Gordon, S., and Nicholas, J. (2008). Why do statistics educators use examples to teach statistics? *Proceedings of the 6th Australian Conference on Teaching Statistics*.

Petocz, P., and Reid, A. (2008). On becoming statistician. *Proceedings of the 6th Australian Conference on Teaching Statistics*.

Reyneke, F., and Fletcher, L. (2014). The impact of an inverted traditional teaching model on first level statistics students, *Proceedings of the Ninth International Conference on Teaching Statistics*.

Richardson, A., and Zhang, A. (2008). Language support for statistics learners. *Proceedings of the 6th Australian Conference on Teaching Statistics*.

Weldon K. L. (2008). Experience early, logic later. *Proceedings of the 6th Australian Conference on Teaching Statistics*.

A socio-public health data-based introductory statistics course

Murray Aitkin

murray.aitkin@unimelb.edu.au

School of Mathematics and Statistics

The University of Melbourne

Australia

Turbulence in the profession!

- A widespread dissatisfaction with the present curriculum.
- p-values banned by a minor psychology journal.
- ASA statement on p-values recommending their replacement – but by what?
- Q-step support by ESRC and the Nuffield Foundation of the development of new statistics or “quantitative methods" courses for social science graduate and undergraduate students by the social science departments themselves, without the participation of statistics departments.
- Statistics departments sidelined? A warning bell ringing!
- Special issue (November 2015) of The American Statistician on statistics and the undergraduate curriculum.
- The Statistical Society of Australia held a two-day workshop (June 2016) to develop proposals for modernising statistics courses at all levels of school and University.

Do we need a new introductory stat course?

The editors of the TAS special issue focussed on second- and higher-level courses:

> Likely the first and most important place to start the curriculum conversation is with the courses that follow an introductory statistics course. (N.J. Horton and J.S. Hardin, Special issue editors)

George Cobb, Mount Holyoke College, did not agree with that:

> Mere renovation is too little too late: we need to rethink our undergraduate curriculum from the ground up.

The Special Issue has a curious absence of discussion of content for the first course, apart from issues like bootstrapping replacing parametric inference.

The SSA subgroup which considered the undergraduate curriculum came to a consensus on the first course: it should be data-, models-, and probability-oriented.

What's in the traditional intro stats course?

- Population and sample; descriptive statistics – mean and variance – of a sample;
- simple probability; the normal distribution;
- sampling distribution of the sample mean for the normal distribution;
- the Central Limit Theorem and the large-sample normal distribution of the sample mean;
- the z-test for a hypothetical mean with known variance;
- the t-test for a hypothetical mean with unknown variance;
- confidence intervals for the mean and for a population proportion;
- the t-test and confidence intervals for the difference of two means and of two proportions;
- simple linear regression and correlation.

What's wrong with this? – what is missing?

- A data base. Students see small samples, and may have to collect data themselves, but do not see a realistic large survey or small population data base.
- The research questions – why do we have these data? Who wants to know?
- The importance of the sample design – (not just for sampling distributions).
- An understanding of probability (though sampling distributions are expected to be understood).
- The idea of a probability model.
- Any principles for statistical inference (the Central Limit Theorem is not an inferential principle).

An ancient syllabus

Apart from the t-test, this intro stat curriculum is pre-1900 (Student's use of the t-test was published in 1908).

How can all these extra topics be fitted into a course which is already overstuffed?

They can't, but space can be made by limiting the range of models and analyses.

The course described here had two components: a data base with research questions, and a general survey of the importance of probability and statistics in daily life.

A successful non-standard course – data base

- The data base – a small population of 1296 families in a Child Development Study at UC Berkeley.
- The research questions – what is the effect of mother's (and father's) smoking on their child's development through
  - birthweight;
  - physical and intellectual development at age 10?
- The study design, selected from a survey of 15,000 families.
- Random sampling from the database.
- Sampling binary attributes: the binomial distribution.
- Inference from sample to population: the likelihood function.
- In a random sample of 20 families, 3 mothers were smoking at the diagnosis of pregnancy. What can we say about the proportion p of mothers in the population smoking at pregnancy diagnosis?

The importance of probability and statistics in daily life

- The dangers of voluntary response and other non-random sampling methods (from David Moore's great book).
- Sample surveys – can we believe what we read? (Shere Hite's non-random surveys of sexual life in American women)
- Randomized clinical trials – why randomize, and why a placebo? What is informed consent, and why is it necessary?
- The ethics of small trials.
- The notorious Sally Clark case – imprisonment for murder from a P-value.

How to handle inference? The 2002 course was frequentist, with a lot of hand-waving.

How to handle inference? – Bayesian

New intro Bayesian book: Statistical Rethinking: a Bayesian course with examples in R and Stan. Richard McElreath, Chapman and Hall 2016.

- Likelihood function conveys the data information. How to turn this into a probability statement?
- Prior distribution on p.
- Finite population of size N, so p must be one of the values $0/N, 1/N, ..., N/N$.
- If no prior preference for one of these over another, all have equal prior probability $1/(N+1)$.
- Bayes's theorem provides the update from prior to posterior.
- How to demonstrate or justify it? – the screening test – a widely used and useful example.

The screening test

- A condition C is uncommon – present in 2% of the population.
- For those people who have the condition, the screening test gives a true positive result 95% of the time.
- For those people who do not have the condition, the screening test gives a false positive result 10% of the time.
- We test a population of 1000 people in a small town. The true positive and false positive rates apply to this town population.

What do we conclude from the results of the screening test?

Venn diagram – contingency table

We write + if the test result is positive, and - if the test result is negative.

We write yes if the person tested has the condition, and no if the person tested does not have the condition.

The test results for the town are:

	condition	present	
test	yes	no	Total
+	19	98	117
-	1	882	883
Total	20	980	1000

What to conclude?

How effective is the screening test?

	condition	present	
test	yes	no	Total
+	19	98	117
-	1	882	883
Total	20	980	1000

- There are 20 (2%) cases and 980 (98%) non-cases.
- Of the 20 cases, 19 (95%) are correctly identified – true +.
- Of the 980 non-cases, 98 (10%) are incorrectly identified – false +.
- Of the 117 + tests, 19/117 (16%) are from people who had the condition.
- Of the 883 - tests, 1/883 (0.11%) are from people who had the condition.

Conclusion

- If your test was negative, you can be reassured that you are very unlikely to have the condition.
- If your test was positive, the probability that you have the condition is only 16%.

So what was the point of the screening test?

Many students find this shocking.

The true positive rate is 95%!

So surely almost everyone testing positive must have the condition!

The fallacy of the transposed conditional.

Bayes's theorem from a 2x2 table, without algebra.

Student response

Very favourable:

- Students enjoyed real dice throwing in class to draw their random samples – convincing randomness.
- The examples from daily life were found fascinating.
- The research questions from the data base were found engaging and relevant.
- One student's comment:

  I can see the relevance of probability and statistics in every aspect of life.

☺ Thank you! ☺

FORMATIVE FEEDBACK: SUPPORTING AND ENHANCING TEACHING AND LEARNING IN STATISTICS

ALDRIDGE, Victoria and WADE, Angie
Centre for Applied Statistics Courses, Institute of Child Health, UCL, London, UK
v.aldridge@ucl.ac.uk

Practice and formative feedback are valuable and practical tools when teaching statistics to non-statisticians. Non-specialist learners of statistics frequently seek statistical training to acquire practical skills and experience, rather than to pass exams; therefore, these learning outcomes should be paramount. Opportunities to apply new knowledge, solve problems, and evaluate these applications are essential to this aim, and can be achieved via formative feedback. Methods and technologies used to produce formative feedback (e.g., voting, embedded practical activities, in-teaching discussion and annotation) enhance traditional lecture-style teaching, creating a higher level of interactivity that helps learners to apply, assimilate, and retain information. Teachers can gauge the understanding of individuals and groups and tailor teaching time accordingly, and the process of evaluation and emphasis on reasoning behind 'correct' and 'incorrect' answers help to amend errors in understanding at an early stage. The use of formative feedback requires a flexible approach to topics, emphasis, and pace, but the benefits for teaching, learning, and practice in statistics training can be substantial; keeping learners involved throughout, ensuring that delivery is appropriate to specific groups, and improving inclusivity across abilities.

WHAT IS FORMATIVE FEEDBACK?

Whilst there are many definitions and discussions around what formative feedback is (Black & Wiliam, 2009), it has been briefly outlined as "*evaluation carried out in the course of an activity in such a way that the information obtained is used to improve learning and/or instruction*" (Coffield, Moseley, Hall, & Ecclestone, 2004). Formative feedback is also known as 'formative assessment'; 'feedback' was chosen for the current paper to emphasise the importance of dialogue and interaction between teacher and student(s) within this process. Broadly speaking, formative feedback encompasses methods and techniques such as practical exercises, quizzes, and problem-based tasks that 1) give students the opportunity to apply what they have learned, 2) enable students and teachers to review and evaluate this application, and 3) modify teaching and learning in response to this information, as appropriate for the individual or group. Methods for gathering and using feedback are relatively non-prescriptive, but importantly, formative feedback takes places during the learning process, at a time where the student still has an opportunity to correct mistakes in their learning before applying it in the real-world (Yorke, 2003). Constructivist theory notes the importance of working with the student during this process to understand how they arrived at a particular response and working from there to modify teaching and learning (Von Glasersfeld, 1989). Unlike summative assessment (i.e., post-course tests/feedback), which is said to offer a résumé of knowledge at the end of a course, literature purports a role for formative feedback in learning development and improving continuity of learning beyond the educational context (Black & Wiliam, 2009; Boud & Falchikov, 2007).

Formative feedback requires input, effort, and flexibility from teachers, but to derive the greatest value to learning, student participation is essential (Meletiou-Mavrotheris & Lee, 2002). A key factor in achieving student engagement is to reduce the fear that many individuals have of getting questions and tasks wrong, by removing the negative consequences of such (e.g., embarrassment, low grades and classifications, failure), and instead presenting them as opportunities to correct and improve learning. As such, the collection and use of formative feedback should be largely informal; review of answers may be private to the individual or may be fed back to teachers on an individual or group basis, the process may be repeated throughout a learning occasion, and teachers share the responsibility for correct and incorrect learning with the students. Students need to be able to identify and communication their strengths, weaknesses and if relevant what is not working for them. Equally, teachers need to be able to encourage this information, absorb the feedback, and be adaptable to student diversity in learning (Ramsden,

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

Margetson, Martin, & Clarke, 1995).

WHY IS FORMATIVE FEEDBACK ADVOCATED IN EDUCATION?

Formative feedback has the potential to improve the learning experience for teachers and learners, increase the satisfaction levels of both parties, and foster retention of knowledge and practical skills beyond the learning environment. Both the nature and practical features of the formative feedback process appear to be beneficial. Research highlights the fact that individuals vary naturally in their learning style, pace, and capacity (Coffield et al., 2004; Kam et al., 2005). As such, single modes of teaching like the traditional lecture model are unlikely to satisfy, or prove effective, for more diverse and expectant students, motivated by social, economic and professional pressures (Bligh, 2000). Furthermore, emphasis on skills development and practical application of knowledge is ever-increasing in education, meaning that students are no longer fulfilled by a single linear opportunity to learn, summarised by a single grade that may not validly reflect enduring learning or the ability to apply knowledge within the wider working world (Garfield, 2006; Garfield & Gal, 1999; Illeris, 2009; Sowey, 1995). The type and level of motivation to learn can also vary greatly among students. This can impact on an individual's ability to achieve thorough understanding of concepts, lasting comprehension, and application of new skills in practical and professional settings, which rely on engagement and motivation to learn (Biggs & Tang, 2011). Fatigued or less intrinsically motivated students, for instance, require variety and novelty in teaching methods to maintain engagement. Light, Calkins, and Cox (2009) concluded that the optimal model for learning was an 'engaged' model, in which the lecturer promotes engagement with students, and the focus of lecturing is on dialogue between teacher and student(s). Techniques used in formative feedback typically include a range of activities, use of multiple media, and a high level of interaction in teaching, which can provide the variety in learning activities required by diverse learners in an effective and enjoyable way, and help maintain wider engagement in learning and knowledge development (Beeland, 2002; Bryant & Hunton, 2000; Lerman, 1996). These methods also help teachers to better listen to students, and to be dynamic, reflective and adaptable to different learners and contexts; characteristics which are said to represent good teaching (Ramsden et al., 1995).

Aside from specific tools and activities, the timing of formative feedback (the fact that it occurs during teaching and learning), also plays an essential practical and pedagogical role. Giving learners opportunities to test new knowledge and skills and to reflect on them within the educational setting can help students to confirm correct learning and, more importantly, to understand where errors have been made, what those errors are, and how to go about adjusting their thinking to correct them. Feedback to the teacher can also inform what topics to revisit, whether alternative teaching methods and explanations may be beneficial, and how subsequent teaching is delivered (e.g., pace, focus, etc.). Simply identifying mistakes may be enough to prompt some students to identify for themselves why and where their thinking went wrong, but for other students it is helpful to move through this process together. Rather than simply knowing that they were correct or incorrect, which has little functional use in professional application, feedback within teaching gives students the means to meaningfully reorganise their knowledge (Garfield & Ben-Zvi, 2007; Von Glasersfeld, 1989). Generally, formative feedback helps teachers and learners to work together in an iterative and interactive way to achieve these goals, and ultimately, more accurate and durable learning.

WHAT ARE THE ADVANTAGES FOR TEACHING IN STATISTICS?

Whilst formative feedback can have a positive impact on teaching and learning outcomes across all subjects and fields of practice, certain factors relating to statistics, particularly teaching statistics to non-statisticians, make formative feedback of particular benefit in this field. Firstly, statistics is a topic shrouded in anxiety for some students, who are (or perceive themselves to be) less number literate. This anxiety can create a barrier to learning and heightened fear around engagement and being asked questions. Research in this field found that it was easy for teachers to underestimate student difficulties in understanding of statistics and overestimate how well they understand basic concepts (Garfield & Ben-Zvi, 2007), possibly in part because of the low profile that anxious students maintain. While formative feedback cannot remove statistics anxiety, it can be used as a means to share the burden of comprehension between teacher and student, and create a more social learning environment, so that the student does not feel isolated in their learning. Research in statistics teaching advocates the use of cooperative, social, and interactive teaching activities, which facilitate more open forms of teaching, to increase learning and retention (Giraud, 1997; Light et al., 2009; Magel, 1998; Meletiou-Mavrotheris & Lee, 2002), and ensure that the student is an active part of their own learning (Kam et al., 2005). Maintaining informality in formative feedback and removing elements of scoring and consequence from activities and dialogues can also encourage engagement from anxious students. Students can learn that getting questions or tasks incorrect doesn't lead to negative outcomes, and can in fact lead to further, alternative, and perhaps preferable explanations from teachers, and additional attempts at the task.

Secondly, for many individuals learning statistics, it is a secondary subject/skill that is required in order to function or progress in their primary field of work or study. Lack of intrinsic motivation to learn statistics can have a negative influence on the level of engagement that learners have for statistics, and therefore the amount of information that is absorbed and comprehended for future application. Statistics is also often considered a 'dry' subject making it more challenging to convey fundamental concepts in innovative and stimulating ways for those not already interested. Formative feedback can help to break-up lecture-style teaching and offer variety in the methods used to teach and learn. While this may not alter an individual's intrinsic motivation, it can help to increase engagement and maintain focus during teaching. Technology is also increasingly adopted within teaching to collect, collate, and convey feedback, which helps to increase the level of active participation in learning and comprehension of concepts (Bryant & Hunton, 2000). Research has demonstrated significant improvements in learning capacity and information retention when active and interactive methods are incorporated into statistics teaching (Giraud, 1997; Magel, 1998; Meletiou-Mavrotheris & Lee, 2002). Evidence also suggests that novelty and enjoyment in themselves can improve retention of learning (Sowey, 1995).

Finally, statistics is widely studied for practical purposes, with learners needing to apply the skills and information that they have acquired in real-world academic, educational, and commercial contexts. Therefore, it is essential that learners can correctly and appropriately apply and retain what they have learned; and hence, it is essential that they are actively involved in learning and are given hands-on experience to construct knowledge and develop skills (Garfield & Ben-Zvi, 2007). Summative assessment may ascertain the extent to which these learning outcomes have been achieved, and perhaps indicate where mistakes were made, but it offers no opportunity to correct these mistakes or improve applied skills. Statistics students have been shown to retain incorrect reasoning for statistical concepts based on inaccurate prior knowledge (Garfield & Ben-Zvi, 2007), which amplifies the need to identify errors as early as possible in order to correct them in the student's lasting comprehension. Garfield and Ben-Zvi (2007) also found that learning of statistics can be enhanced when students are offered constructive feedback on their progress, given consistent advice, and helped to identify their own errors in their learning. These factors all signal the need for formative feedback in statistics teaching. This process can create valuable opportunities for students to put new knowledge into practice, and for results to be fed back to the learner and the teacher at a time when errors can still be identified and corrected (Garfield & Ben-Zvi, 2007; Von Glasersfeld, 1989). Formative feedback also offers the hands-on practice required for skills development, opportunities for immediate 'marking' and feedback, and further explanations that can help to clarify why a wrong response was given and how to restructure existing knowledge to correct it (Illeris, 2009).

PRELIMINARY EVIDENCE FROM AN AUDIT OF FORMATIVE FEEDBACK

For a number of years, methods of formative feedback (practical exercises, group critical appraisal, interactive voting, and live lecture e-annotations) have been incorporated into our teaching of statistics to non-statisticians. Furthermore, interactivity and open dialogue between teachers and students has been actively encouraged throughout all teaching sessions. Despite the long-standing belief that these methods add value to our teaching, we had only anecdotal evidence and personal perceptions on which to base this belief. Therefore, an audit of teaching methods was initiated from summer 2015, which aimed to more objectively assess the value of formative feedback from the learner's perspective. All delegates who had undertaken short-courses featuring several combined formative feedback techniques between July 2015 and June 2016 (N=183, from 8 courses) were contacted and asked to complete a brief questionnaire about their experiences and perceptions of the methods used during the course. For each method, respondents completed a visual analogue scale from *'really disliked' (0)* to *'really liked' (100)*, and were asked to give information about their reasons for the score. The response rate was relatively low for this initial sweep (9.3%) and hence this process will be continued following future courses in order to build a stronger evidence base. However, the information received from these responses was highly consistent, both between-learners and in the context of existing teaching and learning literature, offering indicative early information regarding the perceptions and value of formative feedback techniques. Around two thirds of respondents reported that at least one activity was new to them; with novelty most often relating to interactive voting and e-annotations. These novel activities were enjoyable for learners, with delegates reporting that they hadn't seen them before so enjoyed the change and found them fun. Furthermore, all respondents considered one or more of the specific methods to have been useful to their learning. A selection of direct quotes from respondents is presented below in support of the proposed benefits of each activity.

Using an anonymous voting system to answer applied statistical questions appeared to help increase the level of interactivity and learner engagement in teaching, by encouraging active participation and removing negative consequences of incorrect answers. The system seemed to be enjoyable for learners, with a median liking for this activity of 97/100 (IQR 80-100), and was also reportedly valued by them for facilitating corrections to their own knowledge and practical skills. Little information was reported about the frequency of these types of activities; while this may reflect satisfaction with the current approach, a larger body of data is necessary before we can begin to ascertain the optimal way of integrating interactive voting into teaching:

"Ensured we stayed engaged and didn't fall asleep"
"I enjoyed it very much as I am generally reserved so I have always found difficult to publicly speak. I also appreciated to hear the explanation of all the answers that were displayed"
"Gave people opportunity to voice their thoughts anonymously"
"Very useful, as [tutor] was able to go through the incorrect selections and address why those were wrong"
"Voting pads were useful to make it interactive and to allow the instructor to address misunderstandings"
"I think it really helped to indicate what we understood and the explanation of why an answer was wrong was also very informative"
"Gave a practical application to theoretical knowledge and encouraged discussion. Also served to highlight where you may have misunderstood some elements of teaching to that point"

The use of live annotations, which was novel for just under half of respondents, allows teachers to create their own ad-hoc explanations, with the aim of guiding learners step-by-step through statistical concepts and problems. Provisional reports suggested that the intended aims of this method were met, with students perceiving the benefit of real-time, interactive and individualised explanations, and conveying a high level of liking for this method (Median 92, IQR 76-94). It was also suggested that annotations provided an additional visual dimension to learning:

"Good for breaking down the theory into simpler, digestible terms"
"It has been really useful to use e-pen to personalize the annotation according to the need of the teacher or the classroom"
"It was helpful because it meant we had the lecturer doing the exercise with us, step by step"

> *"Clearer to follow than pre-printed formulae"*
> *"Was good to see the tutor's processes of working out the examples and could see the best method of doing things/how to make your own method better or where you went wrong"*

Within the current sample, practical activities and hands-on critical appraisal were more familiar to many learners. However, they still appeared to be valued and enjoyed, with median liking scores of 88 (IQR 80-100) and 79 (IQR 73.5-91.3) respectively, and >80% of students specifically mentioning practical activities when asked which activities they found useful during their learning. These activities gave delegates opportunities to apply the knowledge they had gained, and to evaluate the success of these applications and correct mistakes in comprehension. This helped to ensure that they were correctly equipped with knowledge and skills to apply in the 'real-world'. The active nature of practical activities also appeared to help motivate student engagement, and the more relaxed learning environment that they create fostered more natural social learning by making it *"fun to interact with course mates"*:

> *"Vital in actually putting the theory into practice and internalising what had been discussed and learnt".*
> *"They tested ones understanding and gave the necessary feedback on responses … they were a valuable learning aid"*
> *"Gave us the opportunity to put into practice what the tutors had just said and as it was fresh in the memory reinforced my learning better. Reviewing the answers was good as it was all worked through – could see what you did right and what you did wrong (and where to correct the problem where necessary)"*
> *"Very much helped to apply the lectured information, also served to encourage discussion amongst students/delegates in an informative and productive manner"*
> *"Good for putting theory into context … understanding how to analyse scientific papers as well as getting a feel for how statistics and research methods are discussed in real life situations"*

It is important to note that respondents appeared to value the more traditional lectures that underpinned teaching, for providing the building blocks on which to learn and develop; *"We need the lectures to get a base understanding"*. However, despite variation in the specific activities that individuals preferred, and the extent to which activities were enjoyed, the current cross-section of learners were unanimous in reporting that the course would not have been as enjoyable or effective without the additional activities and that these were the factors that added value to the course and to their learning. In support of this assertion, the median anticipated liking for the lecture without any activities was just 18/100 (IQR 5-26). This offers provisional support for the continuation of lecture-style teaching when it is supplemented with active and interactive learning components.

> *"The other components allowed us to interact with the material and instructor to better assess and to improve our knowledge."*
> *"…doing the engaging stuff actually made me think and understand, even enjoy, the statistics!"*
> *"I really benefitted from the interaction that these things bring"*
> *"The other elements, apart from being useful in themselves added vital variety which is a proven aid to learning"*
> *"…a mix of activities breaks up the day and caters to differing learning styles"*

CONCLUSIONS

The preliminary evidence offered by the current teaching audit supports existing research in the field in suggesting that interactive activities and technologies are positive and effective ways to gather and use formative feedback, and that formative feedback processes contribute perceptibly to student learning. The available information suggests that variety, anonymity, and the opportunity to practice, were among the key advantages of formative methods reported by learners in statistics teaching. The foundation of evidence we now have to build on, tentatively suggests that benefits of formative feedback are perceived by students, who maintain focus for longer, apply and better retain new knowledge, and enjoy the novelty of the process. In our experience, the benefits are also felt by teachers, who receive a higher level of engagement, feedback, and positivity from students.

REFERENCES

Beeland, W. D. (2002). *Student engagement, visual learning and technology: Can interactive whiteboards help.* Paper presented at the Annual Conference of the Association of Information Technology for Teaching Education.

Biggs, J. B., & Tang, C. (2011). *Teaching for quality learning at university: what the student does* (4th ed.). Maidenhead: Open University Press/McGraw-Hill Education.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education), 21(1), 5-31.

Bligh, D. A. (2000). *What's the Use of Lectures?* San Francisco Jossey-Bass.

Boud, D., & Falchikov, N. (2007). *Rethinking assessment in higher education: learning for the longer term*. London; New York: Routledge.

Bryant, S. M., & Hunton, J. E. (2000). The use of technology in the delivery of instruction: Implications for accounting educators and education researchers. *Issues in Accounting Education, 15*(1), 129-162.

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). *Learning styles and pedagogy in post-16 learning: a systematic and critical review*. London: Learning and Skills Research Centre.

Garfield, J. B. (2006). *Collaboration in statistics education research: Stories, reflections, and lessons learned. In Proceedings of the Seventh International Conference on Teaching Statistics (pp. 1-11).* Paper presented at the Seventh International Conference on Teaching Statistics, Salvador, Bahia, Brazil.

Garfield, J. B., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75*(3), 372-396.

Garfield, J. B., & Gal, I. (1999). Assessment and statistics education: Current challenges and directions. *International Statistical Review, 67*(1), 1-12.

Giraud, G. (1997). Cooperative learning and statistics instruction. *Jnl of Statistics Education, 5*(3), 1-13.

Illeris, K. (2009). *Contemporary theories of learning: learning theorists... in their own words*: Routledge.

Kam, M., Wang, J., Iles, A., Tse, E., Chiu, J., Glaser, D., . . . Canny, J. (2005). *Livenotes: a system for cooperative and augmented note-taking in lectures.* Paper presented at the SIGCHI conference on Human factors in computing systems.

Lerman, S. (1996). Intersubjectivity in mathematics learning: A challenge to the radical constructivist paradigm? *Journal for Research in Mathematics Education, 27*(2), 133-150. doi:Doi 10.2307/749597

Light, G., Calkins, S., & Cox, R. (2009). *Learning and teaching in higher education: the reflective professional*. Los Angeles: Sage.

Magel, R. C. (1998). Using cooperative learning in a large introductory statistics class. *Journal of Statistics Education, 6*(3).

Meletiou-Mavrotheris, M., & Lee, C. (2002). Teaching students the stochastic nature of statistical concepts in an introductory statistics course. *Statistics Education Research Journal, 1*(2), 22-37.

Ramsden, P., Margetson, D., Martin, E., & Clarke, E. (1995). *Recognising and Rewarding Good Teaching in Australian Higher Education*. Canberra: Australian Government Publishing Service.

Sowey, E. R. (1995). Teaching statistics: Making it memorable. *Journal of Statistics Education, 3*(2).

Von Glasersfeld, E. (1989). Cognition, construction of knowledge, and teaching. *Synthese, 80*(1), 121-140.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45*(4), 477-501.

THINKING CRITICALLY ABOUT THE 1936 US PRESIDENTIAL ELECTION POLLS

FINCH, Sue and GORDON, Ian
Statistical Consulting Centre
The University of Melbourne
sfinch@unimelb.edu.au

The story of the rivalry between George Gallup's American Institute of Public Opinion and The Literary Digest in pre-election polling for the 1936 US Presidential race is popularly used in statistical education. Gallup's prediction of The Literary Digest poll result, and the massive size of the error in The Literary Digest's prediction, provide dramatic and memorable lessons. In contrast to The Literary Digest, the Gallup poll is often seen as getting prediction of the 1936 election result "right". The reasons for the failure of The Literary Digest have been discussed by a number of authors. We suggest that there are many useful lessons in thinking further about the Gallup poll that are yet to be highlighted in statistical education. The story of the polls and the 1936 US Presidential election can suffer distortions in the re-telling so here we focus on providing data and insights from original sources, where possible.

AMERICA SPEAKS

America Speaks was a national weekly poll of public opinion, carried out by the American Institute of Public Opinion (AIPO) – the Gallup poll. The results were published in a weekly column, also called *America Speaks*, in syndicated newspapers. George Gallup established the AIPO in 1935, and Gallup and the AIPO quickly became synonymous. The AIPO conducted its polls using mail-outs and face-to-face interviews. When *American Speaks* first appeared in syndicated newspapers on 20 October 1935 (e.g. Pittsburgh Press), Gallup wrote that his "scientifically correct" poll should not be confused with "straw votes" or straw polls which were casual or ad hoc. Such polls claimed to rely on large numbers for their accuracy.

Gallup had given the newspapers subscribing to *America Speaks* a money-back guarantee that the AIPO survey results would be more accurate in predicting the outcome of the 1936 Presidential election than those of *The Literary Digest* (Rich, 1939; Ohmer, 2006). The 1936 Presidential election was a race between Republican candidate Governor Alfred Landon and incumbent Democrat President Franklin Delano Roosevelt.

THE LITERARY DIGEST

Popular and widely read, *The Literary Digest* was a weekly magazine that had conducted opinion polls since 1916. *The Literary Digest* mailed out ballot cards to obtain responses. The polls conducted by *The Literary Digest* were huge. The mail-outs for polls conducted between 1924 and 1936 ranged upward from 10 million; the 1932 Prohibition poll had a mail-out of over 20 million ballots (Lusinchi, 2015). The return rates for these polls ranged from 11.8% to 23.8%; the largest number of ballots returned was 4,806,537 for the 1930 Prohibition poll (Lusinchi, 2015). *The Literary Digest* polls had correctly predicted the winner of the Presidential race from 1920 to 1932.

The Literary Digest reported that its poll predicting the result of the 1936 Presidential election would be the even larger than previous polls. It was claimed that: "The Poll represents the most extensive straw ballot in the field—the most experienced in view of its twenty-five years of perfecting—the most unbiased in view of its prestige—a Poll that has always previously been correct" (*The Literary Digest*, 31 October 1936).

THE LITERARY DIGEST'S PREDICTION

The Literary Digest commenced its polling in August, and published their election prediction on 31 October 1936, providing the following detail with the headline "Final Returns in The Digest's Poll of Ten Million Voters": Landon received 1,293,669 votes and Roosevelt 972,897 votes". This gave Landon 57.1% of the votes for these two candidates, and Roosevelt 42.9%. *The Literary Digest* poll also reported 83,610 votes for William Lemke, and the total number of ballots returned was 2,376,523. Of course, the *Digest's* headline is misleading: ten million ballots were sent out but around 2.4 million were returned. The results were described as "exactly as received

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

from more than one in every five voters polled in our country … neither weighted, adjusted, nor interpreted" (*The Literary Digest*, 31 October 1936).

GALLUP'S CRITIQUE AND PREDICTION OF *THE LITERARY DIGEST* RESULT

Gallup was sceptical about the "straw" polling methods used by *The Literary Digest.* On 12 July 1936 in his column *American Speaks*, Gallup also made a prediction about *The Literary Digest*:

"*The Literary Digest* has announced its poll this year will be larger than ever before. In the past this magazine has sent ballots to as many as 20 million persons.

If *The Literary Digest* were conducting its poll at the present time, following its usual procedure, Gov. Landon would be shown in the lead. The actual figure would be in the neighbourhood of 44 per cent for Roosevelt, 56 per cent for Gov. Landon.

Since the Institute of Public Opinion sends part of its ballots to the same lists covered by *The Literary Digest*, it is possible to predict with a high degree of accuracy the sentiment which *The Digest* will find. The lists comprise telephone subscribers, automobile owners, and registered voters. The upper economic levels are represented to a much grater (sic) extent than the lower levels in such lists, and particularly in the returned ballots. People at the upper levels are more inclined to answer ballots than people at the lower end of the scale. When the lower one-third of the voting population is fully represented, Roosevelt's percentage changes from 44 per cent to 52 per cent." (Pittsburgh Press, 12 July 1936) Gallup's prediction was made weeks before *The Literary Digest* started its poll.

GALLUP'S PREDICTION OF THE ELECTION RESULT

Gallup's final pre-election poll (AIPO survey #55B, interview dates 22nd to 28th October) was reported on 1 November 1936, and gave President Roosevelt 55.7% of the major party vote (minor parties eliminated) to 44.3% for Governor Landon (*America Speaks*, 1 November 1936, Pittsburgh Press).

On the same date, Gallup's prediction of the result of *The Literary Digest* poll and the final *The Literary Digest* results were compared:

"A direct proof that number of ballots is not necessarily the most important item in a poll was established when the Institute on July 12 – one month before *The Digest* began sending out ballots – predicted on the basis of a few thousand returns what *The Digest* would find as a result of its mailing of 10,000,000 ballots.

The prediction was that Landon would be shown winning by approximately 56 per cent of the major party vote. The final report of *The Digest* gives Landon 57 per cent." (Pittsburgh Press, 1 November 1936)

Gallup predicted Roosevelt would win, and additionally he had predicted that *The Literary Digest* would predict that Landon would win. The headlines of the AIPO news articles claimed that the "Vote Will Test Poll Methods" and the "Election to Decide Value of Scientific Sampling vs Mass-Balloting" (Pittsburgh Press, 1 November 1936).

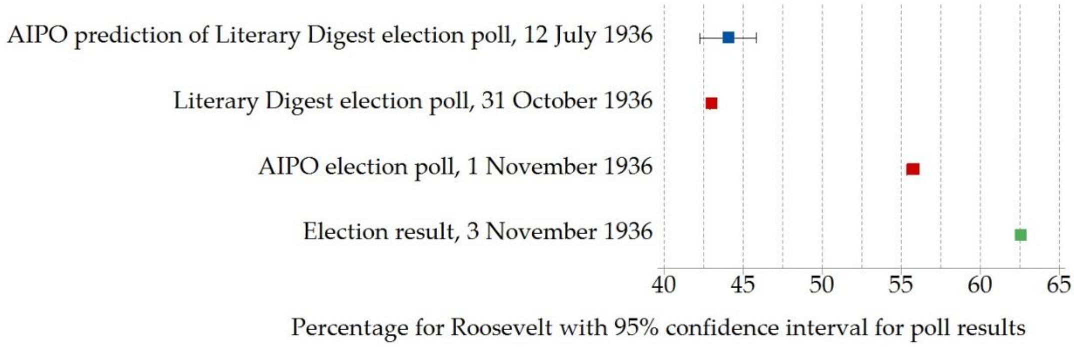
THE 1936 PRESIDENTIAL ELECTION RESULT

The result of the 3 November 1936 election gave Roosevelt 27,752,648 votes and Landon 16,681,862 votes; this gave Roosevelt 62.5% of these two party share results. In total 45,647,699 people voted.

Figure 1 shows the final election result in green. The AIPO and *The Literary Digest's* predictions are shown in red. Gallup's prediction of the result of *The Literary Digest* poll is shown in blue. Figure 1 shows approximate 95% confidence intervals calculated using the assumption of simple random sampling and the reported sample sizes for the various polls, discussed further below. The sample size for *The Literary Digest* prediction was so large that the confidence interval bars are not able to be detected on Figure 1; the notional margin of error was 0.06%.

Gallup's prediction of the election result was out by 6.8%. *The Literary Digest*'s prediction was out by over 19%. Figure 1 illustrates this clearly. It shows the consistency of Gallup's prediction of *The Literary Digest* poll with its result.

Gallup's poll was seen as a "success" as it predicted victory for Roosevelt; *The Literary Digest* did not. *The Literary Digest* poll was a "fiasco" (Feinberg and Tanur, 2001) and it is true that *The Literary Digest* went bankrupt in 1937.



Assumed sample sizes: AIPO prediction of Literary Digest: 3,000; AIPO election poll: 50,000.

Figure 1: Predictions of the 1936 US Presidential election result with approximate 95% confidence intervals

THE LITERARY DIGEST POLL IN STATISTICAL EDUCATION

The failure of *The Literary Digest* poll makes a compelling and engaging story in statistical education, and so the story of the polls and the 1936 US Presidential election usually focusses on *The Literary Digest* poll. Here we recount the lessons which can be taught.

Large sample sizes do not guarantee accurate results

The Literary Digest's extraordinary sample of nearly 2.4 million voters failed to produce an accurate prediction of the election outcome. Clearly a large sample will not guarantee an accurate result, as Gallop repeatedly argued. The magnitude of *The Literary Digest's* sample size *and* error in prediction provide a striking counter-example to challenge the misconception that a large sample is sufficient for statistically valid results. This lesson is important to all learners of statistics, from those aiming to acquire the fundamentals of statistical literacy to those learning about complex data analysis that cannot be divorced from the context in which it is collected.
Possible explanations for *The Literary Digest's* error provide further important lessons.

Beware of biased sampling frames

Gallop's critique of the sampling frame used by *The Literary Digest (*e.g. *America Speaks*, Pittsburg Press, 12 July 1936), suggests that it was biased. He described the differences between the AIPO sampling frame and that of *The Literary Digest*: "The [AIPO] ballots are distributed by mail and by a staff of more than 280 personal interviewers. The interviewers are used to reach the lower one-third of the voters, whose names do not appear on ordinary commercial lists and who do not answer mail ballots in sufficiently large numbers. Mail ballots reach the upper levels through the use of names of residence telephone owners and owners of automobiles." Gallop argued that the sampling frame used by *The Literary Digest* did not include the "lower one-third of voters", meaning people at the lower end of the income scale.

Lusinchi (2012) describes the common critique that *The Literary Digest* relied on a biased sampling frame based on lists including telephone subscribers, automobile owners and registered voters as the "conventional explanation" (also Erikson and Tedin, 1981). This explanation for the failure of *The Literary Digest* poll is popular, and has its roots in Gallup's critique. Lusinchi (2012) contends however, that it is not the primary explanation.

Beware of poor response rates, and non-response bias

A second concern is *The Literary Digest's* response rate of 24%: only 2.4 million of 10 million ballots were returned. When there is a poor response rate, there is a potential for non-response bias where the people who respond to the survey are different, on average, from those who do not. The voters responding to *The Literary Digest*'s invitation to participate, in particular, tended to be Republican (Landon) voters; those who chose not to participate tended to be Democrat (Roosevelt) voters.

There is still debate about the extent to which the spectacular failure of *The Literary Digest* poll was due to non-response bias or a biased sampling frame. Here we do not focus on these arguments. The interested statistical educator can find further useful lessons in examining these arguments in greater depth; see, for example, Lusinchi (2012).

THINKING CRITICALLY ABOUT THE GALLUP POLL

Concerns about *The Literary Digest* poll relate to overreliance on sample size as a guarantee of validity, possible problems with the sampling frame and response bias. How did Gallup's poll compare on these issues?

Why the emphasis on the number of ballots distributed?

Like *The Literary Digest*, Gallup's reporting tended to focus on the number of ballots distributed, rather than the number of responses received. For example, in *American Speaks* on 27 October 1935, Gallup wrote "the American Institute of Public Opinion in order to be absolutely sure of the accuracy of its results, normally distributes 100,000 to 200,000 ballots by mail and by personal interviewers on each issue" (e.g. Pittsburgh Press). In general, the actual number of ballots distributed for the polls during the early years of the AIPO is not easy to find. However, Gallup did report the number of ballots distributed for the final AIPO 1936 pre-election poll: "The number of ballots distributed in the poll was 312,551. They went by mail and by personal interviewers to a scientifically selected cross-section of voters in all states." (The Salt Lake Tribune, 1 November 1936, page 72.)

Gallup (1972b, p.66) also reported on the size of the mail-out for the survey that predicted *The Literary Digest* result: "A sample of only 3,000 post card ballots had been mailed by my office to the same lists of persons who received *the Literary Digest* ballot".

The final sample size in *The Literary Digest* poll was around a quarter of the number of ballots distributed. The focus on the large number of ballots distributed, by both Gallup and *The Literary Digest*, appears to be a mechanism for *asserting* statistical robustness, even though Gallup argued that large samples do not guarantee success.

What were the sample sizes in the 1936 pre-election Gallup polls?

The numbers of responses obtained by return ballot and in interviews for the early Gallup polls are also hard to find. Here the sample size is the number of responses obtained, not the number of ballots distributed. Sample sizes were generally not reported with the poll results (e.g. in *America Speaks*) or in publications of the findings (e.g. Gallup, 1972a).

According to Robinson (1937) the AIPO forecast of the 1936 election was based on a sample size of 125,000. However, Freedman, Pisani and Purves (1978) report the sample size for the final pre-election poll (survey #55B) was around 50,000; presumably this is the number of responses obtained given that the number of mail-outs and interviews for the final election poll was over 300,000.

In relation to the survey predicting *The Literary Digest* result, Freedman, Pisani and Purves (1978) report that "[Gallup] just chose 3,000 people at random from the same lists the Digest was going to use, and mailed them all a postcard **…"** (page 304). Hence for this survey, 3,000 refers to the number of mail-outs; we have not been able to find the sample size obtained for the prediction of *The Literary Digest* result. It may have been far fewer than 3,000.

How far out was Gallup in predicting the 1936 election result, and did it matter?

Although the Gallup poll was regarded as successful in terms of predicting the election winner, the predicted two-party percentage for Roosevelt was out by nearly 7%. Did the magnitude of this difference matter? In contrast with *The Literary Digest* result, Gallup's "error" may have seemed small. However, considering that the margin of error for the result based on a sample size of 50,000 would be less than 0.5%, the Gallup poll result was quite biased. Why was this?

Gallup admitted some flaws in the first post-election edition of *America Speaks* (Pittsburgh Press, 8 November 1936). He argued that *temporal* factors contributed to the inaccuracy in the predicted percentage for Roosevelt. This kind of argument serves to minimize scrutiny of design flaws in the AIPO poll. While it true that AIPO polls shows increasing support of Roosevelt since July 1936, they also showed that his support had waxed and waned since 1934 (e.g. Pittsburgh Press, 1 November 1936). The error of the Gallup poll was not trivial; could it be fully explained by an unmeasured trend?

What was the response rate in the 1936 pre-election Gallup poll?

Like the sample sizes, response rates were generally not reported with the poll results (in, for example, *America Speaks*) or in publications of the findings (e.g. Gallup, 1972a). Without exact information about the number of responses obtained in the final 1936 AIPO pre-election poll, we cannot determine the response rate.

Given the concern about the response rate for *The Literary Digest* poll; what was it for the Gallup poll? If Freedman, Pisani and Purves' (1978) sample size information is correct, the response rate is about 16%. Robinson's (1937) sample size figure would put the response rate at about 40%.

There is some published information about response rates for AIPO surveys around the time of the 1936 election. Benson (1937) gives an overall estimate for 1936 mail ballot returns of AIPO surveys for the total US of 17.3%. So it is quite possible that the response rate for the AIPO's final pre-election poll might have been worse than that of *The Literary Digest*. This is rarely discussed, but leaves the Gallup poll potentially open to problems of response bias. Figure 2 provides a comparison of these response rates.

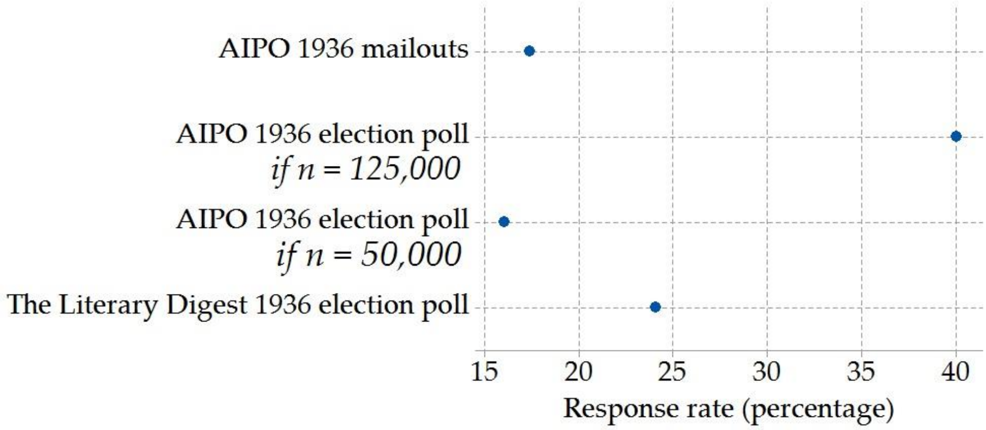


Figure 2: Response rates for various 1936 US polls

Was there non-response bias in the Gallup poll?

Lusinchi (2012) suggests that non-response bias was an important part of the failure of *The Literary Digest* poll: Roosevelt voters were less inclined to respond to the poll than Landon voters. It is plausible that the same kind of non-response bias would have occurred in the AIPO poll; there is no *prima facie* reason why there would not be. However Gallup used a quota system to attempt to ensure that his polls included appropriate proportions of voters of different types.

What kind of characteristics did the AIPO polls consider in setting quotas? Ohmer (2006) describes this in some detail, based on information published in *The Washington Post* in 1935. Quotas were set by considering geography, age, income, race, gender, political affiliation and voting preference in the 1932 Presidential election. This approach attempted to ensure that response bias did not result is disproportionate representation of, for example, Democrat versus Republican affiliates. If Democrat affiliates were less likely to respond, Gallup's pollsters may have had to work harder to fill the Democrat quotas compared with the Republican ones.

Non-response bias arises when there are differences on average between those who do and those who do not respond to a poll. Hence the possibility of non-response bias remains for the Gallup poll, even though quotas were used. Given the magnitude of the discrepancy between the AIPO final pre-election poll and the final election result, response bias should be considered.

Was there sampling bias in the Gallup polls?

The AIPO's method of obtaining a sample involved working out what kinds of personal characteristics related to voting patterns and using these in the design of the sample. Quotas were set for the numbers of individuals needed for each type of respondent, so that the number surveyed would reflect the population distribution of adult voters. This is not a random sampling method. In the first post-1936-election edition of *America Speaks* Gallup suggests the system of setting quotas was not without fault in relation to the quotas set for new and low income voters (Pittsburg Press, 8 November 1936). Indeed, the management of nationwide quotas must have been challenging.

Why was Gallup's prediction of The Literary Digest poll result so accurate?

The AIPO prediction of *The Literary Digest*'s prediction was more accurate than the AIPO prediction of the election result. Why did the AIPO survey result correspond so well to *The Literary Digest*'s result? The population that Gallup and his colleagues wanted to survey was the population of voters who were to be surveyed by *The Literary Digest*, and they understood the sampling frame for this population well. In predicting *The Literary Digest* result, Gallup apparently used *The Literary Digest* sampling frame and replicated their methods. Gallup's poll would have been subject to the same response biases as *The Literary Digest* poll, but for the purposes of predicting *The Literary Digest* poll result, this did not matter.

With 3000 responses, the margin of error of the Gallup poll estimate of *The Literary Digest* result of 44% of the vote for Roosevelt is less than 2%. The actual result for *The Literary Digest* was 43%, well within this margin of error. Of course, this assumes that there was a perfect response rate to the AIPO survey predicting *The Literary Digest* result; it is likely that the margin of error was somewhat greater.

CONCLUSIONS

Compelling stories such as the prediction of the 1936 US Presidential election result can provide engaging and memorable lessons in teaching statistical literacy. Here is a rich source of material, and the apparent success of the Gallup poll should not mean that it goes without scrutiny. The context is unusual for students: estimates can be compared with the true parameter of interest. The inaccuracy of *The Literary Digest* and Gallup's predictions, relative to the respective margins of error should lead to enquiry about sample size, sampling bias, response rates and response bias for *both* polls.

REFERENCES

Benson, L. E. (1946). Mail surveys can be valuable. *Public Opinion Quarterly*, *10*(2), 234-241.

Erikson, R. S., & Tedin, K. L. (1981). The 1928–1936 partisan realignment: The case for the conversion hypothesis. *American Political Science Review*, *75*(04), 951-962.

Fienberg, S. E., & Tanur, J. M. (2001). History of Sample Surveys. *International Encyclopedia of Social and Behavioral Sciences*.

Freedman, D., Pisani, R., & Purves, R. (1978). *Statistics*. Norton.

Lusinchi, D. (2012). "President" Landon and the 1936 Literary Digest Poll. *Social Science History*, *36*(01), 23-54.

Gallup, G.H. (1972a) *The Gallop Poll: Public Opinion, 1935–1971*. New York: Random House.

Gallup, G. H. (1972b). *The sophisticated poll watcher's guide*. Princeton Opinion Press.

Ohmer, S. (2006). *George Gallup in Hollywood*. Columbia University Press.

Rich, W. (1939). The Human Yardstick. *Saturday Evening Post*, *21*.

Robinson, C. E. (1937). Recent developments in the straw-poll field. *Public Opinion Quarterly*, *1*(3), 45-56.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

Let's be informed users of simulation to facilitate learning sampling distribution of the mean

LIU, Yulin
Queensland University of Technology
y68.liu@qut.edu.au

Sampling distribution of the mean is pivotal in an introductory university statistics course. Students find the concepts of sampling distribution of the mean and Central Limit Theorem difficult to understand. Using simulation to teach those concepts has been widely recommended to promote conceptual understanding, though the issue of its effectiveness is far from settled. More importantly, there is recent evidence suggesting that careless use of simulated sampling distribution of the mean could potentially foster misunderstandings. This article supports the opinion that using simulation is an effective way to teach about sampling distributions as long as it is used in a pedagogically sound manner. Based on a review of the research on both the effectiveness and warnings of sampling distribution simulations, this study proposes a four-step agenda to use simulation to facilitate learning sampling distribution of the mean, i.e., a good summary of the theory, explanation of the non-limiting properties, advice about proper interpretation of simulation results, and verifying Central Limit Theorem with a structured simulation experiment. Preliminary effectiveness of this agenda is observed in two classes of business and engineering students. This study contributes to the scholarship of statistics education and informs the practice of adopting simulations

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

Designing Massive Open Online Courses for Educators around the World: The Case of Teaching Statistics

TRAN, Dung[1] and LEE, Hollylynne S.[2]
[1]Victoria University, Australia, [2]NC State University, USA
Dung.tran1@vu.edu.au

Statistics receives attention through global curriculum. Some have designed professional development for teachers to develop their statistical content and pedagogy, typically on a small local scale. Online courses can expand the number of teachers involved and create communities beyond school or district lines. For a "massive" and "open" course, there are many design challenges to meet the needs of participants with varied backgrounds in teaching statistics. We will share how a Massive Open Online Course for Educators (MOOC-Ed) designed in the USA, but offered and taken by educators from around the world, including Australia and New Zealand, presents design challenges for how to best impact teachers' learning and classroom practices. For this paper, we will focus on course design principles, the design and implementation of a framework for supporting students' approaches to statistical investigations and a task guide to assist teachers in evaluating, designing, and implementing worthwhile statistical tasks. We will share participants' engagement with the course and evidence we have collected suggesting impacts on their classroom practice.

Statistics receives attention through global curriculum (ACARA, 2012; CCSSM, 2010). Some have designed professional development for teachers to develop their statistical content and pedagogy, typically on a small local scale (cf. Darling-Hammond et al., 2009). Online courses can expand the number of teachers involved and create communities beyond school or district lines (Kim, 2015). Indeed, with advances in technology and interest in offering alternatives to traditional professional development, the number of online professional development opportunities has increased. The USA National Research Council (2007) claimed that: "Growing numbers of educators contend that online teacher professional development (OTPD) has the potential to enhance and even transform teachers' effectiveness in their classrooms and over the course of their careers." (p. 2)

MOOCs are designed and delivered to serve different target populations and provide diverse experiences for learners (Clark, 2013). For a "massive" and "open" course, there are many design challenges to meet the needs of participants with varied backgrounds in teaching statistics and working in different educational contexts. Capturing the potential for MOOCs to serve as large-scale professional development, teams have created MOOCs for Educators (MOOC-Eds) in the USA. As leaders of one of these teams, we designed a course focusing on teaching statistics to assist mathematics and statistics teachers in developing content understanding and pedagogical strategies for improving practice, and forming local and global communities of educators. This course was offered to and taken by educators from around the world, including those from Australia and New Zealand, presenting design challenges for how to best impact teachers' learning and classroom practices worldwide. Our question guiding this design and research is "*To what extent does a MOOC in teaching statistics offer opportunities for mathematics and statistics teachers to engage in professional learning and impact their teaching statistics practices*?" For this paper, we will focus on course design principles, the design and implementation of a framework for supporting students' approaches to statistical investigations and a task guide to assist teachers in evaluating, designing, and implementing worthwhile statistical tasks. We will share participants' engagement with the course and evidence we have collected suggesting impacts on their classroom practice.

Course Design Principles

MOOC-Eds are specifically designed to help educators meet their professional learning needs, so it is assumed that participants are motivated and self-directed. This MOOC-Ed was built using design principles of effective online learning and professional development (Kleiman, Wolf, & Frye, 2014) that emphasize: (a) self-directed learning, (b) peer-supported learning, (c) job-connected learning, and (d) learning from multiple voices. Most typical MOOC participants review material individually and some engage in discussion forums (Kim, 2015). We will highlight two of the design principles that

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

address this challenge that makes MOOC-Eds different. For *peer-supported learning*, this MOOC-Ed made extensive use of discussion forums for encouraging participants to reflect, exchange ideas and resources, and engage in dialogue and debate to extend their understanding. We value the experience and expertise of the participants and design learning activities such that educators can share their knowledge to further the learning of others. There is also an emphasis on establishing professional connections among MOOC-Ed participants, who are identified by name in all their comments and projects; participants are not able to post or give feedback anonymously.

The course incorporated a number of opportunities to *learn from multiple voices*. MOOC-Eds are purposefully not designed around one or two experts who present online lectures; instead, they offer a rich set of perspectives presented within the context of activities and exchanges. As members of the design teams, we created our own resources and used existing open access resources written by other educators. Discussions that included well-known experts in the discipline were recorded and used throughout the courses. In these videos, the experts discuss relevant topics, share personal experiences and valued resources, and suggest strategies for implementing knowledge gained from research in everyday classrooms. Student voices were brought into the course through videos of teachers and students engaged in tasks in classrooms, and through animated videos based on actual student responses to research tasks. Multiple voices allow participants to learn about the perspectives of other teachers and administrators and those of students, researchers, and experts in the field

The discussion forums were designed for participants to post their thoughts about resources (readings, videos, tools) and discussion prompts, and interact with others. The design teams function as facilitators in forums; we encourage participants to share experiences and connect similar threads from different groups to offer multiple perspectives purposed to support richer discussions.

Statistical Investigation Learning Opportunities

The purpose of the course was for participants to think about statistics teaching in ways likely different from current practices in middle school through introductory statistics. A major goal was for teachers to view statistics as an investigative process (pose, collect, analyse, interpret) that incorporates statistical habits of mind and view learning statistics from a developmental perspective, aligned with guidelines from Franklin et al. (2007). We highlight our effort to design and implement a framework to support students' approaches to statistical investigation and a task guide to assist teachers in analysing, designing, adapting, and implementing worthwhile statistical tasks.

Design and implementation for a framework for supporting students' approaches to statistical investigations. We built upon the GAISE K-12 (Franklin et al., 2007) report by incorporating recent research on students' and teachers' statistical thinking and highlighting productive statistical habits of mind. The GAISE document is extensive (88 pages) with many illustrated examples. Our design challenge was to create useable artefacts and experiences that could communicate the essence and important messages within relatively brief comprehensible formats. Thus, we developed an adapted version of the GAISE framework we called supporting Students' Approaches to Statistical Investigations (SASI), and then designed learning materials to communicate the framework. Two brief PDF documents included statistical habits of mind and explicitly described the framework. We designed a graphic to communicate the investigative cycle, reasoning in each phase at each of 3 levels, and an indication of productive habits of mind for each phase. An interactive diagram was created so that details appear when an aspect is clicked (Figure 1). In a video, the instructor illustrated the framework using student work from research as supportive examples, and another video featured one of the experts illustrating the development of the concept of mean across levels of sophistication. The participants then discuss, in the forums, a task that allows for students to approach their work at varying levels of sophistication, video examples of students' work on the task, and how they could use such a task in their own practice. Participants were specifically asked to discuss how students' reasoning aligned with the SASI framework. Participants would again experience each of the phases of the investigative process when dealing with massive real data from the international CensusatSchool project. They experienced the statistical process as a learner and as a teacher who will teach the process to students. We explicitly highlighted the habits of minds in doing statistics, built upon previous general habits of mind in doing mathematics (e.g., Cuoco, Golden, & Mark, 1996) and statistical thinking frameworks (e.g., Wild & Pfannkuch, 1999). Educators were expected to discuss the habits of mind as they watched students engage in a statistical investigation with CensusatSchool data in the animated videos.

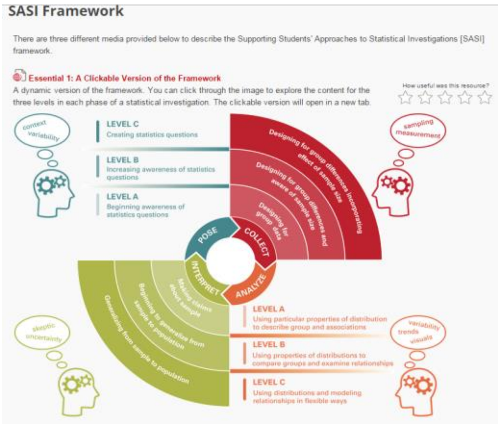


Figure 1. Learning resources for the SASI framework.

Task guide in evaluating, designing, and implementing worthwhile statistical tasks. Researchers have emphasized the roles of instructional tasks as catalysts for student learning (Stein & Lane, 1996). Therefore, we intentionally created learning opportunities around task analysis, adaption, and design. Two emphasised features included the difference between mathematics and statistics, and nature of worthwhile statistical tasks. Statistics educators (Franklin et al., 2007; Gattuso & Ottaviani, 2011) posit that uncertainty, the omnipresence of variability, and importance of context collectively distinguish statistics from the deterministic nature of mathematics. It is crucial for educators to realise the essence of statistical reasoning so that they could support students developing such thinking, which was underscored in the brief summary *Difference between Statistics and Mathematics* for participants to read. In this reading, we highlighted the role of context, measurement, variability and uncertainty to serve as a lens for educators to examine tasks that offer potentials for students to engage in statistical reasoning. We also developed a guideline for educators to examine why tasks are worthwhile in developing statistical thinking, how to improve the tasks, and what considerations need to be taken into account when implementing tasks in the *Considerations for Design and Implementation of Statistics Tasks* document. Participants then analysed tasks by mapping components of the tasks to the task guidelines.

Data Collection and Analysis

Data included registration and click logs of actions taken by participants (e.g., resources viewed, videos watched, forum viewed and posted). All dialogs generated in discussion forums, 5-star ratings of resources, and feedback surveys were collected. Descriptive statistics were generated based on demographic information, survey responses, and click logs. Open coding of forums and survey responses was used to develop themes related to impacts on practice, especially participants' understanding of statistical investigation and the use of an analytic frame to look at statistical tasks.

Engagement in the MOOCs

We will focus on the course offerings that occurred in Spring 2015. The course had 797 participants registered from 43 countries, with 597 (76%) registrants from the USA (see Figure 2 for the global reach for courses). Classroom teachers (64%) constituted the majority of participants, followed by about 10% of participants that worked in mathematics teacher education in university settings or other professional development roles. Interestingly, about two thirds of participants in the course held advanced degrees (masters or doctoral), which indicates that engaged learners in their discipline, valuing advanced educational opportunities were attracted to the MOOC-Eds.

Figure 2: Global enrollment in course in Spring 2015.

Participants were categorised into "no shows" i.e. entered the course after registration, "visitor" who logged into the course and engaged with some aspect of it four or fewer times, and "active participants" – the remaining participants (see Kleiman, Kellogg, & Booth, 2015 for more details). There is a large proportion of "no shows" participants (more than 1/3), and a large number (528) engaged somewhat and more fully. These numbers may not look impressive for a massively scaled course. However, in the context of professional development, a large number of participants are potentially taking the advantage of the learning opportunity offered in the course. Further analysis (see Kleiman et al., 2015) characterized the active participants according to how engaged they were throughout the course resources including videos, readings, tools, visiting, posting, and commenting on the discussion forums. This analysis showed that the 180 active participants had either declining activity (54%), or sustained high activity (46%). These high activity rates, through the final units in the courses, is much higher than typical MOOC completion rates (2-10%), but is aligned with completion rates when participants intend to complete a course (Reich, 2014).

As highlighted above about the peer-support learning in the MOOC-Eds, the course mainly accomplished through opportunities to interact with one another in discussion forums. There were about 33% of visitors and active participants who posted at least twice in the forums with either a new post or comment on a post of a peer. 308 (58.3% of the 528 visitors or active participants) participated in the forums with an average of 7.1 posts each with 930 discussions. There were also many more discussion views than postings, which were done by both active posters and non-posters. This suggests many saw discussion forums as an opportunity for learning, even by merely reading the posts of others. These "lurking" participants are present, but not visible; thus, exactly why they read discussions and what they have learned from them is unknown.

Impact of the MOOC-Eds on Educators' Practices

When asked, at the end of each unit, the extent to which activities supported the application of course content to their professional practice, 98% of participants agreed to some extent (i.e., selected "somewhat agree," "agree," or "strongly agree). As a follow-up question, 100% educators reported they had made changes in their practice as a result of the course. The ways they were applying their learning experience in the course to their practices include: (a) integrating new tools and strategies, (b) implementing course projects, and (c) using course content for instructional coaching and professional development, as indicated in the open-ended survey items.

Participants report positively on triggers for changing teaching statistics practices that were purposefully designed in the course. Participants (n=48 respondents) reported the course (a) strengthened their understanding of how to engage students in a statistical investigation process (97.8%), (b) improved their ability to use the framework to guide their teaching of statistical investigations to promote deeper data explorations for their students (93.7%), (c) improved their ability to use rich data sources to support investigations (95.8%), and (d) improved the ability to use dynamic tools to visualize and analyse data (91.6%). These four triggers also appeared as we examined the forum posts and open responses on surveys. For example, their open ended survey showcased an "overhaul" in their approach to the practice in relation to the use of real data in teaching statistics:

> Since starting the class, I have had my students use richer and messier data in their investigations and I have also put more of an emphasis on understanding the results and being able to analyze findings.

The MOOC-Ed participants appreciated being introduced to a more structured, comprehensive approach to teaching statistics. As a result of their engagement in the course, some participants described a shift in their overall approach to teaching statistics. In commenting on the value of the Supporting Students' Approaches to Statistical Investigations (SASI) Framework, one participant noted:

> The SASI framework was the most useful part of the course. It is incredible. I've been telling the teachers here about it because normally we teach the Intro to Stats class only procedurally, just calculations, with no sense of equations or interpreting. But that has changed now because of using the framework.

When analysing tasks as to whether they are worthwhile for teaching statistics or viewed examples of students' work on tasks, participants recognized four phases of the statistical investigation displayed in students' engagement in tasks or as they engaged in analysing statistical tasks:

> These activities were full of various levels from the SASI framework. Students were posing relevant questions checking the simulations to see what changed and then posing more questions. I love how they spent so little time on the actual calculations and so much time posing and seeking answers to relevant questions. Students were engaged and not frustrated with tedious calculations.

Participants discussed more on the first two phases of the investigation:

> I was more interested in the Television Time data because that data could be collected in far less class time through technology shared by multiple classrooms (Google Docs comes to mind). The data sets contained the actual data and allowed for exploration of variability, comparison of different ways to graph the same data, disaggregating the data, mean vs. median, and several others. Maybe I'm only valuing the variety of available analysis tools rather than a direct opportunity to *physically* collect data. I still think there's data collection concerns that can be addressed even with this activity: units of reporting, discrete vs. continuous, response bias, etc.

They also saw the potential of tasks in addressing different phases of statistical investigations and developing statistical habits of mind.

> I think there is a lot of merit to different statistical tasks that focus on only portions of the cycle. I think it's ok to maybe have a day to brainstorm how to just pose a question. Then another task the just focuses on how to gather data. The other two tasks were mostly analysis, but there is certainly merit in having some tasks that only analyze data as well. Of course we should have some tasks that integrate all 4 parts of the cycle, but I would argue that we cannot make every task that way. So if the goal is to teach the analysis of the data, the other tasks are just fine. If the goal is to engage in more of the cycle, then Coke vs Pepsi does that, but it still lacks the whole cycle. I like that this task uses real world data rather than made up numbers.

Educators also reported wanting to use all four phases of a statistical investigation, rather than their past heavy emphasis on the analysis phase.

> I have changed my planning process for statistics. I will use more technology in my teaching and spend more time on the first 2 phases of the investigative cycle. I will encourage statistical habits of mind and movement through the levels of the SASI framework.

While some of the comments indicate how teachers have already changed, or will change, their practices with their own students, other comments show how elements of the course are impacting how participants encourage their colleagues to change their practices.

Discussion

While the number of participants does not look massive like other MOOCs, in the context of professional development, at least 528 participants take advantages of our purposeful designs in Spring 2015 (the course has been offered 3 additional times with many more participants and a growing global community). The design includes general principles in development of MOOC-Ed, as well as specific productive practices of teaching statistics. These serve as catalysts for self-reflection and change in practice. The research-informed practices in teaching statistics were highly valued and appeared to assist participants in viewing the learning and teaching of statistics more conceptually and comprehensively that focuses on developing statistical thinking and habits of mind. Participants also seemed to be able to shift their perspectives from viewing the importance of teaching and learning statistics as reliant on algorithms or procedures, to a view of statistics as more of a process that has nuanced conceptions that must be developed with extended experiences.

Results show that our MOOC-Ed provided participants the opportunity to engage in professional development to strengthen their content and pedagogy in areas *they personally were interested in improving*. We continue to learn about the affordances and constraints of this model of professional learning for mathematics teachers and are interested examining the long-term impacts on practice. How could such MOOC-Eds sustain the trigger to continuing practices of educators in their jobs? How would participants engage, and would professional learning networks emerge? Furthermore, we would like to explore the possibility of international collaboration in the design of future courses that closely address more specific demands (such as the Australasia education system) and the possibility of offering smaller scale modules that are continuously available.

Acknowledgement
The design, implementation, evaluation, and research of MOOC-Ed courses at North Carolina State University, USA is partially funded by the William and Flora Hewlett Foundation. Any opinions, findings, and recommendations expressed are those of the authors, and do not necessarily reflect the views of the Hewlett Foundation.

References

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2015). *The Australian curriculum: Mathematics, version 8.2*. Sydney, NSW: Author.

Clark, D. (2013, April 16). MOOCs: Taxonomy of 8 types of MOOC [Web blog post]. Retrieved from http://donaldclarkplanb.blogspot.co.uk/2013/04/moocs-taxonomy-of-8-types-of-mooc.html

Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Cuoco, A., Goldenberg, P. E., & Mark, J. (1996). Habits of mind: An organizing principle for mathematics curricula. *The Journal of Mathematical Behavior, 15*(4), 375-402.

Darling-Hammond, L., Wei, R., Andree, A., Richardson, N., & Orphanos, S. (2009). Professional learning in the learning profession: A status report on teacher development in the United States and abroad. Dallas, TX: National Staff Development Council.

Franklin, C., et al. (2007). Guidelines for assessment and instruction in statistics education (GAISE) Report: A Pre-K-12 curriculum framework. Alexandria, VA: American Statistical Association. http://www.amstat.org/education/gaise/GAISEPreK-12_Full.pdf.

Gattuso, L., & Ottaviani, M. G. (2011). Complementing mathematical thinking and statistical thinking in school mathematics. In C. Batanero, Burrill, G., & Reading, C. (Eds) *Teaching statistics in school mathematics-Challenges for teaching and teacher education* (pp. 121-132). NY: Springer.

Kim, P. (Ed.). (2015). Massive open online courses: The MOOC revolution. New York, NY: Routledge.

Kleiman, G., Kellogg, S., & Booth, S. (2015). MOOC-Ed Evaluation: Final report submitted to the William and Flora Hewlett Foundation. Raleigh, NC: Friday Institute of Educational Innovation.

Kleiman, G.M., Wolf, M.A. & Frye, D. (2014). Educating educators: Designing MOOCs for professional learning. P. Kim (Ed.). Massive Open Online Courses: The MOOC Revolution. Routledge.

National Research Council. (2007). Enhancing professional development for teachers: Potential uses of information technology, report of a workshop. Washington, DC: The National Academies Press. Retrieved from http://www.nap.edu/catalog/11995/enhancing-professional-development-for-teachers-potential-uses-of-information-technology

Reich, J. (2014). MOOC completion and retention in the context of student intent. EDUCAUSE review. http://er.educause.edu/articles/2014/12/mooc-completion-and-retention-in-the-context-of-student-intent.

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation, 2*(1), 50-80. doi:10.1080/1380361960020103

Wild, C. J., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review / Revue Internationale de Statistique, 67*(3), 223-248. doi:10.2307/1403699

Blended Learning in Postgraduate Applied Statistics Programs

BHOWMIK, Jahar, MEYER, Denny and PHILLIPS, Brian
Swinburne University of Technology,
Melbourne, AUSTRALIA
jbhowmik@swin.edu.au

The term blended learning refers to an approach to curriculum development where some form of an online learning environment supports and enhances the traditional on-campus or face-to-face experience in an integrated manner (Oliver & Trigwell, 2005). Postgraduate applied statistics programs at Swinburne University of Technology adopted the blended learning almost a decade ago. This allows for flexibility in design approaches, and accommodates the range of blended learning capabilities and experience of teachers and learners. Blended learning design adopted in these programs has involved the thoughtful integration of learning and teaching approaches in both on-campus, face-to-face and online/virtual learning environments by utilising the benefits of each of these environments to enhance the student learning experience. These programs focus on designing learning interactions across formal teaching spaces, informal learning spaces and online learning and teaching spaces. This flexible approach has been well accepted among both online and on-campus students. This paper describes the blended structure adopted in the applied statistics programs at Swinburne and the feedback received from students during recent study periods.

Key Words: curriculum, flexibility, virtual, learning environment and feedback.

INTRODUCTION

Blended learning is a term increasingly used in higher education to describe the way e-learning is being combined with traditional classroom methods and independent study to create a new, hybrid teaching methodology. The main purpose of blended learning is to make learning flexible and effective for the learner. A meta-analysis by Means, Toyama, Murphy, & Baki (2013) has reported that students perform significantly worse with face-to face only learning than with blended learning. Furthermore, student performance was found to be similar for face-to-face and purely online learning.

Statistics is always a challenge for teachers as well as learners because it involves many conceptual and mathematical concepts. It has been observed that students are more engaged with applied statistics subjects when a variety of activities are used in the different instructional methods (Biggs, 2003; Biggs & Tang, 2011; Kember & McNaught, 2007). Statistics teaching has benefited from the development of new technological resources. Several authors claim that teachers need to understand how to integrate the technology effectively within the blended learning structure to maximize its impact on student learning outcomes and support their learning (Park, 2009; Tishkoveskaya, & Lancaster, 2012; Ghahari, 2013).

From the literature review it can be seen that the term blended learning has been defined either in a broad or in a very specific way. However, all definitions of 'blended learning' have one common component - 'an integration of different instructional methods'. A thoughtful integration of different instructional methods (e.g. face-to-face and online components) needs to follow a suitable design approach at the planning stage of blending (Alammary, Sheard & Carbone, 2014). After examining different processes for designing blended learning courses, Alammary et al. (2014) classified three distinct design approaches: (1) Low impact blend: adding extra activities to an existing course, (2) Medium-impact blend: replacing activities in an existing course and (3) High-impact blend: building the blended course from scratch.

At Swinburne University of Technology, the postgraduate applied statistics nested programs (graduate certificate, graduate diploma and masters) started in 1989 with about 20 on-campus local part-time students. The main vision of the programs was to focus on the practical real life based application of statistical theory, statistical tools and techniques rather than concentrating too much on the theory. The program evolved during the 1990's with the addition of a coursework masters program and some full time students. By 2005, due to student demand, these programs were offered online, then in 2008 through Open University Australia (OUA). To satisfy the demands of different student cohorts (on-

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

campus and online), and the course learning objectives, the academic team employed a blended instructional approach for each of the units. The approach used in these programs can be classified as medium-impact blend, as defined by Alammary et al. (2014). Furthermore, according to Sharpe, Benfield, Roberts, & Francis (2006), iterative course redesign should consider student feedback as a critical success factor for course improvement with the medium-impact blending approach. Therefore, since 2005, at the end of every study period (semester), all units have been reviewed and updated based on student feedback and teaching panel members' experiences.

The remainder of the paper is structured as follows: First, the structure and objectives of the applied statistics program at Swinburne is briefly discussed. Next the different instructional methods used in blended instruction are explained along with their benefits. In addition, recent student feedback on the blended learning approach is presented. Finally, some concluding remarks are given.

POSTGRADUATE PROGRAMS IN APPLIED STATISTICS AT SWINBURNE

Many professionals use statistics for routine data collection, data mining analysis and interpretation in order to assist decision-making and ongoing work-related activities. Others, who rely on research articles and reports to stay ahead of developments in their industry, require an understanding of statistical methods to accurately interpret and comprehend reported results and relationships. The broad application of statistics demands that professionals, for example psychologists, market researchers, doctors, nurses, and scientists, have a sound knowledge of the statistical methods applicable to their discipline so that the decisions they make are well-informed. Statistical techniques are regularly under review and the technology available to carry out analysis is constantly developing. As a result, many professionals find the need for further training to keep up to date with the latest developments. To help facilitate this process, Swinburne University of Technology offers flexible postgraduate programs in applied statistics that focus on practical applications. These programs develop competencies in areas ranging from practical and basic statistical knowledge at the graduate certificate level, to the development of higher level statistical and research skills at the master level. Since its inception in 1989 the Swinburne Applied Statistics programs (initially as Social and Health statistics) have built an excellent reputation as a provider of quality statistics training. In 2015 there were a total of 207 enrolled students in these programs with 47 in Graduate Certificate, 30 in Graduate Diploma and 130 in Masters program where 97% of the enrolment was domestic students. About 50% of the enrolled students were female and a majority of them are matured age with a large number of students in their late 30's. More than half of the students work full time and are from a variety of academic backgrounds, including graduates from physical sciences, engineering, health sciences, economics, business and marketing.

The current structure of the post graduate program consists of 4 units (50 credit points) for the graduate certificate program, 8 units (100 credit points) for the graduate diploma program and 16 units (200 credit points) for the master's program, as shown in Figure 1. A major objective of the initiative taken in 2005 to adopt blended tuition was to provide students enrolled in the same courses, but in different modes, with exciting, innovative and flexible opportunities for engaging in learning, to achieve unit and course outcomes with career relevance while gaining life-long learning and development skills and having a positive university experience. For this purpose, initially the programs were reviewed and redesigned to implement learning interactions across formal teaching spaces, informal learning spaces and online learning and teaching spaces through a medium-impact blend approach. This approach has been appropriate for these programs and has been greatly helped by instructors with prior long-term face-to-face experience in teaching the traditional courses. Also, the instructors have excellent technological knowledge for online teaching and have had great support from the university. It is important to note that these changes have not been made overnight, rather they have occurred through an incremental replacement approach involving excellent institutional support which includes technical, technological training, educational designers and workload allocation.

A deeper look at the units' and programs' objectives were considered before selecting those educational technologies that would best meet the students' requirements. To reach a harmonious balance between online and face-to-face components for each of the units a number of changes were made during 2005-2008 along the lines recommended by Alammary et al. (2014). For most of the units on campus tutorials were replaced by virtual class room and other online activities.

The optimum balance has been found to vary at different course levels. In the higher level units, a greater use of face-to-face components than online components have been found to be appropriate. The opposite was found for the lower level units. Added activities such as audio and video clips and Camtasia recordings, were integrated into the graduate certificate level units in order to achieve more of a balance between on-campus and online learning spaces, as suggested by authors such as Chen & Looi, 2007; Kaleta and Skibba & Joosten, 2007. These extra activities have been adopted due to pedagogical need of the units and to fulfil students' demand. The breakdown of units in Swinburne's Applied Statistics programs is shown in Figure 1.

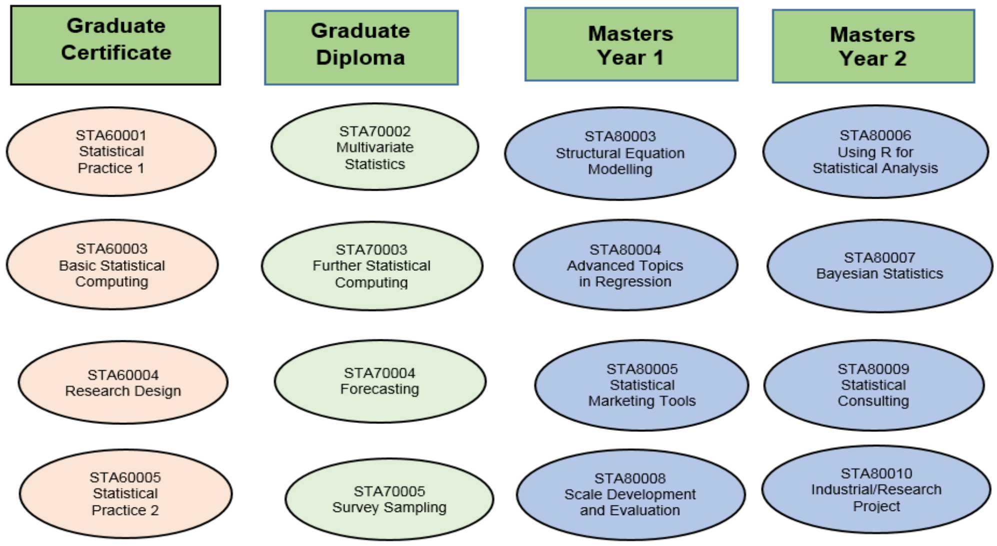


Figure 1: Postgraduate Applied Statistics Program Structure

The underlying mathematics is introduced gradually on a need-to-know basis. A variety of statistical software is used throughout the program; IBM SPSS Statistics software (SPSS) is used for the graduate certificate level units, SPSS and SAS are used for graduate diploma level units and various software including SAS, R, Mplus, Amos, RUMM2030, are used in the master's level units. In 2005, weekly on campus tutorial classes for the graduate certificate level units were replaced by Collaborate/Elluminate Live sessions, Camtasia recordings and Blackboard discussion board activities. For all other units, the on campus weekly classes were supplemented by one or more of Camtasia recordings, discussion board threads and short audio/video clips. During 2010-2015 optional on campus evening classes were replaced by weekend workshops for five of the eight graduate certificate and diploma level units.

MAIN ACTIVITIES (INSTRUCTIONAL METHODS) COVERED IN TEACGING AND LEARNING SPACES

In the selected medium-impact approach, existing units were redesigned by replacing some of the face-to-face activities with online activities. These activities were developed by the teaching staff most of whom had medium to long-term prior experience in teaching the traditional applied statistics courses, while others were mentored by experienced staff members.

The balance of different instructional/technological methods and pedagogies has been considered carefully during the design stage of the blending. To satisfy the demand for blended teaching, a mix of methods have been used in each of the units. Each of the instructional methods (activities) used in the applied statistics postgraduate programs are listed in Table 1.

Face-to-face instruction with Echo recording: On-campus face-to-face classes are offered for all postgraduate units at the Swinburne Hawthorn Campus in Melbourne for those students who can attend. These sessions are offered outside normal business hours. For the graduate certificate level units two day on-campus weekend workshops are run during the study period so that all interested students can join the sessions, even those living interstate. These sessions are recorded through the Echo System and uploaded on Blackboard so that all students can watch the recorded lectures. On-campus evening classes (3 hours weekly) are offered for graduate diploma and masters level units throughout the study period. To improve the level of industry engagement guest lecturers are used in some of the units including the masters level unit called "Statistical Consulting'. These sessions are also recorded through Lectopia/Echo360 and made available for students through Blackboard. Learning occurs in face-to-face sessions through the study period with online recordings of these sessions also available to all students.

Blackboard Collaborate class: **(Formerly Elluminate Live session)** This is a real-time, virtual classroom that gives instructors and students the opportunity to meet online to learn, rather than in a traditional classroom. For many of the units, live Collaborate sessions are offered through Blackboard so that learners can interact with their lecturer and fellow students. These virtual classrooms are very useful for students who cannot attend face-to-face sessions, either because they live remotely, or because tight schedules prevent them meeting at the same time in one place. In addition for the masters level sessions students are encouraged to give powerpoint presentations using Collorate. Using Collaborate means that students can stay in touch and feel part of the unit community. These live discussion with students and these sessions are a partial replacement of on-campus classes. About 4-6 one hour sessions are conducted throughout the study period in many units. These sessions are also recoded then made available for all enrolled students through Blackboard. Collaborate is also used to provide one to one consultations, where students can meet with the lecturer online, which allows for the flexibility of being able to demonstrate software, use apps on the web, draw on a whiteboard and live audio.

Discussion board: Blackboard's discussion board feature allows participants to carry on discussions online, at any time of the day or night, with no need for the participants to be logged into the site at the same time. The discussion is recorded on the course site for all to review and respond to at their convenience. Discussion threads are run through Blackboard throughout the study period for all postgraduate units. Often students post a question which other students and/or the instructor answer. They are also often driven by the instructor by posting questions on a specific learning activity or creating an activity in which students can participate. In some discussion threads instructors provide a link or a series of links and students follow the link(s) and report back through an instructor-defined set of questions. The discussion board builds class community by promoting discussion on unit topics and allowing time for in-depth reflection- students have more time to reflect, research and compose their thoughts before participating in the discussion. This helps the learner to develop thinking and writing skills.

Camtasia recording: Camtasia is a user friendly TechSmith product which allows students to access to computer generated audio visual training via online delivery. They capture both what the lecturer is doing on the screen and the voice. In some of the units a number of Camtasia recordings (audio/video) are used along with lecture recordings. These recordings are mainly used as a replacement of laboratory activities and to summarise topics included in unit content. Sometimes instruction about how to use a specific statistical tool or software are recorded in Camtasia and made available for students through Blackboard. In some the recordings, responses to general queries about the unit are recorded and made available for all students through Blackboard.

Short audio/video clip: In respose to requests from students short audio/video clips are adopted in some units to engage students with weekly learning activities. These materials are supportive for the lecture notes and weekend workshops. Short video clips help students to follow the steps when using statistical software (e.g. SPSS, SAS). These short clips are created using movie maker or similiar software or by an in house publisher called Swinburne Commons. There are many benefits to using video in education as shown in several decades of research. Allam (2006) demonstrated that the creative challenge of using moving images and sound to communicate a topic is indeed engaging and insightful, but adds that it also enables students to acquire a range of transferable skills.

Along with the instructional methods described above, e-mail communications are always appreciated as some students find it easiest to make direct contact with the instructor/convenor by email for urgent issues. The instructional methods described here have been well accepted by the learners and the feedback has been encouraging.

Table 1: Instructional methods (activities) used in medium-impact blend

Units	Delivery Type	Name of the method
All units	On-campus	Face-to-face (lecture/workshop)
All Graduate Certificate level units.	Virtual Classroom*	Blackboard collaborate class
STA60001, STA60004, STA60005	Virtual Classroom*	Elluminate Live session
All units	Online	Echo/Lectopia recordings (Lecture /workshop)
All units	Online*	Discussion board
All units	Online*	Camtasia recordings
STA60001, STA60004, STA6005, STA70002, STA70003, STA70004, STA80006.	Online*	Short audio/video clips

*Just like in a real-world classroom, a student in a virtual classroom participates in synchronous instruction, which means that the teacher and students are logged into the virtual learning environment at the same time. Online refers to having the learning information provided and certain assessments like quizzes available on a website at any time, where the learning and teaching do not occur at the same time, i.e. asynchronous instruction.

FEEDBACK

At the end of each study period (semester) feedback is obtained through a student feedback survey for the units offered through Swinburne and OUA. The vast majority of students, 80% out of 102, responded positively on the current learning structure. The feedback reveals that the flexible course structure with classes outside business hours and the blended learning structure are appropriate in postgraduate applied statistics programs for a mixed-cohort of students. The application of a medium-impact blended learning structure using a variety of extra online activities along with face-to-face on-campus classes has been appreciated by past and current students. The unit satisfaction ratings received through the surveys run by Swinburne through student feedback survey (SFS) and OUA have been excellent. During the last 3-4 study periods the overall mean satisfaction rating was above 80% and for many of the units the satisfaction rate was 100%. The overall mean satisfaction rate was 84.5% (study period 1: 84%, study period 2: 85%) for the period 2010-2014 with a response rate of 60% and it was 100% for international students with a response rate of 74%. Numeric ratings of teaching satisfaction (SFS) and units is very high, out of a maximum possible rating of 6, average over the reporting period (2010-2014), the satisfaction with teaching averaged 5.10 and 5.13 for semesters 1 and 2 respectively, and satisfaction with units averaged 4.81 and 4.82 for semesters 1 and 2 respectively. The response rates for these surveys were between 40% and 55% for student cohorts of 120 and 200. A few randomly selected qualitative responses received from the students during the period 2010-2015 regarding for their experience with these programs are quoted below.

> "It is really good to know that there are so many modes of study in this course i.e. workshops, online discussions, print materials. And it is good to see that you are keeping an eye on our progress. I am planning to attend workshop 3".
>
> "Not sure if you've had any other feedback - just wanted to let you know that I am finding the Lectopia recordings excellent. I haven't been able to attend the lectures in person yet, and I feel that I haven't missed anything! It works really well as I have the recording going as I work through the demonstrations at the same time.
>
> "I pop the recordings on in quicktime or media player - make sure the video file always appears on top, and run the lecture / my R console side by side. Works really well."
>
> "Just wanted to drop you a line to thank you for running such a well-organised course".
>
> "I graduated from the Masters of Science in Applied Statistics last year and I just wanted to thank you for all your efforts in running the course. I enjoyed the course and found it worthwhile, although I didn't make it particularly easy on myself working full time all the way through. However, the online format and the responsiveness of lecturers to any problems made it much smoother though".
>
> "Jahar, thank you very much for your mentoring this term, I have learnt a lot this term. It was tiring, but more importantly it was a very fulfilling time too. What I have taken from this course will definitely help me to succeed in the career which I want to pursue".
>
> "I am graduating after this and just wanted to say that I have thoroughly enjoyed your courses. I have managed to get a …promotion and a step up in the company so the program has definitely paid off for me :-)".
>
> "I am really getting a lot out of what I learned at Swinburne in my Masters course for Applied Stats. I use what I learned every day in my current role, and am really enjoying myself".

"I want to say that I thoroughly enjoyed the unit and that I think I'll find it very useful professionally".

"Thanks again for all your help over the degree; it's been an absolute pleasure. In particular, I feel like I am taking truly practical skills away that I am already using in the real world".

"This course felt like an on-campus course that I was watching online".

As well as such comments, student feedback for each of the units offered through Swinburne and OUA are collected and evaluated at the end of each study period. These are evaluated by the unit and course panel and, where appropriate, changes incorporated the next time the unit runs. This ensures continuous improvement in our program.

CONCLUDING REMARKS

Overall, the medium-impact blended learning structure has provided students with very useful flexibility and a variety of options and tools to engage with unit learning activities. Students have reported that guest lecturers are very useful for professional learning. The flexible blended learning approach adopted for the applied statistics postgraduate programs has been highly appreciated by this mixed cohort of learners and made the courses available to a much larger group. Finally, the sustained level of student ratings and the satisfaction ratings obtained by Graduate Careers Australia (GCA) for these postgraduate programs show that the quality of teaching and overall satisfaction in the blended learning structure has been maintained and is well accepted by the students. Our experience suggests that appropriate choice, integration and balance of different instructional methods, using a combination of technologies and pedagogies, are important in statistical education, especially for part-time post graduate students.

REFERENCES

Alammary, A., Sheard, J. & Carbone, A. (2014). Blended learning in higher education: Three different design approaches. *Australian Journal of Educational Technology*, 30(4), 440-454.

Allam, C. (2006) in Bijnens, M., Vanbuel, M., Verstegen, S., Young C., Handbook on Digital Video and Audio in Education, Creating and using audio and video material for educational purposes, The Videoaktiv Project.

Biggs, J. (2003). *Teaching for quality learning at university: what the student does,* 2nd Edition, Buckingham: Society for Research into Higher Education and Open University Press, UK.

Biggs, J. and Tang, C. (2011). *Teaching for quality learning at university,* 4th Edition, Open University Press, UK.

Chen, W. & Looi, C. (2007). Incorporating online discussion in face to fact classroom learning: A new blended learning approach. *Australian Journal of Educational Technology*, 23, 307-326.

Ghahari, S. (2013). The Effect of Blended Learning vs. Classroom Learning Techniques on Iranian EFL Learners' Writing. International Journal of Foreign Language Teaching & Research, Vol.1, No.3, pp. 1-9.

Kaleta, R., Skibba, K., & Joosten, T. (2007). Discovering, designing, and delivering hybrid courses. In A. G. Picciano & C. D. Dziuban (Eds.). *Blended learning research perspective* (pp. 111-143). Needham, MA: Sloan-C.

Kember, D. & McNaught, C. (2007). *Enhancing University Teaching: Lessons from Research into Award Winning Teachers*, Abingdon, Oxfordshire: Routledge, UK.

Means, B., Toyama, Y., Murphy, R. & Baki M. (2013). The Effectiveness of Online and Blended Learning: A Meta-Analysis of the Empirical Literature. *Teachers College Record, 115, 030303, 1-47.*

Oliver, M. & Trigwell, K. (2005). Can 'Blended Learning' Be Redeemed? *E–Learning,* Vol 2, No. 1, pp.17-26.

Park, S. Y. (2009). An Analysis of the Technology Acceptance Model in Understanding University Students' Behavioral Intention to Use e-Learning. *Educational Technology & Society*, *12* (3), 150–162.

Sharpe, R., Benfield, G., Roberts, G., & Francis, R. (2006). The undergraduate experience of blended e-learning: A review of UK literature and practice. Retrieved from http://business.heacademey.ac.uk/assets/documents/research/literature_review/blended_elearning_full_review.pdf.

Tishkoveskaya, S. and Lancaster, G.A. (2012). Statistical Education in the 21st Century: a Review of Challenges, Teaching Innovations and Strategies for Reform. *Journal of Statistics Education,* Vol 20, No. 2, pp.1-56.

LESSONS FROM INTEGRATING ONLINE AND FACE-TO-FACE LEARNING IN OUR WORKPLACE

WESTBROOKE, Ian and Van DAM-BATES, Paul
Department of Conservation,
New Zealand
iwestbrooke@doc.govt.nz

With the massive increase in technology and the availability of data of all kinds, there is a growing need for greater statistical and data skills in the workplace. A good part of this will be met by upskilling staff with little or no background in statistics or data analysis. Established approaches to upskilling include short courses and one-to-one mentoring, which we have discussed in recent ICOTS papers. Technology also opens up new opportunities and approaches to meeting these needs, and opportunities for more active involvement in learning. We have added a short pre-course module to an existing 3-day course and a post-course tele-presentation where participants present a project based on their own data. We recently established an online-based course titled Data Wrangling, and are developing further courses to strengthen skills in data extraction. We discuss lessons from our experiences, and why we favour mixing online and more traditional modes of delivery.

INTRODUCTION

There is growing demand for statistical skills in the workplace, driven by massive changes in technology and the availability of data. Increasing the skills of staff is one way of meeting these demands (Westbrooke & Ellis 2014). Modern technology also opens opportunities and possibilities to go beyond the traditional face-to-face classroom for training.

STATISTICAL TRAINING IN THE WORKPLACE

The statistics education literature focuses mainly on the educational sector, rather than the workplace. Most of the limited information available is from proceedings of the 4-yearly International Conferences on Teaching Statistics (ICOTS), which include the workplace as a topic area. Barnett (1991) looked at meeting statistical needs in industry; suggesting that either trained statisticians be employed; or the statistical skills of other staff be developed. A common thread is the importance of context for workplace training. There is a need for emphasis on real data and applied problems taught through hands-on computing rather than focussing on mathematical skills. See Westbrooke & Rohan (2014) for more detail and references.

OPPORTUNITIES FROM TECHNOLOGY

The use of technology in teaching statistics and similar subjects can be grouped into 3 main approaches

- Traditional classroom – usually focusses on lectures and presentations as well as referencing written notes, books and papers; computer-based labs/workshops and homework.
- Online – with elements such as written material; video/audio; interactive computer-based tutors; virtual tutorials; forums (some in real time) and organised peer-to-peer interaction.
- Mixed mode – a combination of both traditional and online, balancing the two approaches.

GOING ONLINE - ADVANTAGES AND CHALLENGES

Online approaches provide greater flexibility.

- They can provide greater scope for active learning, where the learner takes greater control of what they learn and how they learn it. Students can work through exercises and problems at their own pace, engaging in actual problem solving, learning how to access information and become self-sufficient.
- The ability to choose when and where time is allocated to learning provides great advantages for adult learners in the workplace. This is especially important in a dispersed

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

organisation like ours where learners are spread all over New Zealand, including some of the most remote places, such as Chatham Island.

- Flexibility means that the one course can provide for learners with a wide range of backgrounds and existing skills. Those with a strong background can work through material they are familiar with rapidly, while those with little or no background can work through more slowly and methodically.
- Trainers can provide material and links to additional material in different forms which provides for multiple learning strengths of different learners. This can include written, video and audio material, plus access to online quizzes and exercises for extension and reinforcement.
- In place of physical gatherings at lectures, tutorials and labs come flexible approaches to learner interaction, through online meetings and forums.
- Freeing up trainers' time can allow focus on individual learners expressing needs

Challenges in online training
With greater flexibility comes substantial challenges.

- The flexibility in where and when to learn can make maintaining engagement more difficult. Inevitably there will be tension for learners balancing time for training with normal work tasks.
- There is less contact between trainers and learners; and between learners unless they happen to be co-located.

CASE STUDIES:
Adding an online module to an existing course

Our first venture in using online approaches involved adding an introductory online module to our three-day classroom course *Statistical Modelling* (See Westbrooke 2011). The module provided text and an exercise we wrote, together with links to data and pre-existing online video resources, ensuring participants have a basic familiarity with software – R and R Commander – and that they have experience accessing, exploring and graphing data before coming to class. This can free up a quarter of a day, and more importantly helps to level the playing field between those who have little or no previous experience with the software, and those who have more substantial experience. It also provides the trainers with feedback on the level learners are starting from before we meet up. The module has worked well over several years now, and has become built into the course. Recently we have trialled successfully adding a post course presentation a month after, with a teleconference where participants from the course present an application to their own data.

Developing an online course
Recently we have developed an all-online course called *Data wrangling* to develop skills using the statistical software R, manipulate and explore data and learning best practices to create reproducible research. There are three chapters or modules. The first is an introduction to coding in R. This includes loops, functions, logic, and data structure. The second provides the core of the course, involving tools for manipulating data in R such as *data.table, reshape* and *dplyr*. The last chapter introduces graphing in R in the base environment and using ggplot. Each chapter provides a written guide to work through with mini-quizzes to test the student's understanding. In addition, we recorded some simple video instructions, and provide links to other videos and online introductory tutorials for R.

Each chapter ends with an assignment, to be completed using tools within R to create reproducible reports. These assignments are to test understanding but also to force the student to go beyond what was taught in the chapter and find the resources online to solve the problem. Students submit their assignment through the organisation's standard document management system through an RMarkdown (.RMD) file. Completing the course involves a final project based around 'wrangling' data related the student's work. For support and to provide for peer-to-peer interaction, we set up an online forum through the workplace intranet, which has encouraged a statistics network and students working together. This helps alleviate some of the issues with reduced peer to peer interaction from an online course.

We initially envisaged most course participants would complete all three chapters in a session of about eight weeks. An initial pilot showed us that a more flexible approach was needed – so that each chapter could be completed as a milestone, provided it included an exercise demonstrating skills learnt applied to their own data. Further chapters can then be completed in later sessions of the course. By modularising the course, we can accommodate both new and more advanced users. This is another advantage of an online course where we can deliver material at many different levels.

We have found there are real challenges to sustain student involvement at a distance and over time. The pilot course that was run in our own location worked well, as we could reinforce and sustain interest directly; there were greater opportunities for trainer/learner feedback. We gathered students together for a final course presentation in person. To help with sustaining involvement, we've set each of 3 chapters as successes/milestones in themselves with flexibility to complete one, or all, in future sessions. A virtual tutorial session by teleconference helped engagement at first but had only partial success as students got busy with their other work. We will need to continue to develop approaches to sustain involvement. One aspect for consideration could be classroom-based sessions within the course.

Active learning without online content

Another course on *Designing studies* will remain as a three-day class room course without online components. This already has active learning integral to its approach with group exercises and discussions focussed on design examples from their own work that each participant brings to the course. The discussion aspect in this course is the main element as we teach the students how to define their research objects and work through designing a study.

HOW MUCH TO DEVELOP IN-HOUSE

With the critical importance of context for workplace learning, there are advantages to developing material in-house to meet learners needs. However, this must be balanced against the resources required for development, especially as the in-house audience is likely to be limited in size. Both generic and very specialised needs are likely to be best out-sourced; and maximum use made of existing material and resources when developing in-house. We made a deliberate decision not to provide basic Stats 101 courses internally, but to point staff in the direction of extra-mural courses or local courses available from universities. Some needs were not readily met from existing providers. We have developed two core courses in-house, both focussed on observational data typical in conservation monitoring: statistical modelling covering linear and generalised linear models and a designing observational studies course.

CONCLUSIONS

Active learning can be facilitated in several ways in a workplace context. More traditional and online approaches as well as mixes of both have worked for us. We have found sustaining involvement in a new venture into an online-based course remains a challenge. In a workplace, there are natural busy periods such as summer activity for field research when scheduling training is inadvisable. Future offerings of this course will aim for timing that is optimal for staff engagement. The timing and energy invested in development for a limited potential audience in-house, means it is advisable to use, adapt, or include existing resources as much as possible – not start from scratch. Ideally, resources developed here will be shared and built upon by groups involved in similar types of statistical training.

ACKNOWLEDGEMENTS

We would like to thank Shannan Mortimer and his colleagues in the capability development team in our workplace for their support for our online development, and the participants in our various courses that we have learnt so much from.

REFERENCES

Barnett, V. (1991), Statistical trends in industry and in the social sector. In *ICOTS3: 3rd International Conference on Teaching Statistics: papers and abstracts.* International Statistical Institute, 440-445. http://iase-web.org/documents/papers/icots3/BOOK1/C8-3.pdf

Westbrooke, I. (2011). Statistics education in a conservation organisation—towards evidence based management. 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, http://iase-web.org/Conference_Proceedings.php?p=ICOTS_8_2010

Westbrooke, I.; Ellis, P. (2014) Training to develop modern statistics in the workplace using R and R Commander - experiences from the New Zealand government sector *Proceedings of the Ninth International Conference on Teaching Statistics. Flagstaff, Arizona* 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.* Voorburn, The Netherlands: International Statistical Institute. http://icots.info/9/proceedings/pdfs/ICOTS9_5F3_WESTBROOKE.pdf

Westbrooke, I.; Rohan, M. (2014). Statistical training in the Workplace. In MacGillivray, H., Martin, M., and Phillips, B. (eds) *Topics from Australian Conferences on Teaching Statistics: OZCOTS 2008-2012*, Springer Science+Business Media, New York.

MASTERING INTRODUCTORY STATISTICS: EXPERIENCES AND OUTCOMES FOR A PUBLIC SERVICE COHORT LEARNING INTRODUCTORY STATISTICS

ASCIONE, Judith[1], QURESHI, Sumaira [1], RICHARDSON, Alice[2] and THANDRAYEN, Joanne [1]
[1]University of Canberra
[2]Australian National University
Judith.Ascione@canberra.edu.au

A unique combination of circumstances culminated in late 2015 in thirty Public Service employees learning introductory statistics as part of a graduate degree at the University of Canberra. Those unique circumstances were (1) a cohort of graduates (2) employed in a single Australian Public Service agency (3) studying introductory statistics in a graduate degree (4) using a Mastery learning model. This model involves assessment of small quantities of material, with a high (80%) pass mark set, but multiple attempts allowed in order to achieve a pass. We will describe the history and rationale of the Mastery approach to learning in general, and give details of the circumstances that gave rise to this unique course. We will also look at the quantitative and qualitative experiences of the students learning statistics in this way. Finally, we will offer suggestions for those planning to teach introductory statistics using Mastery learning, based on our experiences.

MASTERY LEARNING AND THE FLIPPED CLASSROOM

In 2013-2014 a Structural Adjustment Fund for Flexibility, Innovation, Retention and Engagement (SAFFIRE) was made available at the University of Canberra, and one of the adjustments identified for implementation was a Mastery Assessment Program in Science, Technology, Engineering and Mathematics (including Statistics).

Mastery learning is not new, going back to the Keller Plan of the 1960s (Eyre 2007) for teaching first-year mathematics that was still in use in the 1980s in New Zealand. Wong and Kang (2012) found that Mastery learning continues to be an effective strategy in university courses where it is important to understand previous topics in order to master subsequent ones.

Mastery learning involves the design of learning objectives, aligned with carefully chosen and developed learning material and with assessment items that address those objectives. When students complete a segment of work they demonstrate that they have learned the content by taking an unsupervised Mastery assessment. If a student fails to achieve the standard they go back and re-learn concepts, seek advice from a tutor and so on. They then re-take a different version of the Mastery assessment.

Students work at their own pace through the objectives within the course, possibly tackling more than one objective simultaneously. Students can monitor their progress on the university's Learning Management System (LMS).

In more recent years, several publishing houses have produced online systems to support Mastery learning in a variety of disciplines. Carey, Christie and Granger (2015) report on the use of Pearson MyWritingLab for pre-service school teachers. Shotwell and Apigian (2015) report on the use of a statistics product from McGraw-Hill in a business statistics course. Nicola Petty writes about replacing traditional statistics lectures with an online mastery assessment system too, at http://blog.testsoup.com/replacing-traditional-lectures/.

The arrival of these online systems has greatly enhanced an instructor's ability to proceed with another popular teaching and learning strategy, namely the flipped course This involves delivering lecture-style content online and using face-to-face time for more active learning experiences. Flipping can take place on the micro scale, with regular swapping between online and face-to-face experiences as a semester progresses (Winquist and Carlson 2014). Mills and Raju (2011) reviewed a decade of literature on teaching statistics online to find that at the end of their period under study (2006 – 2009), appropriate technology, continued interaction between students and instructors, and careful monitoring of the teaching and learning process were all important contributors to a successful course.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

STATISTICS FOR MANAGERS G

In late 2015 the authors at the University of Canberra encountered these unique circumstances: (1) a cohort of graduates (2) employed in a single Australian Public Service agency (3) studying an introductory statistics course in a graduate program (4) using a Mastery learning model. The course that forms the focus of this paper was called Statistics for Managers G, and is introductory in content and graduate in level.

All students (n = 33) had at least a bachelor's degree, sometimes with statistics, the majority without (and with majors ranging from Economics to History). The employment status of the students brought time constraints to their learning which often took place after-hours, or in concentrated bursts during employer-approved study leave. On the other hand it also brought a cohort of highly motivated students, sponsored by their employer, who have been chosen through a competitive process involving nomination by their supervisors. Education and training in the Australian Public Service is generally handled at the agency level. Specialised agencies conduct in-house education or, as in the case of the agency involved in this course, tender for outside organisations to provide it.

The online Mastery component was delivered by Pearson's MyStatLab, based on the textbook by DeVeaux et al. (2014). Most introductory statistics textbooks claim they are useable at both the graduate and undergraduate level. However research has shown that graduate students can bring their own baggage to the statistics course as often it is unrelated to their previous coursework at which they did well, leading to anxiety (Pan and Tang 2004).

Students read twenty prescribed chapters of the textbook, watched animations (voice over Powerpoint) and other short videos (often to provide a humorous motivation to the statistical topic). When students were ready they took a test (multiple choice questions, with randomly generated numbers) for which the pass mark was 80%. Each test was expected to take about 30 minutes to complete. Students could take as many attempts as they required to reach the pass mark. Once a week an optional "virtual tutorial" using Blackboard Collaborate was held, where students could interact with the lecturer and each other with voice and/or video. Students were expected to complete the MyStatLab component in seven weeks. This represents a raised intensity of learning over the usual semester-long (twelve to thirteen week) presentation of a university course.

The flipping in this course was on a macro scale, with seven weeks of independent and online learning followed by a three-day residential. As noted by Bakker and Akkerman (2013), a residential can enhance the level of integration between the online component of the course and more work-related knowledge that we expected students would construct. A key part of the residential was group work on a presentation, with simultaneous peer marking. A week after the residential, students submitted individual written projects based on the same data as was analysed in the presentations. The experience of staff and students in analysing this data will be reported elsewhere.

The aim of this paper is to examine the effectiveness of mastery learning of introductory statistics. The outcomes of the course will be examined in terms of data analytics generated by the Mastery system itself, student results and student evaluations of their experiences. The specific objectives of this paper will be to

- Examine the characteristics of the time spent on each chapter
- Study student perceptions of the Mastery learning system

Finally we will offer some thoughts for those planning to teach introductory statistics using Mastery learning, based on our experiences.

METHOD

Characteristics of the time spent on each chapter, test scores and overall course mark and its relationship to the test score were examined using descriptive statistics.

Student perceptions of MyStatLab were analysed using the CRiSP instrument (Richardson et al. 2015) whose items are given in Table 1. Ten of the original 27 items were omitted as they were specific to classroom response systems rather than online learning systems. The tool was administered online though the LMS and was anonymous. To carry out the factor analysis, five items were reverse-coded (Wasted time; Tech problems; Too difficult; Expectations hard; Correct but not understand), so that large scores reflected favourable outcomes for all items and so that all loadings were expected to be positive. Due to the small sample size, it was not possible to extract the factors all at once.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

Table 1.The 17 ordinal-scale items used in the questionnaire. All items are answered on a five-point ordinal scale, from Strongly Disagree to Strongly Agree.

Short description	**Complete question**
Wasted time	Using MyStatLab wasted too much time
Recommend use	I would recommend that the lecturer continue to use MyStatLab
Overall value	The use of MyStatLab helped increase the classes' overall value
Motivation	MyStatLab used in this course motivated me to learn
Instant feedback	MyStatLab helped me get instant feedback on what I knew and didn't know
Understand concepts	MyStatLab allows me to better understand key concepts
Enhanced learning	Using MyStatLab enhanced my learning of the subject
Control over learning	I believe that MyStatLab provided me with more control over my learning than in courses that do not use MyStatLab
Think deeply	Using MyStatLab helped me think more deeply about course materials
Correct but not understand	I often selected the right answer without really understanding
More confident	Using MyStatLab made me more confident to participate in class
Increased participation	MyStatLab increased the frequency of my direct participation on the course
Easy to use	For me it was easy to use MyStatLab
Too difficult	For me MyStatLab was too difficult to use
Expectations hard	It was too hard to know what was expected of me using MyStatLab
Tech problems	There were too many technological problems using MyStatLab
Increased enjoyment	Using MyStatLab has increased my enjoyment of lectures

Other questions

Short description	**Complete question**
Age	Age (25 – 29, 30+)
Gender	Gender (Male, Female)

RESULTS

Analysis of the MyStatLab internal data

On average students sat 33.18 tests across the 20 chapters; in other words, most students took one or two attempts at a chapter. Three students were aiming for straight 100% scores and contributed more heavily to the extra tests.

Half of the tests were open for one hour or less, and three-quarters of them were open for two hours or less. Students were able to walk away and leave a test open for an unlimited amount of time, and they may not have been working on the material for all that time either. Anecdotal student feedback suggests that some quizzes did genuinely take a long time to complete, bur time actively on the system is not available. This is a flaw in the data collection that should be remedied in future releases of the mastery system. Figure 1 shows that there was considerable variation between chapters in terms of the length of time taken for a test, with medians ranging from one to three hours. The quizzes that took the longest were on conditional probability, sampling distributions, inference for one mean and comparing groups.

The mean best score in the 20 quizzes across all students was 84% with 47% of students scoring 90% or more. The mean final score in the course was 84.4% (95% CI 82.2% to 86.5%).

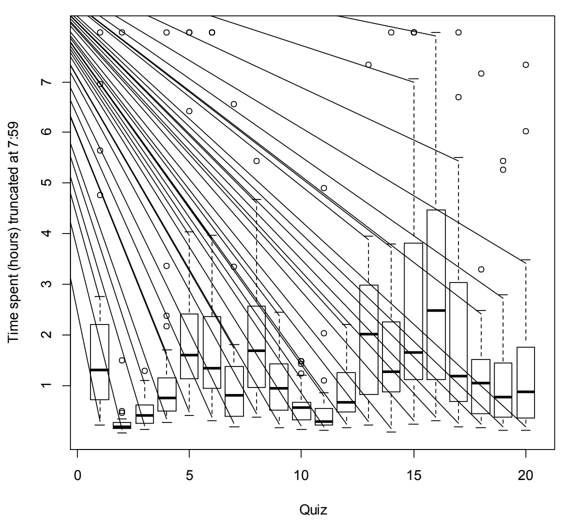
Item analysis of the CRiSP questionnaire

Among the four items identified as making up the Engagement factor (Richardson et al. 2015), the value of the Kaiser-Meyer-Olkin statistic is 0.80. Among the nine items identified as making up the Learning factor, the value of the Kaiser-Meyer-Olkin statistic is 0.76. Among the four items identified as making up the Engagement factor, the value of the Kaiser-Meyer-Olkin statistic is 0.63. These figures all exceed 0.5 indicating that factor analysis is suitable based on the data structure (Pallant 2002). The Spearman correlations between the 17 items range from -0.12 to 0.86; this

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

suggests that no two items are so highly correlated that they should be removed before starting the analysis (Pallant 2002).

Figure 1. Boxplots of amount of time by chapter, that the test with the highest mark was open. Times over 8:00 hours have been truncated to 7:59.



Quiz contents

1. Getting ready (basic mathematics)
2. What is statistics?
3. Categorical data
4. Quantitative data
5. Comparing distributions
6. Normal distributions
7. Correlation
8. Regression
9. Diagnostics
10. Sampling
11. Experiments
12. Probability
13. Conditional probability
14. Sampling distributions
15. Inference for one mean
16. Comparing groups
17. Paired samples
18. Chi-square
19. Regression inference
20. ANOVA

We chose to employ an oblique rotation method (oblimin), as the impact of a technology on engagement and learning was not expected to be orthogonal. A parallel analysis (Horn 1965; Pallant 200, p. 193) suggested that three factors should be extracted.

Nineteen of the 33 students responded and the distribution of the responses by item are shown in Figure 2. All the respondents noted or responded positively to the instant feedback provided by MyStatLab, and the majority thought that the MyStatLab enhanced their learning of the subject, added to the class's overall value and they would recommend its continued use. On the other hand, about 40% of the respondents thought that MyStatLab did not help them think more deeply about course material and about 65% though that Using MyStatLab wasted too much time.

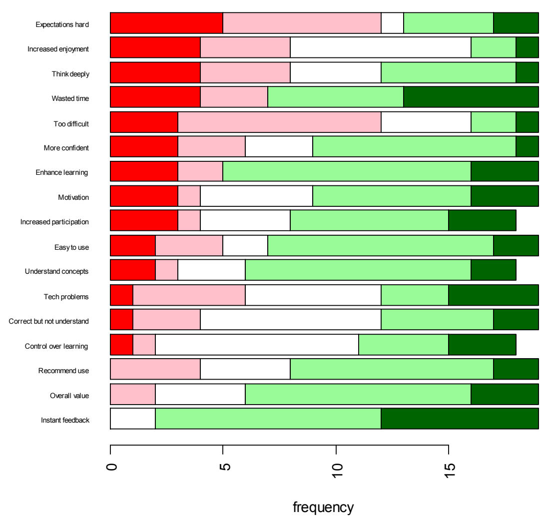


Figure 2. Bar chart of responses to CRiSP items (see Table 1 for the full text of items), ordered by most "strongly disagree" responses to least. Full text of items is given in Table 1. Extreme left = strongly disagree; left = disagree; centre = neutral; right = agree; extreme right = strongly agree.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

Factor analysis of the CRiSP questionnaire

Factor analysis using factanal (Revelle 2015) in R (R Core Team 2013) from the 19 respondents produced the loadings shown in Table 2. All loadings exceeded 0.4 so all are reported. Three factors were extracted, consistent with Richardson et al. (2015). The value of Cronbach's alpha (Cronbach 1951) and Guttman's lambda-6 (Guttman 1945) suggest the three sub-scales have high reliability.

The three factors each explained between 50% and 75% of the variation in the data (the sum of the percentages exceeds 100% because the factors were extracted one-at-a-time). The proportions are higher than achieved by Richardson et al. (2015), which is likely to be due to the small sample size and the homogeneity of the cohort of students. All but one student was in the 30+ age group, no test for age as a confounding variable is reported. No evidence of differences according to gender on the three scales (Engagement, Learning and Usability) is evident using Mann-Whitney tests.

Table 2. The three factors obtained one-at-a-time from 19 students. An oblimin rotation was used, and all loadings exceeding 0.4 are listed. Item names followed by an asterisk were reverse coded, so that large scores reflect favourable outcomes for all items.

Item name	Engagement	Learning	Usability
Increased enjoyment	0.905		
Increased participation	0.839		
More confident	0.864		
Motivation	0.841		
Control over learning		0.657	
Correct but not understand (*)		0.472	
Enhance learning		0.942	
Instant feedback		0.455	
Overall value		0.763	
Think deeply		0.830	
Understand concepts		0.965	
Wasted time (*)		0.519	
Recommend use		0.691	
Easy to use			0.958
Tech problems (*)			0.431
Expectations hard (*)			0.524
Too difficult (*)			0.801
Variance explained	74.4%	52.2%	50.5%
Cronbach's alpha	0.91	0.88	0.76
Guttman's lambda-6	0.91	0.95	0.77

Table 3. Summary statistics of CRiSP scales (n = 19).

	Engagement				Learning				Usability			
	Mean	Med	SD	IQR	Mean	Med	SD	IQR	Mean	Med	SD	IQR
	3.11	3.5	1.12	1.31	3.36	3.67	0.83	1.0	3.30	3.5	0.94	0.88

MyStatLab rated highest on the Learning scale, followed by Usability and Engagement. All means were above the mid-point of 3, suggesting a broadly positive experience for students.

CONCLUSION

Statistics for Managers G used a flipped classroom with a mastery learning approach for the online learning part of the course. This unique combination of circumstances is both a limitation and a strength of this paper, in the sense that the baseline data analysed here has no natural comparison groups available.

The students rated the mastery learning tool just above average on the Learning, Usability and Engagement scales of the CRiSP instrument. In terms of the special characteristics of the cohort using the mastery system we found that (1) graduates engaged well with the concept of mastery learning, with the competitive streak of 10% of the class being fully engaged (2) the mastery system well accommodates intensive learning, such as is demanded by full-time employees studying part-time (3) statistics as a discipline is well-suited to Mastery learning when supplemented by boundary-crossing activities such as the three-day workshop and data analysis project that featured in this course. As with any new tool, we recommend that instructors allow plenty of time for setting up and testing the system before it is released to students. We also recommend that instructors curate a rich collection of animations, videos and text to offer experiences for learners of all styles. Finally, the usefulness of face-to-face contact, whether by "virtual tutorial" or in a workshop, cannot be underestimated.

REFERENCES

Bakker, A. and Akerman, S.F. (2014). A boundary-crossing approach to support students' integration of statistical and work-related knowledge. *Education Studies in Mathematics, 86,* 223 – 237.

Carey, M.D., Christie, M. and Grainger, P. (2015). What benefits can be derived from teaching knowledge about language to preservice teachers? *Australian Journal of Teacher Education, 40(9)* [online] http://ro.edu.edu.au/ajte/vol40/iss9/2.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16(3),* 297–334.

De Veaux, R.D., Velleman, P.D., and Bock, D.E. (2014). Intro Stats 4th edition. Pearson.

Eyre, H.L. (2007). Keller's personalised system of instruction: was it a fleeting fancy or is there a revivial on the horizon? *The Behavior Analyst Today, 8,* 317-324.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10(4),* 255–282.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179–85.

Mills and Raju (2011). Teaching statistics online: a decade's review of the literature about what works. *Journal of Statistics Education, 19(2)* [online] http://www.amstat.org/publications/jse/v19n2/mills.pdf.

Pallant, J. (2002): *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS*. Crows Nest: Allen and Unwin.

Pan, W. and Tang, M. (2004). Examining the effectiveness of innovative instructional methods on reading statistics anxiety for graduate students in the social sciences. *Journal of Instructional Psychology, 31*, 149-159.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org/.

Revelle, W. (2015) psych: procedures for personality and psychological research. Northwestern University, Evanston, Illinois, USA. https://cran.r-project.org/web/packages/psych/index.html. Richardson, A.M., Dunn, P.K., McDonald, C. and Oprescu, F. (2015). CRiSP: an instrument for assessing student perceptions of classroom response systems. *Journal of Science Education and Technology, 24,* 432 – 447.

Shotwell, M. and Apigian, C.H. (2015). Student performance and success factors in learning business statistics in online vs. on-ground classes using a web-based assessment platform. *Journal of Statistics Education, 23(1),* [online] www.amstat.org/publications/jse/v23n1/shotwell.pdf.

Winquist, J.ER and Carlson, K.A. (2014). Flipped statistics class results: better performance than lecture over one year later. *Journal of Statistics Education, 22(3)* [online] http://www.amstat.org/publications/jse/v22n3/winquist.pdf .

Wong, B. S., and Kang, L. (2012). Mastery learning in the context of university education. *Journal of the NUS Teaching Academy, 2(4),* 206-222.

COLLABORATIVE TEACHING AND LEARNING IN LARGE QUANTITATIVE UNITS

KAUR, Charanjit and BROOKS, Robert
Department of Econometrics and Business Statistics,
Monash University, Caulfield East, Vic,
charanjit.kaur@monash.edu

Massification of education has brought about new challenges in the fore of teaching and learning. This paper investigates the challenges facilitators encounter when teaching a large postgraduate statistics unit with heterogeneous student cohorts. Postgraduate students with varying levels of statistical literacy/numeracy and understanding of basic concepts as well as low level interaction in classes are identified as some of the key teaching and learning problems. Managing multiple tutorials in a traditional setting is also causing difficulties in terms of inconsistency in delivery and application of key statistical concepts. There seems to be a lack of support provided for the professional development of less experienced tutors. This paper proposes an approach to combining group work and "super tutorials" that produce an effective and interactive learning environment for quantitative units. The use of an 'audience response system' within a large tutorial that caters for group work allows facilitators to effectively engage with students and obtain real-time feedback. The use of such technology can enhance student collaboration and student-instructor interaction. This paper also explores challenges faced in implementing changes and the effectiveness of co-teaching in super tutorials as a "check and balance" element.

Key words: *collaborative teaching and learning; super tutorials; audience response system.*

1. INTRODUCTION

In recent years, education has become increasingly diverse and complex. This is partially due to increased mobility of educators and students across nations. The significance of these developments is critical given the importance of the role of quantitative skills in most disciplines at both undergraduate and postgraduate levels. A business degree would almost always include an introductory quantitative unit (Brooks, et. al 2014). The ability to analyse data is known to be a skill desired by most employers.

As teachers and students come from different cultural and historical background, they may have different aspirations, values and expectations about the learning process in the classroom. This heterogeneous nature of international classrooms has consequences not only on the practice of teaching and learning, but also on teachers' roles, identity, and professionalism (Tran & Nguyen, 2015). There are several challenges when teaching a large quantitative unit. This paper explores the challenges faced in implementing changes and the effectiveness of co-teaching in a large tutorial setting.

In a large class, facilitating students' heterogeneity is a big challenge. Tishkovskaya & Lancaster (2012) state that "teaching statistical courses is challenging because they serve students with varying backgrounds and abilities" (p.1). We argue that facilitating the effective delivery of a statistics unit requires exploration of teaching strategies and approaches that can cater to varying levels of statistical literacy/numeracy and understanding of basic concepts. In order to enhance the learning experience in statistics, this paper investigates the challenges facilitators encounter when teaching a large postgraduate statistics unit with heterogeneous student cohorts.

As explored by Tran and Nguyen (2015), pedagogy and teaching practices are significantly influenced by cultural differences since culture is vital in shaping the way people learn. This view inspires the basis of the notion of a *culturally appropriate pedagogy*, which is an educational practice that takes into account cultural circumstances of both learners and teachers, and at the same time addresses competencies needed at a more global level (Nguyen, Terlouw, & Pilot, 2006).

Another educational approach that is potentially applicable to facilitate students with diverse backgrounds is constructivism. This perspective defines learning as a process within which individuals construct their own knowledge by giving meaning to their experience with their external world (Brown & King, 2000). This process of meaning making is very contextual and influenced by

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

individuals' cultural background which equip them with particular ways of thinking. The constructivist view is that learning can be facilitated by providing a learning situation that is relevant to the learners' contexts and real-world situation.

Identifying students' basic skills in statistics and computer as well as their cultural background is the first step we take in this unit. In semester one of 2016, we found that ninety seven percent of 750 students enrolled, were international students who have various acquisitions of essential statistics and computer skills. Approximately fifty six percent of them confirmed that they have studied statistic before and ninety per cent of them only have basic knowledge of the use of Microsoft Excel. These varying levels of statistical literacy/numeracy and understanding of basic concepts are some of the key learning problems identified in this unit. This problem is worsened by negative experiences with statistics and mathematics (Garfield, 1995)

Another major challenge in facilitating this unit is low levels of interaction in classes due to large enrollments, which we try to address by providing tutorials after lectures. However, managing multiple tutorials in a traditional setting where there are twenty five students and one facilitator has raised additional challenges in the consistency of delivery and application of key statistical concepts. There also seems to be a lack of support provided for the professional development of less experienced tutors.

We propose a collaborative approach in which we combine group work and "super tutorials" that produce an effective and interactive learning environment. Co-teaching in a "super-tutorial" setting also supports the collaboration between facilitators. This approach is inspired by the idea that to address students' diverse needs and various acquisitions of basic statistic and computer skills, teaching statistic should go beyond the traditional approach which focuses on "developing knowledge and on methodological skills, procedures, and computations" (Tishkovskaya & Lancaster, 2012, p.4). It highlights a "strong synergy among content, pedagogy, and technology" (Tishkovskaya & Lancaster, 2012, p.5).

In facilitating the engagement of facilitators and students, an 'audience response system' is used within a large tutorial that caters for group work and real-time feedback. New technology is increasingly used in the teaching of statistics in many reputable universities (Chance et al. 2007).

The proposed changes in the teaching plan are inspired by a range of philosophical theories. The sociocultural perspective lays the basis of cultural understanding of multicultural nature of this unit. We try to be aware of dissimilarities and different ways of beings and doings among students (Dantas, 2007) as well as gaps in cultural values including teacher-students power relationships, individualism-collectivism orientation in working on tasks, gender roles embedded in daily activities, tendency to avoid uncertain situations and time orientation (Hofstede & Hofstede, 2005 as cited in Nguyen, Terlouw, & Pilot, 2006). This awareness is needed to avoid cultural conflicts within this unit.

The plan of this paper is as follows. Section two outlines the proposed changed and methodology used, section three explains data collection techniques, section four is a review of the results and the final section provides concluding remarks.

2. METHODOLOGY

We propose an approach to combine group works and "super tutorials" as follows:

1. Organise "super-tutorials" which comprise of a two-tutor model with fifty students and two facilitators. The arrangement of small and large tutorials as well as in-class activities are informed by the constructivist perspective that suggests teachers to provide a "constructivist learning environment" (Brown & King, 2000, p.246) within which a group of learners (in the classroom setting) collaboratively construct and share basic knowledge as a learning community. A key principle in establishing this learning community is collaboration among learners (Brown & King, 2000). Group work caters for collaboration. Dividing students into smaller groups with five students in each group allows them to have more opportunities to contribute to group work.

2. Incorporate co-teaching in "super-tutorials" Co-teaching is one example of a model of effective professional development. In co-teaching, educators work together to examine student work and determine ways to improve teaching and learning practices (Desimone, 2009)[i]. A core strength of

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

co-teaching is the collaboration among two or more teachers who may have different ways of thinking and beliefs about teaching. Essentially, the individuals working together (i.e., the preservice teachers and cooperating teachers) establish trust, develop and work on communication, share responsibilities, and problem-solve to overcome the challenges of the diverse classroom environment.

Research shows that students in classrooms where co-teaching occurs become more engaged by working in smaller groups, receiving more individual attention, getting their questions answered faster, and get papers, assignments, and grades back faster (Teacher Quality Enhancement Center, 2010).

3. Establish a Tutors' Network in order to enhance and support learning as a communal activity. A Tutors' Network is a forum through which tutors share their weekly classroom experiences. This allows for the formation of a learning community in which tutors can learn from each other by sharing good practices. As suggested by Huffman et al. (2016), a community of learning can be established when educators are engaged in a culture that continuously enforces collaborative work so that everyone is teaching *and* learning at the same time. The Tutors' Network also allows tutors to reflect on their teaching practice. According to Farrell (2015), this reflective practice is vital for teachers to develop their capacity to be effective teachers, especially in the context of international education.

4. Incorporate the use of an audience system (Learning Catalytics) for all tutorials to allow facilitators to effectively engage with students and obtain real-time feedback. Technology is increasingly used in the teaching of statistics (Chance et al. 2007). Personal Response Systems also referred to as Electronic Voting Systems or Audience Response Systems) involve equipping students with a handset allowing them to send responses to questions put to them by the facilitator. The use of an 'audience response system' within a large tutorial that caters for group work allows facilitators to effectively engage with students and obtain real-time feedback. The use of such technology can enhance student collaboration and student-instructor interaction.

3. DATA COLLECTION

For the purpose of this study, data was collected in semester one of 2016. In week 0, students were sent a link to an online quantitative survey. The survey was administered through Moodle and was open for one week, with reminders being sent to all students. Using a five-point Likert scale, the survey assessed their background, learning experience regarding the use of Microsoft Excel, online learning and group work.

Toward the end of the semester, another survey was conducted to assess their feedback regarding group work and the effectiveness of Learning Catalytics.

4. DATA ANALYSIS AND RESULTS

We present descriptive statistics from the survey for semester one, 2016, providing further separate results for small tutorials as well as "super-tutorials". There were a total of 411 students who participated in the survey. Preliminary findings suggest that half of them have studied statistics before. Although ninety percent of them possess prior knowledge of basic Microsoft Excel, a majority are only familiar with basic functions in Microsoft Excel. These results can be seen in Figure 1 below:

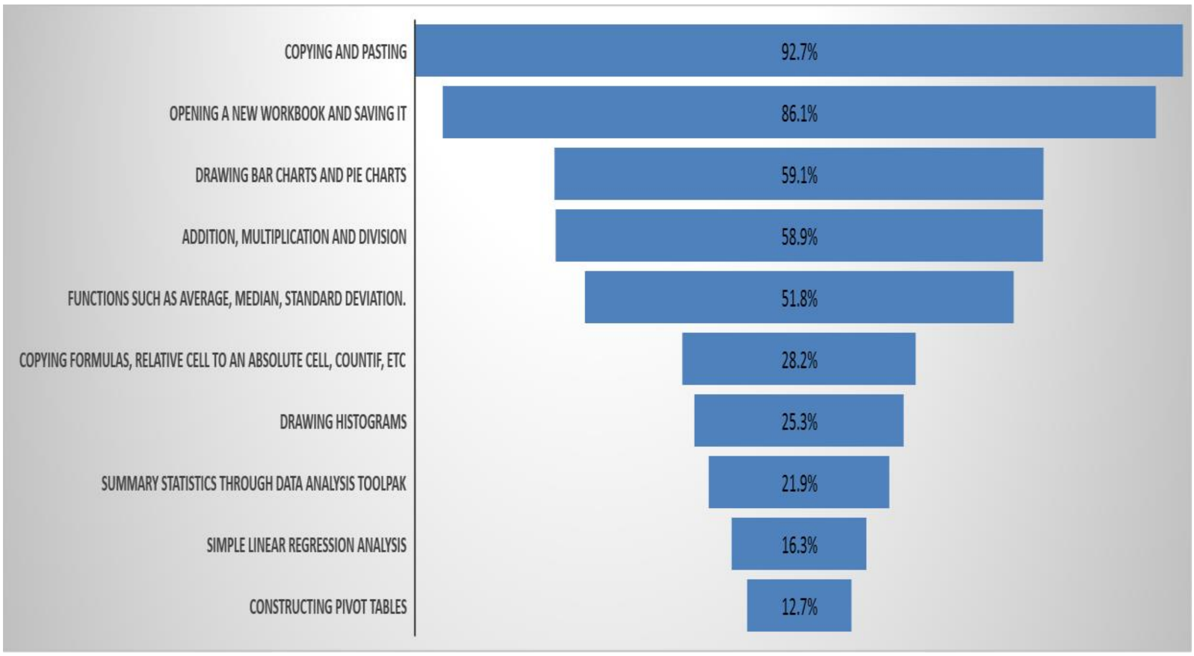


Figure 1: Microsoft Excel Skills

The results from the survey also show that although eighty percent have had experience submitting online assessments/quizzes, only fifty percent have ever been involved in online polls and less than forty percent have been involved in flipped learning. Table 1 below provides an extract of the results of the survey conducted toward the end of semester for both large ("super-tutorials") and small tutorials.

Table 1: Extract of Survey Results from Semester 1, 2016

QUESTION	CLASS SIZE	Strongly agree	Agree	Neutral	Disagree	Strongly Disagree
Group work during tutorials helps me understand the material better	LARGE	27.1%	42.8%	21.1%	6.6%	2.4%
	SMALL	25.5%	47.3%	19.7%	6.7%	0.8%
	TOTAL	26.2%	45.4%	20.2%	6.7%	1.5%
I want to have more group work	LARGE	14.0%	32.9%	34.8%	14.0%	4.3%
	SMALL	13.0%	35.3%	37.0%	11.3%	3.4%
	TOTAL	13.4%	34.3%	36.1%	12.4%	3.7%
Learning Catalytics in tutorials helps me understand the material better	LARGE	36.6%	49.4%	12.2%	0.6%	1.2%
	SMALL	34.7%	50.6%	11.7%	2.1%	0.8%
	TOTAL	35.5%	50.1%	11.9%	1.5%	1.0%

Overall, the results suggest that there is strong evidence to support group work with seventy percent of students agreeing that it has helped improve their understanding of the content. Almost fifty percent of students want more group work to support their learning activities. The results confirm that group experiences have been shown to contribute to student learning.

The results also suggest there is strong evidence to support the use of Learning Catalytics as a real-time feedback tool. We received a very positive result where almost all of students (above eighty percent overall) consider this tool as having helped them understand the material better. The main advantage of using this tool is that it enhances interactivity in class and encourages students to be involved in the learning process. This benefit is not necessarily inherent to the technology, but more of the shifting from passive to active learning that it encourages.

A Chi Square test was conducted to investigate if there was a significant difference in the responses received from small and "super-tutorials" regarding the effectiveness of group work and Learning Catalytics as a learning tool. The chi-square results show that there is insufficient evidence that the size of tutorials have an influence on the effectiveness of group work and Learning Catalytics as a tool to evaluate the learning processes.

Feedback obtained from tutors suggests that co-teaching in "super-tutorials" has contributed positively to the professional development of less experienced tutors. Tutors have also found communication and collaboration within the Tutors' Network effective in helping build their teaching skills.

5. CONCLUSION

This paper explores the challenges facilitators encounter when teaching a large heterogeneous cohort in a postgraduate statistics unit. We investigate an approach to combining group work and "super-tutorials" that produce an effective and interactive learning environment for large quantitative units. The use of an 'audience response system' within a large tutorial that caters for group work allows facilitators to effectively engage with students and obtain real-time feedback. The use of technology can enhance peer collaboration and student-instructor interaction.

Based on the results, we found the "super-tutorial" model with larger room sizes has been as effective as small tutorials. The use of small groups coupled with an audience response system encourages collaborative learning. This approach is well received in both small and large tutorials. The adoption of a common teaching approach (group work) and the use of a common feedback tool (Learning Catalytics) has also encouraged better interaction and collaboration between tutors. The collaboration encourages tutors to collectively reflect on their teaching practice. This reflective practice is vital if tutors are to develop effective teaching skills in an ever growing landscape of diverse learners.

Within a large group of students, it would be possible to incorporate team-based assessment within tutorials. This task is planned for future semesters.

REFERENCES

Brooks, R., Booth, R., Wright, J., & Suntah, N. (2014). Problem-based learning of statistical sampling concepts using fantasy sports team data. In H. MacGillivray, M Martin and B. Phillips (eds.), Topics from Australian Conference on Teaching Statistics: OZCOTS 2008-2012, Springer New York.

Brown, S.W., & King, F.B. (2000). Constructivist pedagogy and how we learn: Educational psychology meets international studies. *International Studies Perspectives. (*1), pp. 245–254.

Chance, B., Ben-Zvi, D., Garfield, J., and Medina, E. (2007). The Role of Technology in Improving Student Learning of Statistics, *Technology Innovations in Statistics Education*, *1(1).* Retrieved from: http://www.escholarship.org/uc/item/8sd2t4rr

Dantas, M. L. (2007). Building teacher competency to work with diverse learners in the context of international education. *Teacher Education Quarterly, Winter 2007*, pp.70-94

Desimone, L.M. (2009) Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures. *Educational Researcher,* 38(3), pp181-199,

Farrell, T. S.C. (2015). It's not who you are! It's how you teach! Critical competencies associated with effective teaching. *RELC Journal 2015*, *46*(1), pp. 79 – 88, DOI: 10.1177/0033688214568096

Garfield, J. (1995). How students learn statistics. *International Statistical Review/Revue Internationale de Statistique.*63(1), pp. 25-34.

Hofstede, G. & Hofstede, J. (2005) Cultures and organization-software of the minds (2nd edition). New York:McGraw-Hill.

Huffman, J.B., Olivier, D. F., Wang, T., Chen, P., Hairon, S. & Pang, N. (2016). Global conceptualization of the professional learning community process: Transitioning from country perspectives to international commonalities. *International Journal of Leadership in Education, 19*(3), pp. 327-351, DOI: 10.1080/13603124.2015.1020343

Nguyen,P., Terlouw, C., & Pilot, A. (2006). Culturally appropriate pedagogy: the case of group learning in a Confucian Heritage Culture context. *Intercultural Education, 17 (1), March 2006, pp. 1–19.*

Teacher Quality Enhancement Center. (2010). Benefits of co-teaching [Web content]. St. Cloud, MN: St. Cloud State University. Retrieved from http://www.stcloudstate.edu/soe/coteaching/benefits.asp

Tishkovskaya, S. & Lancaster, G. A. (2012). Statistical Education in the 21st Century: a Review of Challenges, Teaching Innovations and Strategies for Reform. Journal of Statistics Education, 20(2), pp 1-56.

Tran, L. T. & Nguyen, N. T. (2015). Re-imagining teachers' identity and professionalism under the condition of international education, *Teachers and Teaching, 21*(8), pp. 958-973, DOI: 10.1080/13540602.2015.1005866

Villa, R. A., Thousand, J. S., & Nevin, A. I. (2008). *A guide to co-teaching: Practical tips for facilitating student learning* (2nd ed.). Thousand Oaks, CA: Corwin.

[i] Co-teaching is defined as two or more teachers working together with groups of students, sharing the planning, organization, delivery, and assessment of instruction as well as the physical classroom space (Teacher Quality Enhancement Center, 2010; Villa, Thousand, & Nevin, 2008).

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

WHO NEEDS STATISTICS?

KHAN, R. Nazim
School of Mathematics and Statistics M019,
The University of Western Australia,
Western Australia
nazim.khan@uwa.edu.au

The use of statistics as a tool has greatly increased over the last twenty years, mainly because computers and statistical packages have made the collection, storage and analysis of data more accessible. This has meant that researchers with a wide variety of statistical preparation and backgrounds are now using statistics. But what is the quality of the statistical work? And has the teaching and learning of statistics experienced a reciprocal increase? This research covers two aspects. First, we examine PhD thesis in Australian and New Zealand universities regarding the quality of the statistical analysis. PhD theses were selected as they are the stepping stone to research and innovation. The support available to research students at each university is also investigated. Analysis of data indicates which variables are associated with the correctness of statistical methodology. We also investigate the level of statistics covered in undergraduate science degrees in the same Australian and New Zealand universities. The results reveal a scenario much worse than perhaps anticipated. The findings of the study are relevant to Australia's future in a competitive global market.

Keywords: Applications of Statistics, Statistics in research, PhD theses, Statistical methodology, Statistics in Undergraduate, Statistical support.

INTRODUCTION

The last three decades has seen a revolution in the availability and power of computers. More importantly, modern operating systems have made computing much more accessible. Statistical software has also become widely available, easy to use, and cheap with free and open software such as R. Consequently, the use and importance of statistics as a tool in research has increased over the last two decades. Data is now easy to collect, store and process. Almost every discipline uses statistics to some extent. However, the traditional areas of statistical application such as the life sciences, agriculture, pharmaceuticals and medicine still dominate.

With the increase in the use of statistics, one would expect a corresponding increase in statistical knowledge and understanding. In particular, a commensurate increase in the teaching of statistics at the undergraduate level would be expected. The evidence for these two aspects can be judged from the quality of research output and in undergraduate programmes. Anecdotal evidence does not seem to support either of these two aspects. Experience in a statistics support centre for postgraduate students suggests that they are seriously lacking in the level of statistical knowledge required to conduct experimental design and analysis of the resulting data for their research projects. Similarly, experience with course design and degree structures shows that the number of statistics courses required, available for and taken by students is declining. In addition, the number of students in such courses is also declining.

Literature on the decline in statistics courses taken by students at universities is sparse. Most of the literature focuses on the teaching and learning of statistics and development and appropriateness of the curriculum. However, a review by The Statistical Society Australia Inc. (2005, p 11, Fig. 1) showed that between 1995 and 2002 the number of mathematics and statistics academic (non research-only) staff in universities fell by around 100. University funding models have contributed to this decline. In particular, school funding by effective full-time student units (EFTSU) encourages client schools to either remove statistics from their programmes, or teach some form of statistics themselves. In general such in-house statistics courses are taught by staff with low level knowledge of statistics. Such courses are often a tokenism to satisfy accreditation or avoid criticism. In addition, the tendency for faculties to no longer distribute funding to schools has resulted in staff losses in statistics not being replaced. Staff losses was also noted by Thomas

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

(2002), who commented that "…it defies common sense when other areas of science and technology which are dependent on advanced level mathematics are supported."

The motivation for this study was to address two questions.

1. What is the level and quality of statistics in research in Australia?
2. What is the level of statistics education in undergraduate degrees in Australia?

This paper is organised as follows. In the next section we discuss the methodology, followed by the results and conclusion.

METHOD

To assess the quality and level of statistics in research in Australia, we examined PhD theses from fifteen universities across Australia and New Zealand. New Zealand has a higher level of statistics in high school and was included to provide a comparison. The Australian universities included the G08 as well as other major institutions. Four major New Zealand institutions were included. The theses were selected at random from online repositories, filtered by whether they included statistics. Theses from science, medicine, social sciences and psychology were selected as these disciplines represent the major areas of applications of statistics. Preference was given to more recent theses. Each institution had several thousand theses available for selection, although the exact numbers were difficult to determine.

The availability and level of statistical support for postgraduate research students was obtained from the university websites. The sex of each student was also recorded. The level of statistical support provided directly to the student was obtained by examining the supervisors' research profile, and any statistical support acknowledged in the dissertation. Undergraduate science programmes were perused at each university and the number of statistics units required by each was determined. Science was selected in particular as it is the largest area for statistics application, and a science degree is a pre-requisite for a medical degree in many universities.

RESULTS

Six PhD theses from each of the fifteen universities were selected. Table 1 is the frequency distribution of the year of submission of the theses. Some institutions placed embargoes on recent theses, and this constrained the sampling. The statistical methodology in each thesis was examined and judged for correctness. The results are summarised in Table 2. Also shown in the table for each university are the availability of any support for research students, the level of such support, and whether statistics is a requirement in the undergraduate science programme. The availability of support was not considered when selecting theses. Note that the institutions have been de-identified. (However, details are available from the author.)

Table 1. Frequency distribution of year of submission of the theses.

Year	95	00	01	03	04	06	07	08	09	10	11	12	13	14	15	16
Count	1	1	2	3	1	1	4	1	1	7	4	3	3	7	19	32

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

Table 2. Summary results for the 15 universities: percentage of theses with correct statistical methodology, statistics support, and statistics courses in science undergraduate programme.

University	% correct (No.)	Support	No. of Staff	Ugrad Courses: Stats required
U1	67% (4)	Statistical Support Unit. Support for research students, including honours.	7	No
U2	17% (1)	Advice to HDR students and staff, short courses.	1	Some: Molecular Biosc., etc.
U3	33% (2)	Statistics Consulting Platform. Consultations for HDR students, staff.	2	Some: Agriculture, life sc
U4 NZ	50% (3)	Advice to HDR students and staff, short courses. Charges apply.	5	Yes. Mastery
U5	17% (1)	Statistical Consulting Service, 3 hours per semester per project.	1	No
U6	83% (5)	Some training available.	??	Some: Environ Sc.
U7	0% (0)	Data Management and Analysis Centre. Provides support to Faculty of Health Sciences, related to papers, grants and research students.	5	Yes: Agri etc.
U8 NZ	33% (2)	Statistical Consulting Centre. Short courses and support for research students.	8	Yes: some
U9	33% (2)	Statistics Consulting Centre, provides short courses and statistical advice.	9	Yes:Agric etc
U10	50% (3)	Research Training Statistical Consulting Service	2 hours/ week	Almost all science
U11	33% (2)	In school of medicine	8	Yes. BioInformatics etc.Most science. Some have up to third year units.
U12	50% (3)	Statistical Consulting Service	1	BioInformatics:Two math units at level 1, one stats unit. One further stats at higher level. Plant Sci: one stats unit.
U13	17% (1)	Centre for Applied Statistics. Short courses and statistical advice to research students.	7	One first level stats in some majors (marine biol, agriculture etc.)
U14 NZ	33% (2)	Advice to HDR students and staff. Short courses.	??	First level stats in All biol sciences, Forestry.
U15 NZ	100% (6)	Advice to HDR students and staff. Fee charged after 3 sessions.	1	ALL BSc requires 1 first level stats.

* The four New Zealand universities have been indicated (U4 NZ, U8 NZ, U14 NZ and U15 NZ).

New Zealand had 13/18 (54%) theses with correct statistics, while Australia had 24/66 (36%). Of the fifteen institutions, only one (institution U15 NZ) had all 6 theses with correct statistical methodology. One institution (U7) had none with correct methodology. Nine of the universities had less than half the theses with correct methodology. Overall, only 41% of theses had correct statistical methodology. The theses were from various disciplines, summarised in Table 3. Business has 100% and Ecology 83% theses with correct statistical methodology. No other discipline scores above 50%. Engineering, Environmental Science, Social Science and Medicine had lower proportions of theses that contained correct statistical methodology.

Table 3. Summary of number of theses by broad discipline. Note, Medicine included dentistry.

Discipline	No. of theses	No. Correct
Agriculture	10	2 (20%)
Biological Science	11	5 (45%)
Business	6	6 (100%)
Ecology	6	5 (83%)
Engineering	5	1 (20%)
Environmental Science	3	1 (33%)
Medicine	38	13 (34%)
Psychology	4	2 (50%)
Social Science	7	2 (29%)
Totals	**90**	**37 (41%)**

All the universities in the sample provided some level of statistical support for research students. Some institutions had university wide dedicated statistical support centres, while others had faculty-based support only. The mode of operation varied from unlimited free service, to limited free service with additional assistance attracting a fee, to only provided training. Almost all provided short courses. The number of staff varied from 1 to 9. In some institutions, support was through the Department of Statistics (or equivalent), and in these cases the school staff were involved in providing support. No clear dependence is obvious between of the number of staff in the support centre and the number of theses in error. In particular, the institution with 100% correct theses only has one staff, while one with eight staff only has 33% correct.

Almost none of the theses described the variables in detail. Similarly, almost none conducted any data exploration. Description of the methodology was sparse and incomplete. Statements such as "The data were analysed using Anova" or "A linear regression model was fitted" were common. There was no immediate indication of which variables were involved and which was the response variable. None of the theses showed evidence of any model diagnostics. Typical errors are listed in Table 4. The most common errors were: fitting simple linear regression to non-linear data; fitting several pair-wise linear regressions; conducting several one-way Anova; ignoring correlations in repeated measures or time series data; and fitting Anova to count data. Figure 1 exhibits some typical data and analyses. The contexts and variables have been de-identified.

Table 4. Typical errors in statistical methodology.

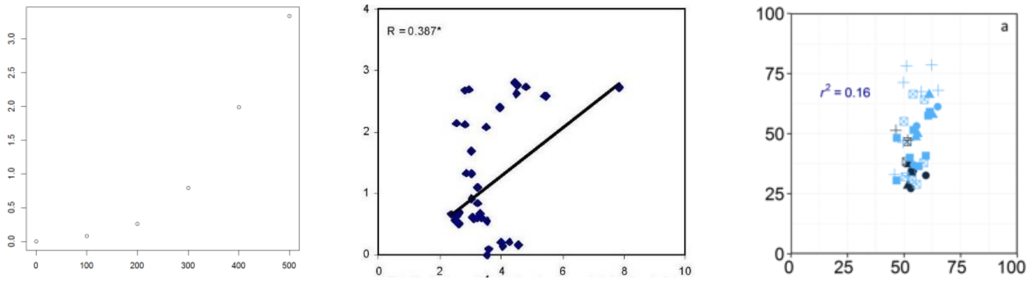
1. Fitting simple linear regression to data that is non-linear.
2. Performing several independent t-tests or Anova at each time point and between time points in a repeated measures experiment, separately for each covariate.
3. Performing several independent t-tests instead of Anova.
4. Performing several simple linear regressions and Anovas instead of a linear model.
5. Ignoring dependence and seasonality in time series data.
6. Fitting linear regression models to time series data, ignoring correlation and seasonality.
7. Several chi squares tests based on two way contingency tables for categorical variables and linear regression for continuous variables, instead of a linear model.
8. Fitting linear regression models where the explanatory variable is a factor.
9. Inference for binomial data based on examining percentages.
10. Testing each variable for normality before fitting any models.
11. Fitting simple linear models for data that has clear groupings.
12. Simple linear regression for data with heteroscedasticity.
13. Repeated measures analysed by examining graphs over time. Covariates ignored.
14. Survey data analysed item-wise only.
15. Only qualitative analysis without fitting any models.
16. Lifetime data analysed as Anova.
17. SE of mean $= s/\sqrt{n-1}$.
18. Survival data analysed based on Kaplan-Meier estimates only, ignoring covariates.
19. Anova for count data.

Table 5. Summary of results by statistical support provided to student.

	Statistical Support		
	No	Yes	
Correct	11 (12.2%)	26 (28.9%)	37 (41.1%)
Incorrect	47 (52.2%)	6 (6.7%)	53 (58.9%)
	54 (64.4%)	32 (35.6%)	90

As described earlier, we further investigated if statistical support was provided directly to the student by supervisors or other sources such as statistically qualified staff. The results are summarised in Table 5. Only 36% of students received specialised statistical support. It appears that students who received such support are much more likely to have correct statistical methodology in their theses.

Exhibit 1 "Regression analysis indicated that the (relationship) was linear." A plot of the data (a) revealed a non-linear relationship.
Exhibit 2 "These data best fitted a linear regression demonstrating a significant linear relationship at the 0.05 level of probability (R=0.387)." Oblivious to the effect of the point of high leverage (b).
Exhibit 3 Response as a function of covariate (c). The symbols represent genotypes. The analysis ignores the group structure in the data, perhaps concealing any patterns or relationships.



(a) (b) (c)

Exhibit 4 Seasonal data over 16 months, of temperature measurements at several sites for several species, during periods of the day (morning, midday, afternoon). Analysis:

- One-way Anovas for differences in response at different times of day within site, within species, between months.
- Simple linear regression between response and a covariate, separately for month, for two levels of a factor separately, with data at period of day averaged over the month.

Figure 1. Examples of incorrect analyses.

A logistic regression model was fitted to the data, with the binary variable *correct* (Y or N) indicating if the statistical methodology was correct, and covariates *Year*, *University*, *Discipline*, *Sex* and *Country*. Two further variables that indicated the level of statistics support were also included: *Support* (Y, Some, N) indicating if statistical support was available to postgraduate students at an institutional level, and *SSupport* (Y, N), indicating if the student had direct statistical support from supervisors or other sources. The model indicated that the thesis was more likely to have correct statistical methodology if statistical support from supervisors or other sources (usually statisticians) was provided (log OR = 3.6, p-value $= 3.47 \times 10^{-6}$). Further, compared with other disciplines Ecology theses had a higher probability of containing correct statistical methodology (log OR = 3.1, p-value= 0.0361). No other variables had any significant effect. We acknowledge that the sample size of six theses from each institution is small and could show wide variation, given that several thousand theses are available from each. Nonetheless, the results are unequivocal, and are supported by the qualitative assessment of the theses.

The second aspect considered was the statistics component in undergraduate science courses. Some institutions have general overriding rules for the science degree, while others have discipline-specific requirements for statistics. Institution U15 NZ required a level 1 statistics course in all science degrees. Interestingly, this institution had 100% theses with correct statistics. Another institution, U4 NZ, also required level 1 statistics in all science degrees, including mastery level assessments. Institution U10 required level 1 statistics in almost all science degrees. Most other institutions required a level 1 statistics unit in selected majors. Two institutions (U11 and U12) required higher level statistics units, while two others (U1 and U5) did not require any statistics in their undergraduate science degree. Compare this with an Indian university, The University of Punjab, Chandigarh (2014). The BSc in Agriculture includes two units of mathematics (one in each semester in first year), and two statistics units including univariate statistical methods and design of experiments. While this is only one institution, it compares favourably with the best Australian institutions, but is ranked only 273 in India (http://www.4icu.org/in/).

CONCLUSION

Statistics is an all pervasive discipline, with applications burgeoning over the last two decades as technology makes data ever easier to obtain, store and analyse. However, the knowledge and understanding of statistical ideas is still very poor, as evidenced in PhD theses from several universities in Australia and New Zealand. Data analysis reveals that the key determinant for correct statistical methodology in a thesis is the provision of direct specialist statistical support to the student. Interestingly, the institution with all correct theses also required a level 1 statistics unit in all science degrees, but had only one staff member in its statistical support unit and only half the theses had received direct statistical support. It seems that this institution has a culture of nurturing statistical literacy amongst its staff and graduates.

Level 1 statistics courses typically cover basic statistical idea, including data exploration, probability, t-tests, Anova, simple linear regression and chi squared tests. This does not adequately prepare graduates for PhD research. The techniques of experimental design, repeated measures, generalised linear models, and perhaps mixed linear models are typically required for research in many disciplines. However, these topics are typically covered in level 2 and 3 statistics courses.

It is imperative that universities include at least up to level 2 statistics courses in undergraduate degrees in disciplines that require statistics. The alternative is for all graduate research students to take at least two level 2 statistics units during their tenure. In addition, a supervisor with a strong statistics background should be included, especially if the research involves higher level statistics such as time series, linear mixed models or multivariate statistical techniques. The level of research is greatly enhanced by use of the correct statistical techniques. Using "bad" statistics severely compromises the quality and results of the research. In most cases the results are at best doubtful and at worse completely incorrect.

Statisticians should be involved from the beginning of projects that have a large statistical component. Data analysis begins before that data is collected! University funding issues and "guarding turf" should not be at the cost of compromising quality. Co-operation cannot be expected in a paradigm of competition for funds and research scores. A cultural change is required to address these issues. It takes only a small change in thinking to produce large results and outstanding quality. Innovation and entrepreneurship cannot be established on false foundations. Our near neighbours may already be ahead of us in the higher education game, as evidenced by international rankings. If no change is made, the illusion of our superiority will soon be replaced by the reality of our mediocrity. The effects on our global competitiveness are obvious.

ACKNOWLEDGEMENTS

The author thanks two anonymous referees whose suggestions greatly improved this paper.

REFERENCES

Statistics Society of Australia Inc. (2005) Statistics at Australian Universities: An SSAI sponsored review. http://www.statsoc.org.au/wp-content/uploads/2013/06/ReviewofStatsFinalReport.pdf (Retrieved 5 October 2016)

Thomas, J. (2002). Mathematical Sciences in Australia: Still Looking for a Future, Briefing paper, *Aust. Math. Soc.*, http://www.austms.org.au/AustMath/braindrain_2002.pdf (Retrieved 5 October 2016)

The University of Punjab, Chandigarh, BSc. Agriculture syllabus. (2014) http://puchd.ac.in/includes/syllabus/2014/20140924163522-b.sc.-four-year-agriculture-semester-system-2014-2015.pdf?201607425608 (Retrieved 5 October 2016)

PREPARING STATISTICAL CONSULTANTS

MEYER, Denny[1], EARNEST, Arul[2], MCKENZIE, Dean[3], BHOWMIK Jahar[1], CRITCHLEY Christine[1], HILLER, Janet[1], QUINN, Stephen[1], GRIGO, Jennifer[1], PHILLIPS, Brian[1]
[1]Swinburne University of Technology, [2]Monash University, [3]Epworth HealthCare
dmeyer@swin.edu.au

Statistical consultants often play the role of collaborators on grants and tenders. This means that they need to be trained in much more than just statistics. In particular, statistical consultants need to have experience with the processes for grant and tender applications, collaboration and reporting. However, unlike other researchers, statistical consultants also have to work with many individual clients, and juggle numerous projects often from very different areas and disciplines. Thus statistical consultants need a diverse knowledge of statistical methodologies, software, data and disciplines or content areas. These complexities have been recognised in a new statistical consulting subject within Swinburne University's post graduate MSc (Applied Statistics). This challenge has been met with a mix of lectures presented predominantly by applied statisticians specialising in a variety of important areas (such as health, social studies, psychology, marketing and management), routinely working with data from sources such as medical registries, patient databases, online open access databases, telephone interview data and government survey/census data. Assessments involved both playing the client and responding to a client's brief, while the exam included a tender application. This paper describes this subject as well as the feedback received from a small class of mostly online students after the subject first being offered.

INTRODUCTION

It was decided in 2013 that a Statistical Consulting subject should be included when the Swinburne MSc (Applied Statistics) course grew from 12 to 16 subjects in order to meet new Australian Qualification Framework requirements (AQF Council, 2013). This master's is a blended program, offered simultaneously in both on-campus and online mode to Swinburne students, some of whom are enrolled through Open Universities of Australia. The nested program aims to provide students with a comprehensive toolbox in terms of statistical concepts, methods and software. The four initial Graduate Certificate subjects assume only the most elementary of statistical backgrounds, introducing basic statistical concepts and techniques. One of the subjects provides an in-depth introduction to the IBM - SPSS Statistics software (SPSS) and another provides a first course in research design, including both questionnaire and experimental design. The remaining two subjects provide foundation subjects in univariate and bivariate statistical methods, stressing application and interpretation rather than mathematical theory. The extra four subjects needed to obtain the graduate diploma build on this foundation with a Multivariate Statistics subject and a SAS statistical software subject, with SAS also used in the Forecasting and Sample Surveys subjects in order to allow the analysis of more complex data structures. There are eight further subjects required for the master's level. Four cover scale development and structural equation modelling, data mining/visualisation methods and advanced regression methods, including mixed models and multi-level modelling. Two subjects cover maximum likelihood and Bayesian analyses using R, while the final two linked subjects are a statistical consulting subject and a project subject which allows for an industry placement.

What makes the program particularly interesting is the wide range of degrees and working experience of our students, making it essential that we cater for a mix of interests and examples in all subjects. The online nature of the program and the mixture of student backgrounds made the content of the Statistical Consulting subject a special challenge. We referred to the literature in order to find out what statistical consultants usually do, what their clients perceived as their strengths and weaknesses, and how others had taught such subjects in the past. A review of what statistical consultants actually do and our own experience suggested that what is commonly included in courses for statistical consultants is not sufficient. In particular, protocol, proposal, tender and grant writing skills are needed for statistical consultants involved with team collaborations. This paper describes how we developed a statistical consulting subject suitable for online/on-campus (blended) teaching,

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

with a mixed cohort of students, addressing the need for collaboration with both individual clients and research teams.

What do statistical consultants do?

In 2006 Gullion and Berman published an article in the American Statistician describing what statistical consultants do. Their survey was approved for use with the members of the Australian Statistical Association (ASA), and more than one-third of the ASA members responded to the survey in 2000. Of the 485 member respondents 421 were currently engaged in paid consulting. The highest qualification for these consultants indicated 65% for PhD's, 31% for Masters and 4% for a bachelor's degree. Of these statistical consultants, 226 were salaried but not self-employed, 95 were only self-employed and 100 were both salaried and self-employed. The most common employer for the 326 salaried consultants was an educational institution (45%), with lower numbers for medical/health care centres (11%), government agencies (8%) and pharmaceutical companies (7%), and the remaining 29% employed by companies specialising in manufacturing, business/finance, contract research, statistical consulting or other research or consultancy. These high levels of consulting activity confirm that a statistical consulting subject is desirable in any Applied Statistics master's program. The Gullion and Berman (2006) survey found that the work of self-employed consultants differed markedly from salaried consultants as illustrated in Table 1.

Table 1: Activities of Salaried and Self-Employed Consultant Roles

Activities (* p<.05)	% Responses as Salaried Consultants (N=326)	% Responses as Self-Employed Consultants (N=191)
Data Analysis	98	94
Interpreting Results	90	74
Planning Analysis	89	71
Designing Studies	73	53
Reviewing Research Protocols	53	35
Writing Reports for Clients	69	64
Designing Questionnaires/Tests	51	32
Survey Design	43	28
Power Analysis	66	44
Writing for Publication	66	33
Performing Statistical Research	50	28
Auditing Data	29	20
Evaluating Protocol Performance	19	8

From Table 1 we see that consultants with a salaried role are more heavily involved in the complete research process, including the designing of studies and writing of research proposals and reports, while consultants with a self-employed role focus mainly on data analysis and to a lesser extent on planning analyses and interpreting results. Careful consideration of these topics in relation to the content of our master's program suggested that our students required additional experience with the designing of studies, reviewing of research protocols, writing research publications and evaluating protocol performance. These activities were therefore ear-marked for inclusion in our new Statistical Consulting subject.

How Do Clients Evaluate the Services of Statistical Consultants?

Johnson and Warner (2004) conducted a survey of 129 clients at two American universities in 2000 and 2001 in order to determine how useful clients found their statistical consultants. The perceived level of statistical expertise of the clients was relatively low, with a mean score of 5.18 on an expertise scale increasing from zero to ten. On a 5-point scale the consultants scored above four on the seven performance measures described below, suggesting a high level of satisfaction overall. A factor analysis of the seven items suggested two factors. After a *varimax rotation* the first factor extracted was called *Stats Collaborator* because it loaded strongly on the items "The consultant understood my problem", "The consultant was competent in statistics", "I was able to understand the statistical advice presented to me" and "Overall the consultant provided a direction or solution to my problem". The second factor was labelled *Personality* because it loaded strongly on the items; "The consultant was interested in my research problem", "The consultant was personable and easy to talk to", and "The consultant was enthusiastic". An ordinal logistic regression based on these two factors was used to predict responses to the question "Overall, I was pleased with the consulting services provided to me". After controlling for the university and whether or not the client requested the services of their particular consultant, it was found that higher scores for both the *Stats Collaborator* factor and the *Personality* factor improved the likelihood of a positive response significantly.

These results suggest that it is not enough for a statistical consultant to just be competent in statistical methods. In addition, consultants need to provide an understandable solution for clients and show a genuine interest and enthusiasm for a client's problems (Lynn, 2016). This suggests that any subject on Statistical Consulting also needs to provide tuition on communication. Without good communication the consulting relationship cannot prosper (Boen and Zahn, 1982) and, as Cabrera and McDougall (2002) emphasise, there is a danger that the wrong problem will be solved, sometimes described as "a Type 3 error". Earnest (2016) stresses the need for good scientific written communication, using frameworks such as the CONSORT (Consolidated Standards of Reporting Trials) and STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statements as a guide. (See http://www.consort-statement.org/, http://www.spirit-statement.org/ and http://www.strobe-statement.org/). Another important aspect of communication dealt with by Earnest (2016) involves the management of client expectations.

What is taught in other Statistical Consulting subjects and books?

The literature suggests that, in comparison to the mix of consulting environments likely for our student cohort, statistical consulting subjects are often designed for a specific discipline. For example, Newton and Rudestam (2013) target only consulting for the social sciences while Ader, Mellenbergh and Hand (2008) focus on the social and behavioural sciences in addition to medicine and epidemiology. However, Hand and Everitt (2008) widen the net to also include finance, marketing and many areas in science and industry. As explained by Mann et al. (1999) understanding "the what and why driving a client's request for service" is critical, so a mix of consulting environments poses a special challenge for statistical consultants. Earnest (2016) stresses the need for statistical consultants to help the client work out the aims and hypotheses for their projects and also to help the client prepare clean and coded datasets, often de-identified in order to ensure data privacy and security. A good understanding of the research context is needed in order to make this possible.

Most courses take a very practical approach, recommending data driven learning (Gibbons and MacGillivray, 2014). Newton and Rudestam (2013) address more than 100 questions that a consultant might be expected to answer. These questions were collected from students and colleagues. Most courses (e.g. Ader et al., 2008) dwell on design questions and general methodological problems, such as missing data and data quality. However, the data analysis itself is kept at a somewhat basic level. Dissemination and publishing are mentioned but usually only in the context of journal articles or conference presentations.

But how is the issue of communication addressed? Joiner (2005) suggests teaching students to "listen carefully and ask probing questions". Rustagi and Wolfe (1982) and Taplin (2007) have stressed the importance of role playing when training statistical consultants, in order to develop empathy with clients. Joiner (2005) recommends that consulting sessions be recorded so that they can be reviewed with the Statistical Consulting class, suggesting that watching how an experienced consultant communicates is the best way to teach this skill. Sharples, Yeend, Francis, Ridall and Booth (2010) also recommend this observational approach for a post-graduate class, whereby a student shadows a member of the academic staff during consultancy meetings. Alternatively, they recommend that students learn by doing, take varying levels of ownership of a client and their problems, with staff merely providing guidance. Derr (2000) argues that clients as well as experienced statisticians, must work together in order to train statistical consultants.

What did we do?

The review of existing statistical consulting subjects and books seemed to focus on the relationships between consultants and individual clients. Clearly this is important, however, as indicated by what statistical consultants actually do in practice, team collaboration also needs to be addressed. In addition, as described in the introduction, the special nature of our program and our students also impacted on the priorities for our consulting subject. In summary therefore, the new subject had to fit well with the rest of the subjects in our master's program and this meant following a similar teaching and assessment approach - an approach which we knew from experience should work well within a blended teaching environment. Secondly, we wanted to be sure that the subject would be relevant and of interest to all our students regardless of their particular background or discipline. Thirdly, we wanted to ensure that this subject would provide a helpful initiation for the associated project subject. Finally, we wanted to ensure that this subject would prepare our students for working with individual clients and that, in addition, our students would be able to pull their weight in team research projects, contributing fully to the administrative responsibilities of the team in terms of ethics applications, writing research proposals and grant applications, designing clinical trials and writing journal articles. We therefore chose the following aim and learning outcomes for our new Statistical Consulting subject.

Aim: " to assist students develop effective consulting strategies and skills for dealing with clients who ask for statistical help, by using case studies and a range of real clients with real problems".
Learning Outcomes(LO): After our students successfully complete the subject they will be able to:-
LO1. Formulate research questions and hypotheses in a specific consulting context
LO2. Distinguish between different data collection strategies and research designs in a specific consulting context
LO3. Appraise clients in regard to appropriate data analysis using a range of statistical procedures
LO4. Support the client with the interpretation of the results of the analysis
LO5. Formulate directions for further investigations by the client
LO6. Compose findings and conclusions using presentations and written reports
LO7. Contribute to team projects with research proposals, ethics applications and journal articles.

Figure 1. Learning Aims and Outcomes

An important aspect of our master's subjects is the standardised nature of the assessment across all subjects. Use of weekly quizzes facilitates engagement, especially for the online students, as do the two assignments, one in the middle of the semester and one at the end. In our master's subjects one of the assignments typically incorporates a presentation. Finally, there is an exam as well as a practice exam. It was decided that this same assessment structure would be followed for the Statistical Consulting subject. Our subjects all have a comprehensive set of notes and/or good text books so, starting from scratch, a suitable text seemed the way to go. Several texts were considered before Cabrera and McDougall (2002) was chosen, largely because it uses case studies with real data to get its message across. In particular, there are case study exercises which were very useful for the preparing of some of the early quizzes. The majority of our students study online, which means that it is not possible for them to learn how to communicate with a client by shadowing an experienced consultant. However, in recorded lectures, the lecturers described their consulting experiences with some candour. In addition, a limited form of role play was used in the first assignment which was designed to give the students some empathy for the client as they played the role of client themselves. In their first assignment, addressing LO1-LO3, students were required to produce a client brief and, in a recorded online Collaborate session, they had to describe to the rest of the class the problem they wanted solved. This involved providing the context for the problem and their requirements, also describing the data that had already been collected and some suggestions regarding their expectations.

Table 2: Content and Assessment for Module 1 of our Statistical Consulting subject

Week	LO	Lecture Topic	Brief Summary
1	1	Introduction	Types of Consulting Environments
2	4-6	Communication	Client Interactions
3	2	Data Collection	Observational, Survey, Experimental
4	3	Statistical methodology	Statistical Analysis Plan (inc.Trans, para and no-para tesing)
5	1-6	A Consulting Project	1st Meeting, Research Questions, Analysis, Present Results
6	1-6	Student Presentations	Client brief: research questions + data

The subject was split into two modules with the first dealing with the key components of a consulting relationship as illustrated in Table 2, while addressing LO1-LO6. As shown in Table 3, in Module 2 a new statistical method was introduced in each week, evidencing the need for statistical consultants to continually upskill while addressing LO3-6 at an advanced level. Additionally, targeting LO7, one important research collaboration requirement was addressed in each week (e.g., writing ethics applications, research protocols, tender applications, research proposals and journal articles). Finally, while targeting LO2 at an advanced level, an important new source of data was introduced in each week of Module 2 (e.g., systematic reviews, ABS data, medical subsidy data, registry data, telephone interview data and open access online data). In module 2, a different lecturer took each class, bringing a new perspective on statistical consulting and a different experience as a statistical consultant to each week of lectures. We are fortunate in having staff with diverse backgrounds and interests. This allowed us to address the mix of student interests in our class, providing real examples from a broad variety of areas and disciplines.

Table 3: Content and Assessment for Module 2 of our Statistical Consulting subject

Week	LO	Lecture Topic	Research Requirement	Data Introduced
7	1-7	Sys. reviews	Ethics approval	Relevant Literature
8	1-7	Observational studies/maps	STROBE guidelines (in 2017) and QGIS software demonstrated	ABS data
9	1-7	Time Series models	Writing research proposals	MBS/PBS subsidy data
10	1-7	Latent Class Analysis	Writing tender applications -Public opinion research on genetic biobanking	Telephonic Interview (CATI) data
11	1-7	Hospital Consulting	CONSORT and SPIRIT guidelines for clinical trials	Registry data/ clinical data
12	1-7	Data Mining	Writing a research article – predicting suicide ideation	Open access online data
13	1-7	Practice exam	Short grant application + Multiple Choice/short questions	Solution Provided
14	1-7	Final Exam	Short tender application +Multiple Choice/short questions	Open Book Exam

In Module 2 the weekly online quizzes related directly to the content of each lecture. For the second assignment in 2016, the students simply exchanged their assignment 1 client briefs and data, providing a written report for the client summarising results and conclusions. Instead, in future years, the students will be asked to prepare proposals for their own project/placement, scheduled for the following semester. A special marking rubric linked to the learning outcomes will be provided for the marking of this second assignment. The weekly quizzes address LO1-LO4 and the second assignment addresses LO1-LO7. The Practice and Final Exams also cover LO1-LO6 with LO7 covered by way of a grant application and a tender application respectively. In 2016 the grant application was directed at the Royal Australian College of General Practitioners (RACGP) Foundation, and the tender application was directed at a government agency. Both these applications covered real but shortened projects and the questions were designed to stress the methodology and statistical aspects of the applications. A solution was provided with the practice exam to assist the students with their preparation for the final open book exam. Much of what is covered in this subject will assist the students in their ensuing projects and placements, allowing a second opportunity to examine what has been learnt in the Statistical Consulting subject using the project performance of the students.

Student and Advisory Committee Feedback

Our advisory committee encouraged the introduction of this consulting subject, in particular stressing the importance of communication skills for our students. The initial feedback from students suggests that they enjoyed the first offering of this subject: “Great subject”, “Privileged to receive lectures from the amazing lecturers”, “Hearing about the professional history of lecturers incredibly valuable”. In particular, it seems that a mix of lecturers with a range of backgrounds and interests was a good decision. It has meant that we were able to provide a variety of perspectives and interests that reflect the mix of our students' backgrounds. At the same time, the use of our standard blended teaching format and assessment processes has meant that both online and on-campus students have been well catered for. From this small group of seven students we scored 10/10 on the student feedback survey (SFS) for this subject, but the response rate was (as usual) very low at 29%. While this is encouraging it is far from conclusive and much more work is needed in order to measure the success of this subject.

Conclusion

This paper describes our current thinking on how a master's level statistical consulting subject should be structured for our program. Our goal is to prepare our students for work as a statistical consultant in a future where research team collaborations are as much of a reality as individual client consultations. In addition, we hope that this subject will better prepare our students for their projects and placements in the following semester. As suggested by one of the referees, in the future we need to better measure the knowledge gain of our students, by testing for improvement in their research proposal writing skills and achievement of the other learning objectives. However, a good start has been made.

References

Ader, H.J., Mellenbergh, G.J. & Hand, D.J. (2008). *Advising on Research Methods: A Consultant's Companion.* Huizen, The Netherlands: Johannes van Kessel Publishing.

AQF Council (2013). *Australian Qualifications Framework*. Second edition.

Boen, J.R. & Zahn, D.A. (1982). *The Human Side of Statistical Consulting*. Belmont, California: Lifetime Learning Publications.

Cabrera, J. & McDougall, A. (2002). *Statistical Consulting*. New York: Springer-Verlag.

Derr J. (2000). *Statistical Consulting: A Guide to Effective Communication.* Pacific Grove, California: Duxbury Press.

Earnest, A. (2016) *Essentials of a Successful Biostatistical Collaboration*. Boca Raton, Florida: CRC Press, Taylor & Francis Group.

Gibbons, K.S. & MacGillivray H. (2014). *Education for a Workplace Statistician*. Topics from Australian Conferences on Teaching Statistics: OZCOTS 2008-2012. Eds: H. MacGillivray, M.A. Martin and B. Phillips. New York: Springer Proceedings in Mathematics and Statistics.

Gullion, C.M. & Berman, N. (2006). What statistical consultants do: Report of a survey. *The American Statistician*, 60(2), pp. 130-138.

Hand, D.J. & Everitt, B.S. (2008). *The Statistical Consultant in Action.* New York: Cambridge University Press.

Johnson, H.D. & Warner, D.A. (2004). Factors relating to the degree to which statistical consulting clients deem their consulting experience to be a success. *The American Statistician*, 58(4), pp. 280-289.

Joiner, B. (2005). *Statistical Consulting.* In S.Kotz, N. Balakrishnan, C. Read & D. Vidakovic (Eds.). Encyclopaedia of Statistical Science (2nd. Ed.) New York: Wiley.

Lynn, H.S. (2016). Training the next generation of statisticians: from head to heart. *The American Statistician*, 70(2), pp. 149-151.

Mann, B.L., Quinn, L., Boardman, T., Bishop, T. & Gaydos, B. (1999). What my mother never told me: Learning the hard way. *The Statistical Consultant.* 16(3), 2-9.

Newton, R.R. & Rudestam, K.E. (2013). *Your Statistical Consultant: Answers To Your Data Analysis Questions*. Second edition. Thousand Oaks, California: Sage Publications.

Rustagi, J.S. & Wolfe, D.A. (1982). *Teaching of Statistics and Statistical Consulting*. New York: Academic Press.

Sharples, S., Yeend, E., Francis, B., Ridall, G. & Booth, J. (2010). *Developing Statistical Consultancy Skills in Post-Graduate Students; a Case Study*. OZCOTS Proceedings.

Taplin, R. (2007). Enhancing statistical education by using role-players, *Journal of the Royal Statistical Society, Series A*, 170, Part 2, pp. 267-300.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

EVOLUTION OF TEACHING STRATEGIES IN A FRENCH ODL UNIVERSITY COURSE

AVALOS Marta[1,2,3], LE GOFF Mélanie[1,2], JOLY Pierre[1,2],
JUTAND Marthe-Aline[1], ALIOUM Ahmadou[1,2]
1 - University of Bordeaux, Bordeaux School of Public Health, France
2 - INSERM U1219 Bordeaux Population Health INSERM Research Center, France
3 - INRIA Research Centre Bordeaux Sud-Ouest, SISTM team
marta.avalos-fernandez@u-bordeaux.fr

The university course on statistical methods in health at the Bordeaux School of Public Health, University of Bordeaux, has been run as an Open and Distance Learning (ODL) program since 2004 on the basics of statistical reasoning in the health field. The course is mainly for professionals. In more than ten years, about 1,000 people have been trained with over a third coming from sub-Saharan Africa. The program aims to meet a growing demand for statistical training from professionals from the south whose mobility is limited. Each year a satisfaction survey is sent to students with a view to improving the program. Even though participation in the survey is anonymous and not compulsory, it is a valuable source of comments and ideas. These have led to innovative pedagogical practices such as "tutored exercises" with individual correction, the use of new statistical software, summary sheets and flipped classrooms. However, benchmarking of the program has shown that more could be done. Teaching strategies should evolve within the framework of distance learning in terms of content, form and interactivity. This article discusses the development of these new educational strategies from their inception as well as future projects.

CONTEXT

The Bordeaux School of Public Health at the University of Bordeaux (France) set up online "university diplomas" in 2001 in response to a request for proper education, training and adequate certification in public health from a French-speaking audience that could not be present for face-to-face training for geographical or professional reasons. University diplomas are institution-specific degrees presented as single subject courses that reflect the strengths of individual universities and offer students opportunities to gain university-level education in specific fields. The Open and Distance Learning (ODL) program was born out of political will, the commitment of the teaching staff and thanks to financial opportunities that, among other things, allowed the establishment of an information technology department that manages the learning platform.

Currently, ten online French university courses are available (http://ead.isped.u-bordeaux2.fr). University courses in Spanish are also proposed. Among the online university courses, two concern the biostatistics track (*Statistical methods in health* and *Statistical regression methods in epidemiology*) and four the epidemiology track. Indeed, epidemiology and biostatistics are one of the strengths of the Bordeaux School of Public Health, which is recognized internationally. The courses are run entirely online except for the final session exam, which is conducted in accredited centres around the world such as universities, French embassies, French high schools or centres within the University Agency for French-speaking communities (AUF). In addition, the online master's program in Public Health at the University of Bordeaux became effective in fall 2007. The master's degree is a more advanced program that has exam sessions but also two-week face-to-face courses organised in a few accredited centres (currently the University of Bordeaux in France, the University of Abomey-Calavi in Benin, and the Institut Pasteur in French New Caledonia in the Pacific Ocean). The master's degree is offered online- and campus-based while only an online program is proposed for the ten university diplomas.

The minimum admission requirements for the university course in *Statistical methods in health* are an undergraduate degree or at least five years of relevant work experience in the health sector. The university course *Statistical regression methods in epidemiology* and the master's degree are more selective with specific admission requirements. The subject courses programs are mostly included in the educational program of the Masters in Public Health. Thus, validating the single subject courses implies automatic validation of statistical or epidemiological courses in the interdisciplinary first year of the Masters.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

STRUCTURE, FUNCTIONING, AUDIENCE

The university course on statistical methods in health at the Bordeaux School of Public Health, University of Bordeaux aims to teach the basics in statistical reasoning and methods in the health field. The course, which has been run since 2004, allows students to develop their ability to conduct data analysis as end-users.

The teaching period runs from mid-October to late May. The program involves 100 hours of teaching, plus personal study/work time (exercises, graded and non-graded assignments, and review for the final exam). This corresponds to a weekly work load of 4 to 5 hours. The course contains four modules:

- Descriptive statistics, introduction to general terminology
- Introduction to the fundamental concepts: random variable, basic discrete and continuous probability laws, population, sample and sampling fluctuation concepts, sampling distributions, central limit theorem, point estimation, confidence interval
- General principles of statistical hypothesis testing, one and two sample models, main parametric and non-parametric tests
- Introduction to simple linear regression and one-factor ANOVA

Each module is structured in three parts. First, the course introduces concepts, which are illustrated by several examples from health studies. Definitions and calculation methods are introduced as the need arises in a convenient, self-contained and intuitive way. Second, written exercises are proposed. Finally, lab exercises can be solved using software. Introductions to recommended statistical software and MS Excel spreadsheets are also provided.

The course schedule including the online availability of documents, theoretical courses, exercise corrections, assignment corrections and deadlines for graded and non-graded assignments, past years' exams etc., is announced at the beginning of the academic year. News and updates are posted using the news column or are distributed by email. The Bordeaux School of Public Health online degrees are offered using an in-house customised learning management system called Plei@de that is an intuitive and ergonomic platform. Like other learning platforms, for example Moodle or Dokeos, it provides technical support for exchange between teachers and students: repository for course material or web pages, upload of assignments, data files, quizzes, forums, technical assistance, space to share or store files, online resources and tools, student marks, etc. Plei@de offers a high degree of fluidity in the interaction between students and teachers. The use is simpler than Moodle in that sense that it offers, in the same space, educational, technical and administrative information. The platform also provides user statistics: number of visits, number of finished exercises and corresponding scores, downloaded documents, uploaded assignments, date of last access, etc. This can be used as a tool to assess the availability, motivation and serious-mindedness of students.

The final grade for a course is determined by a weighted average of the marks students receive on all of their course assignments and exams. To pass the course, students must have a weighted average of 50% or better on the exam and the graded assignments. There are two planned assignments (one for the first two modules and another for the last two modules) that are weighted equally and a final exam that accounts for two thirds of the grade. The final exam is taken in specific centres all around the world, where the presence of the student is required to ascertain his/her identity. After a decade, more than nine hundred students have participated, the number of students per year varying between 60 and 100. More than one-half of students are from Africa, the most numerous coming from sub-Saharan Africa (taken as the country of residence). Approximately a third of the students come from Metropolitan France (the part of France situated in Europe). In fact, a partnership has been established with the AUF which finances several students per year. Between 45 and 90% of the students per year pass the course, and the annual dropout rate (taken as the number of students absent for the final exam) varies between 12 and 30%. Table 1 presents differences in dropout rates and pass rates according to the region of residence for the last three years. Dropout rates are highly variable from year to year and intra-group variability is higher than intergroup variability. Metropolitan France residents group showed the highest pass rates; however the potential for confounding is obvious.

Table 1: Annual dropout rate [1], taken as the number of students absent for the final exam to N, the total of participants, and annual pass rate, taken as the number of students having a weighted average of 50% or better on the exam and the graded assignments to n, the number of participants present for the final exam.

	2013-2014		2014-2015		2015-2016	
Residence	Dropout rate (N)	Pass rate (n)	Dropout rate (N)	Pass rate (n)	Dropout rate (N)	Pass rate (n)
Sub-Saharan Africa	8% (49)	56% (45)	24% (34)	65% (26)	21% (38)	50% (30)
Metropolitan France	26% (38)	75% (28)	20% (30)	92% (24)	12% (42)	81% (37)
Other regions	13% (15)	62% (13)	11% (19)	71% (17)	47% (15)	63% (8)
All	16% (102)	63% (86)	19% (83)	76% (67)	21% (95)	67% (75)

[1]Rates are computed for participants coming from sub-Saharan Africa, Metropolitan France, and other regions for the last three years (course in *Statistical methods in health*, Bordeaux School of Public Health, University of Bordeaux, France, 2013-2016).

The vast majority of the students are in the health sector. For instance, in 2014-2015, only 13% of the students came from non-medical professions, 53% came from medicine and the remaining percentage came from dentistry, midwifery, pharmacy, nursing or associated health professions. Some of them are already experienced health professionals while others are still studying and not yet qualified.

INSTRUCTIONAL DESIGN, EVOLUTION

Once the purely online approach was adopted, the pedagogical thinking had to change entirely. We could not limit ourselves to downloadable educational content but had to explore the best possible ways of supervising or advising students throughout their gradual learning process. We had to take into account the traditional constraints associated with e-learning student profiles: e.g. professionally active adults who want to acquire an advanced degree or new professional skills but do not have time to attend regular classes. In addition, constraints specific to the targeted audience had to be considered, i.e. professionals in the public health sector located in different countries worldwide with variable internet connection quality and living in different time zones. We also realised we needed to develop interactive web pages. Animations could be used to reinforce conceptual understanding by illustrating particular points and active learning that made the course dynamic. However, downloadable documents (handouts, exercises, etc.) were required because of unequal access to Internet.

A flexible schedule, freedom of action, reduced indirect training costs and a cooperative learning environment are all advantages of online teaching compared to traditional face-to-face classes. However, factors such as student isolation may lead to a high dropout rate (Packham et al. 2004, Park et al. 2009) and a high failure rate in distance learning. The most conventional ways to overcome this problem are forums, which we offered as part of each teaching module, or the use of email for direct contact with the teacher. In general, forum discussions dealt with misunderstanding of mathematical formulas or concepts, understanding the use and sense of statistical calculations and reasoning, particular problems when using statistical software, questions on particular steps when resolving exercises, etc. Forums are also conducive for exchange between students. Students can answer each other's questions, share difficulties, and give explanations and advice, which is beneficial for all students concerned (Hall et al., 2010).

To assess the quality and effectiveness of our courses and understand what has not worked in order to change it, a systematic annual evaluation is conducted. Mills and Raju, 2011 summarized effective practices for teaching statistics online, and in particular, they recommend conducting evaluations to monitor the teaching and learning process. The satisfaction survey that is sent to students contributes to improving the program. Questions relate to learning objectives (for each module and software training), interactivity (utility, effectiveness), visual impact, technical functionalities (interface and navigation effectiveness) and training tools, assistance quality (technological, pedagogical or administrative problems), time (real availability, real personal study/work time), habits and environment (internet accessibility, material preferences). Tudor, 2006 showed the impact of interactivity on student satisfaction.

Since participation in the survey is not compulsory, a selection bias effect is more than likely, so caution is required when interpreting these results (Zumrawia et al., 2014). Roughly, students responding to survey questions are globally satisfied. Since participation in the survey is anonymous, comparisons in reactions to course innovations or in survey responses between different groups (according to residence region, dropping out, passing the course, etc.) are not possible. The interest is to gather comments and ideas that have led to the implementation of new pedagogical practices. First, the forum alone cannot meet the expectations of all students. As in a classroom, shy or unsure students may be reluctant to openly formulate their questions on a forum. After two years of experience, we decided to introduce the notion of "tutored exercises" which are non-graded assignments in every module that lead to personalized feedback. They allow students to train, to evaluate themselves and to receive personalized correction so that they can identify potential errors or misunderstandings and improve their writing. By providing personalized feedback that guides students towards improvement, student motivation is fostered throughout the year and the feeling of isolation is somewhat reduced. Furthermore, teachers receive quick feedback on learners'' difficulties. In fact, there is a close relationship between doing “tutored exercises” and results in graded assignments and at the final exam (Figure 1). This relationship should be interpreted cautiously since further research is required to establish whether doing “tutored exercises” has a positive effect on motivation by keeping students involved and active or whether it is simply a marker of motivation.

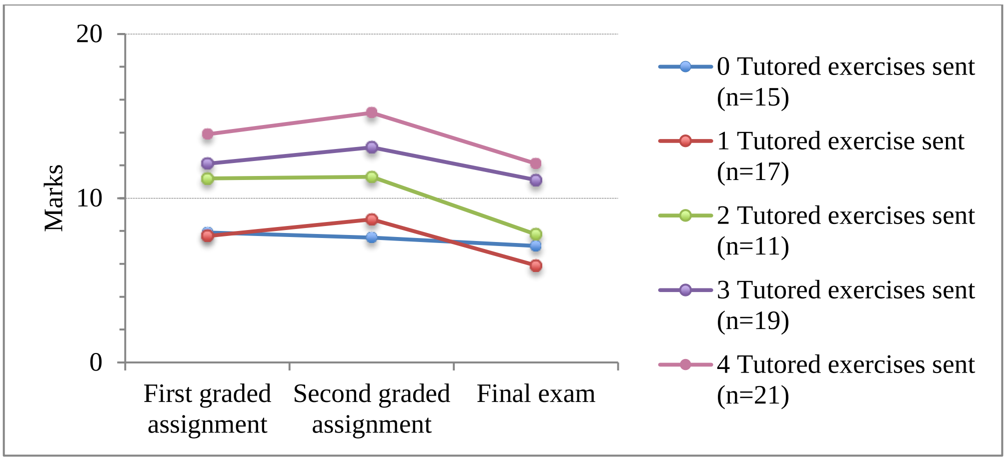


Figure 1: Average marks (the higher the better) of graded assignments and final exam per number of “tutored exercises” sent among the four “tutored exercises” (course in *Statistical methods in health*, Bordeaux School of Public Health, University of Bordeaux, France, 2014-2015, 83 students). n indicates the number of participants in the group.

Examples used in the course, “tutored exercises”, graded assignment and final exam questions are based on real publicly available data or real data with restricted use. Also, simulated data based on health studies are used. Real data with restricted use came from studies conducted in the Bordeaux School of Public Health, e.g. the intima-media thickness measurement study (Mercié et al., 2002) or the PAQUID cohort of individuals aged 65 years or older followed from 1988 until present to study the effects of different environmental, behavioural, and social vectors of age-related medical conditions and diseases (Lemeshow et al., 1998). Simulated data are based on epidemiological research articles or on studies conducted in the Bordeaux School of Public Health that are not available for educational purposes, such as data coming from the i-Share cohort aimed to gain better knowledge and understanding of students' state of health over a period of at least 10 years (Guichard et al. 2016).

All the modules now start with a video presentation of the teacher who concisely summarises the course content. This provides learners with eye contact with the teacher. We recently introduced virtual classrooms by means of the Adobe Connect web conferencing software using a flipped classroom style. In general, students appreciate real-time contact and oral communication to ask questions (instead of writing questions using the forum, which needs precision and involves non-instant feedback). Furthermore, discussions are more fruitful and students' errors of interpretation or understanding are identified more quickly and easily. At least one virtual class per module and statistical software are planned and announced from the beginning of the academic year. Different Internet connections or equipment quality, time zones and availability are still a problem with synchronous online classes in which all students need to be online at the same specific time in order to participate. A brief summary of questions and discussions is posted on the forum especially for absent participants.

Initially, EpiInfo, an easy-to-use database and statistical analysis program widely used by epidemiologists, and MS Excel spreadsheet were the only programs taught in this university course. While they are well suited to our audience, neither EpiInfo nor MS Excel spreadsheet allow the implementation of all the methods taught in this course. R (R Core Team, 2013) offers greater scope and its fast-growing popularity makes its use mandatory. In recent years, the introduction to the R module has been improved and virtual classes allow direct explanations. It has become more and more difficult to propose multiple software solutions to resolve data analysis exercises. Consequently, we now focus on R mastery. However, students' results are still unsatisfactory so considerable effort needs to be made by students to master this software.

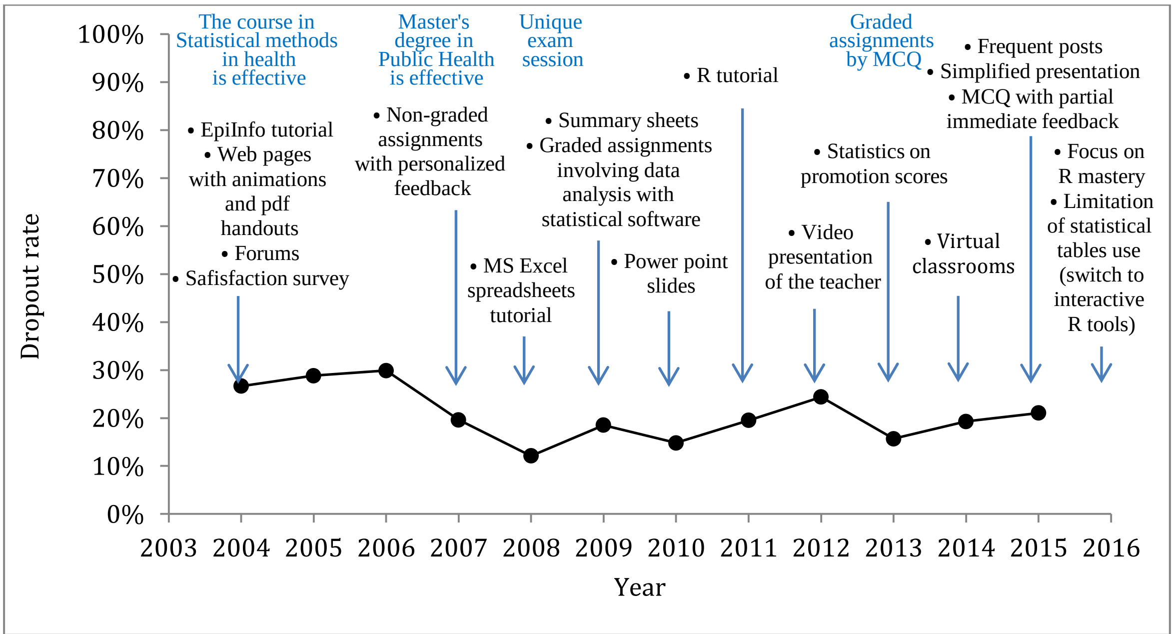


Figure 2: Dropout rates and innovations adopted over time from 2004 to 2016 (course in *Statistical methods in health*, Bordeaux School of Public Health, University of Bordeaux, France).

As a result of students' requests or remarks concerning their habits, we introduced changes to the presentation. First, summary sheets were provided for each module. In addition, the redundant content available as web pages and power point slides was deleted and only downloadable pdf handouts were kept. Web pages were still useful for the animations they contained. This led to a simplified presentation. The homepage is now animated and frequently renewed to encourage users to return regularly. Multiple-choice questions (MCQ) have been added to each module with partial immediate feedback that encourages students to search for the right answers in the course. However, compared to popular e-learning platforms like Moodle which allowing for fast developments and the integration of new functionalities, our platform lacks modernity. Since Moodle has been widely adopted by the University of Bordeaux, we now need to create a link between the platforms in order to capitalise on what Moodle has to offer, e.g. the wide variety and potential of its evaluation tools.

Figure 2 shows annual dropout rates from 2004 to 2016 and summarizes innovations adopted over time. Also, organizational modifications are indicated. For example, the suppression of the second session exam (because of the organisational burden and small effectiveness) and the graded assignments using a MCQ system that improved timing of notification of results to students. In these

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

graded assignments students are expected to solve each MCQ by conducting data analysis using software. In general, the relationship between innovations adopted and dropout rates are not conclusive and confounding is more than likely. However, adoption of "tutored exercises" shows a positive effect on the decrease of dropout rates.

CONCLUSION

Distance learning has become a great tool for gaining access to knowledge and diploma courses. It particularly encourages the professional development and improvement of public health in Southern countries while contributing to maintaining the integration of their health managers in them. Since its inception, the educational strategy of the *Statistical methods in health* course has not been limited to the launch of downloadable educational content but also included exploration of the best possible ways to supervise and counsel students throughout their gradual learning process. Thanks to feedback from annual training satisfaction surveys sent to students, we have introduced several changes, e.g. "tutored exercises", virtual classes, new statistical software training and supplementary MCQ. Because of the importance of provision to Southern countries, it would be useful to analyse feedback by region of residence. The assessment of R courses is still a challenge for students attracted by the course. Furthermore, it is difficult to keep abreast with other learning platforms such as Moodle that are developing innovative tools at a fast pace thanks to the heavy financial investments made in them. Nevertheless, after a decade, we now have sound experience in ODL, even though this sector requires constant re-evaluation, re-adaptation and re-thinking of the best educational strategies and learning tools.

REFERENCES

Brendel K.A. et al. (2011). EpiInfo™, a database and statistics program for public health professionals. CDC, Atlanta, GA, USA, 2011.

Dokeos. Extracted from http://www.dokeos.com/, last assessed on 24th October (2016).

Guichard E., Montagni I., Tzourio C., Kurth T. (2016) Association Between Headaches and Tinnitus in Young Adults: Cross-Sectional Study. *Headache,* 56(6):987-94.

Hall S., Vance E. A. & Tech V. (2010) Improving Self-efficacy in Statistics: Role of Self-explanation & Feedback. *Journal of Statistics Education* Volume 18, Number 3.

Lemeshow S., Letenneur L., Dartigues J.F., Lafont S., Orgogozo J.M., Commenges D. (1998) Illustration of analysis taking into account complex survey considerations: the association between wine consumption and dementia in the PAQUID study. *Am J Epidemiol.,* 148(3).

Moodle. Extracted from https://moodle.org/, last assessed on 24th October (2016).

Mercié P. et al. (2002) Evaluation of cardiovascular risk factors in HIV-1 infected patients using carotid intima-media thickness measurement. Ann Med., 34(1):55-63.

Mills J.D. & Raju D. (2011) Teaching Statistics Online: A Decade's Review of the Literature About What Works. *Journal of Statistics Education*, 19(2).

Packham G., Jones P., Miller C. & Thomas B. (2004) E-learning and Retention: key Factors Influencing Student Withdrawal, *Education + Training*, Vol. 46 Nos 6/7, pp. 335-42.

Park J. H. & Choi H. J. (2009) Factors Influencing Adult Learners' Decision to Drop Out or Persist in Online Learning. Educational Technology & Society, 12(4), 207-217.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Tudor G. E. (2006) Teaching Introductory Statistics Online – Satisfying the Students. *Journal of Statistics Education*, 14(3).

Zumrawia A. A., Batesa S. P. & Schroedera M. (2014) What Response Rates are needed to make Reliable Inferences from Student Evaluations of Teaching? *Educational Research and Evaluation: An International Journal on Theory and Practice* Volume 20, Issue 7-8.

APPLETS TO SUPPORT REASONING ABOUT EXPLAINED AND UNEXPLAINED VARIABILITY

MORPHETT, Anthony and GUNN, Sharon
School of Mathematics and Statistics
The University of Melbourne
a.morphett@unimelb.edu.au, sharonkg@unimelb.edu.au

Reasoning about overall variability in a data set, and the partitioning of variability into explained *and* unexplained*, is a key aspect of statistical reasoning. It also provides a powerful conceptual framework for unifying several sometimes disparate topics of introductory statistics such as linear regression, ANOVA, and significance testing of a fitted model. This type of reasoning creates a platform for comparing statistical models and hence lays a foundation for inferential thinking. We will describe a pair of applets that have been developed with the explicit aim of helping students to reason visually about the partitioning of variability into explained and unexplained variability. The first applet deals with the case of a numeric explanatory variable (regression for bivariate data), the second with the case of a categorical explanatory variable (single-factor ANOVA). We highlight some aspects of the applets' design which are of pedagogical relevance, and discuss how the applets may support possible learning trajectories for reasoning about variability in the topics of regression, ANOVA and statistical modelling.*

INTRODUCTION

The partitioning of variability into a portion that is *explained* by a statistical model (a.k.a. 'signal', 'between-groups') and a portion that is *unexplained* (a.k.a. 'noise', 'within-groups') is at the heart of statistical modelling. It is the foundation of analysis of variance and inference using the F-test. Nevertheless, it is not often treated explicitly in educational resources. One exception is the work of (Reid & Reading, 2010), (Reid, Reading, & Ellem, 2008), who suggested a framework for describing the progression in thinking and reasoning about explained and unexplained variability and developed associated assessment items to measure the progression.

Given the importance of partitioning variability, and the seeming lack of relevant, accessible resources online, a pair of applets were developed for visualising the partitioning of variability in two contexts: data with a numeric response and numeric explanatory variable (regression), and a numeric response and categorical explanatory (one-way ANOVA). The applets were developed as part of a wider project (Morphett, Gunn, & Maillardet, 2015) and are available at www.melbapplets.ms.unimelb.edu.au . As well as providing visualisations of the partitioning of variability in each case, we wanted visual representations that would reveal the underlying links between the two contexts – in particular, that one-way ANOVA and regression are two instances of the general analysis of variance procedure. The applets' design was informed by research such as the theory of multimedia learning (Mayer & Moreno, 2003). Many of the principles which guided the applets' design were consistent with those discussed in (Wild, Pfannkuch, Regan, & Parsonage, 2013) and (Arnold, Pfannkuch, Wild, Regan, & Budgett, 2011).

The applets share some features with earlier related applets, such as those of the CAST (see http://cast.massey.ac.nz/core/index.html?book=agExper&page=sec_oneFactorAnova , http://cast.massey.ac.nz/core/index.html?book=regn&page=sec_regnAnova) and WISE projects (see http://wise.cgu.edu/portfolio/demo-regression-and-correlation/), but differ in purpose and design. Our applets are intended to support *visual reasoning* (Dreyfus, 1991), as a complement to symbolic reasoning. We believe that visual and symbolic reasoning, married with a narrative, can potentially lead to more complex knowledge structures and consequently, deeper learning. Hence, the applets were designed from the outset to support *conversations*, in lectures or small-groups, and we envision their use accompanied by a spoken narrative. They are not intended to be used in isolation, nor to be a comprehensive treatment of the concept. In the next sections of this paper, we will describe the applets and highlight some aspects of their design which are pedagogically relevant for partitioning of variability.

The applets are designed to support reasoning about explained and unexplained variability, and are most effective when situated within a curriculum that focuses upon statistical modelling with

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

models comprised of deterministic and non-deterministic components. In the final section, we describe a possible learning trajectory, making use of the applets, for developing an understanding of modelling in this perspective.

APPLET 1: VISUALISING PARTITIONING OF VARIABILITY IN ANOVA

The applet initially shows an individual value plot of bivariate data with a categorical explanatory variable (group – on horizontal axis) and numerical response (vertical axis). The data is divided into three groups, shown in red, green and blue, with 5 observations $y_{i1}, \dots, y_{i5}$ in each group (i = 1, 2, 3). The group mean $\bar{y}_i$ of each group is shown as a triangle aside the group data. (The triangle can be thought of as the pivot of a see-saw, visually re-enforcing that the mean is a 'balance point' for the data.) To the left of the plot's vertical axis is a dotplot (running vertically) of the pooled data. The points in the dotplot are coloured according to group. The grand mean $\bar{\bar{y}}$ is shown as a dashed horizontal line running across the plot and as a black triangle beside the dotplot. To the left of the dotplot is a "variability meter". It is a bar whose height is proportional to the variance $\mathrm{Var}(y_{ij})$ and represents the total variability in the pooled data. It is initially coloured grey. The components just described make up the region labelled 'Overall variability' in Figure 1.

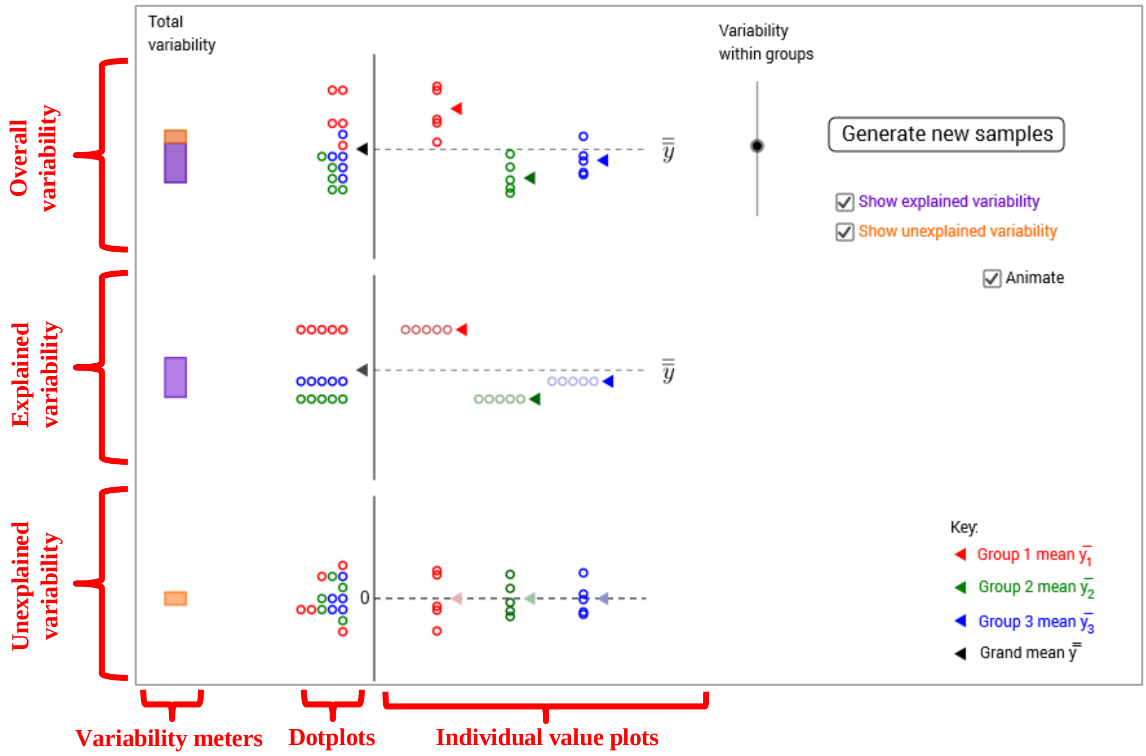


Figure 1: Screenshot of ANOVA applet after all animations have finished. Sections corresponding to overall, explained and unexplained variability are indicated at the side; the variability meters, individual value plots and scatterplots are indicated beneath.

The user may modify the group means by clicking and dragging the triangles representing the mean for each group, or may change the within-group variability using a slider labelled 'Variability within groups' to the right of the plot. A button "Generate new samples" produces a new set of data with the same group means and within-groups variance. When the data is modified in any of these ways, the individual value plot, dotplot and variability meter update automatically.

There are two checkboxes to the right of screen. When the first checkbox, labelled 'Show explained variability', is enabled, an animation is triggered. First, a new vertical axis appears in the middle section of the display below the 'Overall variability' section described above. Second, the points in each group of the individual value plot move into a line next to the group mean. Next, the aligned points, group mean triangles, and grand mean line $\bar{\bar{y}}$ descend, maintaining their relative positions, to form a plot by the new axis. The resulting plot shows the group means and 'predicted' values $\hat{y}_{ij} = \bar{y}_i$ of the response, and the grand mean $\bar{\bar{y}}$. Finally, a dotplot of the predicted values $\hat{y}_{ij}$

appears to the left of the new axis, with the points coloured by group; a variability meter for the 'explained' variability appears to the left, and the original data points and means reappear on the top plot. The height of the variability meter is proportional to $\mathrm{Var}(\hat{y}_{ij})$, or equivalently, to $n\,\mathrm{Var}(\bar{y}_i)$ where $n = 5$ is the number of observations in each group. The variability meter is coloured purple, and as it appears, a corresponding portion of the 'overall' variability meter from the top section, equal in height to the new variability meter, is also coloured purple. The animations just described produce the region labelled 'Explained variability' in Figure 1; its structure mirrors that of the overall variability section.

When the second checkbox 'Show unexplained variability' is checked, a set of axes appears in the bottom section of the display, and the data points from each group, and their group means, descend until the group means align with the horizontal axis of the new set of axes, thus forming a plot of the residuals e_{ij} in the ANOVA model $y_{ij} = \bar{y}_i + e_{ij}$. The data points retain their original positions relative to the group mean. To conclude the animation, a dotplot and variability meter appear, and the original data reappears in the top plot. The variability meter in this case is coloured orange, and its height is proportional to $\mathrm{Var}(e_{ij})$. When it appears, the remaining portion of the overall variability meter from the top section, equal in height to the new variability meter and until now coloured grey, also turns orange. A screenshot of the applet, with all animations complete, is shown in Figure 1.

APPLET 2: VISUALISING PARTITIONING OF VARIABILITY IN REGRESSION

The applet initially shows a scatterplot of bivariate numerical data (x_i, y_i). The sample size of $n = 15$ is fixed. The individual data points can be changed by dragging each point within the scatterplot. Also shown on the scatterplot is a horizontal dashed line representing $\bar{y}$, the grand mean of the observed responses y_i, and pale grey vertical line segments from each point to the grand mean line, representing overall deviation from the grand mean. To the left of the scatterplot is a dotplot of the responses y_i, with a triangle indicating the mean. On the far left is a variability meter, initially grey, whose height is proportional to the variance $\mathrm{Var}(y_i)$. Each of these components updates automatically when any of the data points are changed by dragging.

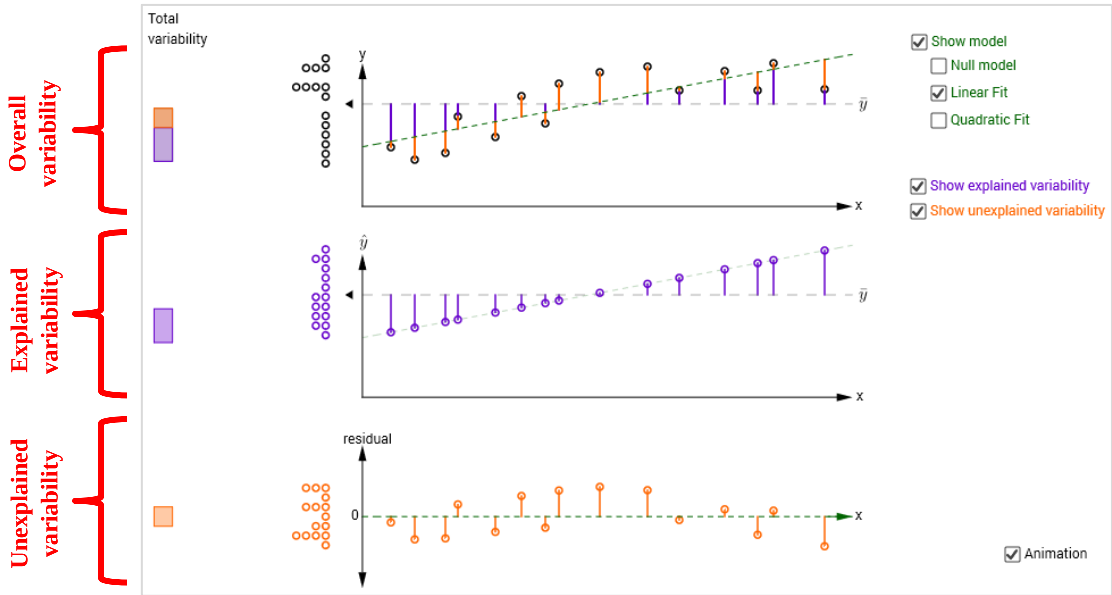


Figure 2: Screenshot of the regression applet with all components shown and all animations complete.

Checkboxes to the right allow more elements to be added to the display. The first checkbox, labelled 'Show model', adds a curve representing a regression model to the scatterplot. The user may choose between three options for the regression model: a linear model (of the form $y_i = \alpha + \beta x_i + e_i$, where α, β are constants and e_i a random residual), a quadratic model (of the form $y_i = \alpha + \beta x_i + \gamma x_i^2 + e_i$), and the null model ($y_i = \bar{y} + e_i$). When the "Show explained variability" checkbox is enabled, the scatterplot points animate and ascend/descend to the grand mean line. At the same time, vertical line segments, coloured purple, appear between the grand mean line and the

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

regression curve. These depict the individual deviations of the predicted values $\hat{y}_i$ about the grand mean $\bar{y}$. Next, copies of the points, grand mean line, and purple segments descend from the original plot to a new set of axes below the original scatterplot, to form a scatterplot of the predicted values $\hat{y}_i$. A dotplot of the $\hat{y}_i$'s appears to the left of the axes, and a variability meter, purple, appears to the far left, representing the variability in the $\hat{y}_i$.

When the final checkbox "Show unexplained variability" is enabled, orange vertical line segments appear between each point and the regression curve, depicting the residuals. The points, grand mean, regression curve, and orange line segments animate and descend from the top plot to a new set of axes at the bottom of the display, to form a scatterplot of the residuals. A dotplot of the residuals appears to the left of the axes, an orange variability meter appears at far left, whose height is proportional to the variance $\mathrm{Var}(e_i)$, and the remaining section of the overall variability meter is coloured. To complete the animation, all visual components reappear in the original scatterplot. Figure 2 shows the applet with all checkboxes enabled and after the animations are finished.

DESIGN CONSIDERATIONS

In both applets, the dotplots show the variability in overall data (top section), variability in means (middle section), and variability about means (bottom section). The variability meters on the left clearly show the partitioning of a total variability into parts corresponding to the explained and unexplained portions. This is a visual representation of the 'partitioning of sum of squares' theorem $SS_{total} = SS_{error} + SS_{treatments}$, and a subtle reinforcement of the independence of the random and deterministic components of the models. The animations show the act of partitioning; they can be repeated over and over to *cement* the abiding image. The movement is 'eye-catching'. However, as in (Wild et al., 2013), we see the animation as valuable when the focus is on understanding the *nature* of partitioning variability, but it may become an undesirable distraction once the nature is understood and the focus shifts to the *effects*. To mitigate this, an additional 'Animation' checkbox allows the lower panels to be shown or hidden without the delay of the animation.

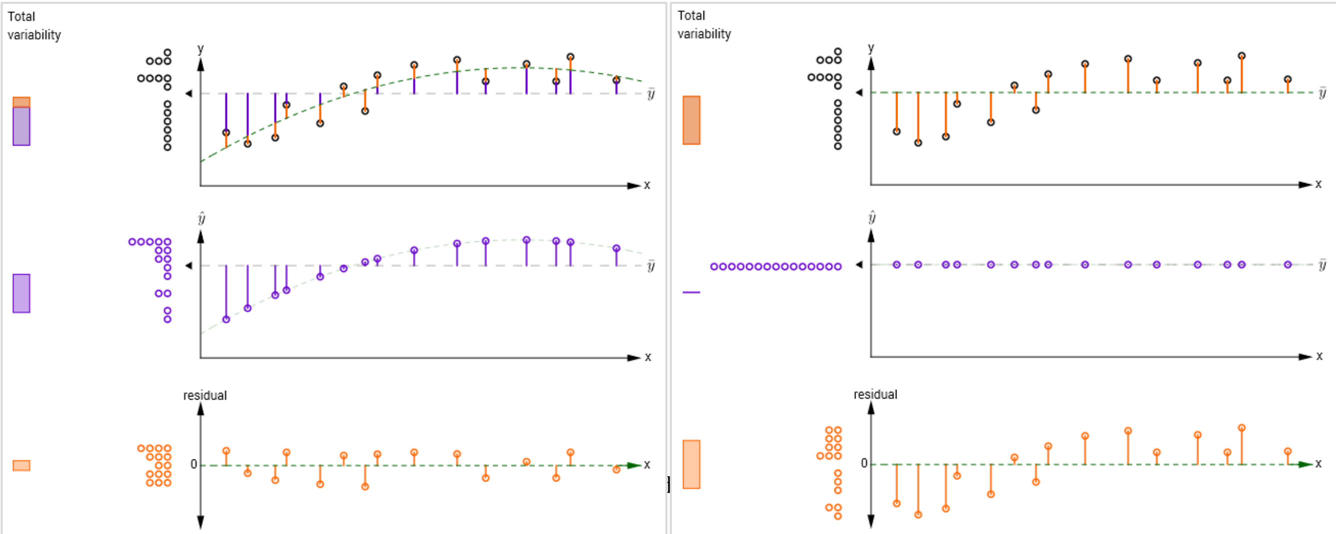
In order to focus on the visuals, we have chosen not to include any numbers or formulas and to use minimal labels in the applets' displays. By omitting a numerical scale or axis label for the variability meters, it is left agnostic whether the variability is measured as a sum of squares, a variance, or perhaps some other measure of variability. This helps to target a more fundamental concept – "the variability can be partitioned" – rather than a specific mathematical realisation of this idea – "the sum of squares can be partitioned". Our aim was a visual representation that would produce *abiding images* of the concept of partitioning variability. Abiding images are encouraged by visuals that are dynamic and *stripped back* so that features of the concept are clearly visible (Wild et al., 2013). The stripping back is about reducing cognitive load (Mayer & Moreno, 2003) and involves the removal of all unnecessary distractions, such as formulae and numbers (in our case), as well as non-essential complexity, such as additional user interface elements to change the sample size or number of groups.

The overall layout of the regression applet mirrors that of the ANOVA applet discussed earlier. The use of a common layout and visual components is intended to help convey the unity of underlying concepts: that regression and one-way ANOVA are two variations of the common idea of modelling data with deterministic and random components. (Indeed, they are two instances of the more general concept of general linear model.) The intent is to "convey the 'sameness' of what is happening [...] to reveal the unity of the underlying way of thinking" (Wild et al., 2013, p. 10).

These are the three features that, we feel, most distinguish these applets from previous applets such as those of CAST and WISE cited earlier: the stripped back visuals for supporting visual reasoning, the use of animations to further make explicit the process of partitioning, and the common layout and visual elements to reinforce the unity of the underlying concepts.

One notable difference between Applets 1 and 2 is the enhanced ability to compare models in the regression applet. By switching between 'linear fit', 'quadratic fit' and 'null model', the user can observe changes in the *nature* of explained and unexplained variability (as seen in the purple and orange line segments comprising the deviation from the mean in the top section), in the *quantity* of 'explained' and 'unexplained' variability (as reflected in the size of the 'explained' and 'unexplained' variability meters or the spread of the dotplots), and in the partitioning of overall variability (as reflected in the break-up of the overall variability meter into purple 'explained' and orange 'unexplained' components). This is particularly powerful for the null model, where it is clear from the variability meters and explained/unexplained dotplots that, under this model, 'all variability is unexplained'. The user can also observe the preservation of total overall variability, as seen by the total size of the overall variability meter remaining unchanged. This can be seen in Figure 2 (a linear fit), Figure 3a (a quadratic fit) and Figure 3b (null model). The data is the same in these three figures; only the choice of model is changed. We note that the 'null model' is visible in the ANOVA applet also, in the form of the $\bar{\bar{y}}$ line, though it is not explicitly labelled as such.

Another difference between the two applets is the way that users can manipulate the data. In Applet 1, the user can manipulate the group means and within-group variability, but not the



individual data points. This allows direct manipulation of the explained and unexplained components of variation, and avoids any issues with potential violation of the equal-variances assumption. In Applet 2, the user can drag individual data points directly. This allows the user the flexibility to choose data with a linear, quadratic or other relationship (although it admits the possibility of heteroskedasticity – a trade-off we feel is acceptable).

These applets have been used in an introductory statistics course for first-year science students taught by the second author for several semesters. The majority of students in this course are not specialising in statistics and many do not have advanced mathematical backgrounds. We feel that the emphasis on visuals taken by the applets is appropriate for such students. For some results from an evaluation of the applets (as part of a suite of applets) in this introductory statistics course, we refer the reader to (Morphett et al., 2016).

MAPPING THE APPLETS INTO A POTENTIAL LEARNING TRAJECTORY

We now illustrate how the applets map to a possible learning trajectory focused on developing an understanding of modelling, which is largely followed in an introductory statistics subject taught by the second author. The learning trajectory for this subject begins by (1) exploring random variation, including sampling variation and other sources of variation. This is followed by (2) modelling variation in outcomes of a random experiment using probability models, eg binomial and normal distributions. (3) With these tools we look at random variation about a single mean. Random variation about the mean is modeled using a distribution. Discussion then moves to competing models, i.e., $\mu = 50$ vs. $\mu \neq 50$, and (informally) the idea of variability that can be explained by the null model versus variability that can't be explained by the null model. (4) We next move to variability about two means (separate means model). This involves modelling with a deterministic component which depends on the explanatory variable. (5) The next step is variability about several means. We now start talking more formally about variability which is *explained* and *unexplained* - by (the deterministic part of) the model, in each case. (6) We now progress to variability about a more complex model – a regression model – in which the deterministic component of the model is a linear or other function of the explanatory variable. In all steps of this trajectory we are building the notion of variability that can be explained by a proposed model versus variability that is not explained by the model. By steps (5) and onwards, this idea is becoming more formal with the introduction of the analysis of variance ANOVA table.

We first introduce the ANOVA applet at step (5) of this trajectory, however aspects of the applet reflect elements from earlier stages. For instance, sampling variability (1) is seen when generating a new sample, and the 'null model' – shown as the $\bar{\bar{y}}$ line – reflects variation about a single mean (3). We then lead into the regression applet at stage (6), but it also reflects earlier stages, particularly (1) and (3), and is a generalisation of (5). From here, we could lead into hypothesis testing as the comparison of two (nested) models – a possible stage (7).

In this trajectory, the commonality of stages (5)-(7) is seeing ANOVA not just as one single method for comparing several means, but as a powerful and general analysis technique – *analysis* of variance – that involves partitioning variability into 'explained' and 'unexplained' components. The unity of design of the applets, noted earlier, complements this.

We have discussed a pair of applets intended to support visual reasoning about variability and statistical modelling. The design of the applets provides several different ways of viewing the variability in the data. Common visual elements and layout between the applets help re-inforce the unity of underlying concepts, particularly when nested within a curriculum that focuses on statistical modelling, comparison of models and analysis by partitioning variability into explained and unexplained components. In future work, we hope to investigate student use of the applets with regard to the framework of (Reid & Reading, 2010).

On reflection, many of the principles which guided our design of the applets match those discussed in (Wild et al., 2013). The fact that we, largely independently, reached broadly similar considerations as Wild et al when working with different concepts suggests that these principles are sound, robust principles for educational software design in statistics.

ACKNOWLEDGEMENTS

The basic partitioning of variability diagram, and in particular the idea of repeating all sample points at the group mean in the section which illustrates between groups variability, is based on unpublished work by Dr Robert Maillardet, who we would also like to thank for valuable contributions and suggestions during the detailed applet design process.

REFERENCES

Arnold, P., Pfannkuch, M., Wild, C., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: From hands-on to "movies". *Journal of Statistics Education*, *19*(2).

Dreyfus, T. (1991). On the status of visual reasoning in mathematics and mathematics education. In Furinghetti, F. (Ed.), *Proceedings of the 15th PME International Conference*, Assisi, Italy.

Mayer, R., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, *38*(1), 43-52. doi: 10.1207/S15326985EP3801_6

Morphett, A., Gunn, S., & Maillardet, R. (2015). Developing interactive applets with GeoGebra: processes, technologies. In Blignaut, R. & Kizito, R (Eds.), *Proceedings of Elephant Delta: 10th Southern Hemisphere Conference on the Teaching and Learning of Undergraduate Mathematics and Statistics*, Port Elizabeth, South Africa.

Morphett, A., Maillardet, R., & Gunn, S. (2016). Conceptual Learning with Interactive Applets project report. The University of Melbourne. Available at http://www.melbapplets.ms.unimelb.edu.au/?page_id=201

Reid, J., & Reading, C. (2010). Developing a framework for reasoning about explained and unexplained variation. In Reading, C. (Ed.), *Data and context in statistics education: Towards an evidence based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS-8)*, Ljubljana, Slovenia.

Reid, J., Reading, C., & Ellem, B. (2008). Developing assessment items to measure tertiary students' reasoning about explained and unexplained variability. In Hays, T. & Hussain, R. (Eds.), *Proceedings of the 2nd Annual Postgraduate Research Conference: Bridging the Gap between Ideas and Doing Research*, University of New England, Armidale, Australia.

Wild, C. J., Pfannkuch, M., Regan, M., & Parsonage, R. (2013). Next Steps in Accessible Conceptions of Statistical Inference: Pulling ourselves up by the bootstraps. Unpublished draft, available at http://www.stat.auckland.ac.nz/~wild/TEMP/bootstrap.pdf

TO R OR NOT TO R – WHAT SHOULD WE BE CONSIDERING?

GUNN, Sharon and MORPHETT, Anthony
School of Mathematics and Statistics
The University of Melbourne
sharonkg@unimelb.edu.au, a.morphett@unimelb.edu.au

When choosing a statistical software package for introductory statistics, educators are faced with a range of choices. The choices include GUI-driven software such as Minitab, primarily command-driven packages such as R, and variations in-between. The justification for choosing a particular package may vary from a simplistic argument based on pragmatic considerations (practicalities such as costs, accessibility, functionality and design/technical aspects of the package) to more sophisticated arguments that also consider the educational affordances of the software package.

In this paper, we compare and contrast our experiences in two different introductory statistics subjects. In one subject, Minitab has been used for several years; in the other, Minitab was recently replaced with R and the Deducer GUI. We will consider the pedagogical affordances offered by each package, in the context of the curricula for these introductory statistics subjects. The comparison will focus on how well the software packages complemented the subjects' curricula – what worked well and what didn't. From our reflections we suggest some important considerations to make when choosing a pedagogical software package for an introductory statistics course.

INTRODUCTION

Which statistical package to use in introductory statistics courses (ISC's) is an issue that many statistics educators are currently re-visiting. Recently, there has been an increasing uptake of R (www.r-project.org) in ISC's (and textbooks). Some possible reasons for R's recent upsurge in popularity include its perceived superiority for working with 'big data' or for running simulations, as well as more pragmatic reasons such as R's being freely available, open-source and cross-platform. While consideration of these (and other) pragmatic concerns is important, a more informed choice would also include consideration of what the package, potentially, brings to the learning environment. How can the package help students to achieve the desired learning outcomes? Does use of the package lend itself to developing the type of thinking we are trying to encourage in our students?

Backwards Design principles (Wiggins & McTighe, 2005) say that we should start with the desired learning outcomes (LO's) and then choose the tools/activities/materials accordingly. In a similar vein, at the recent eCOTS conference, Bethany White commented, "Even though it is tempting to adopt new technologies as they emerge, we must be careful not to let the technology dictate the learning that occurs; rather it's the desired learning outcomes that should drive which technologies are used and how they are used." (White & Gibbs, 2016) With this in mind, when choosing a software package we should ask ourselves, how could a given package potentially serve our desired LO's?

A useful concept when addressing this question is the idea of *pedagogical affordances*. Pedagogical affordances (Gibson, 1979) are what the technology (potentially) offers in terms of learning. We note that affordances are not purely the possibilities that an object offers, but depend also upon the subject's ability to perceive how these possibilities may be realized. That is, the affordances need to be both transparent and accessible to the user: a software package may afford sophisticated computation, but only for a user who can perceive how it may be used.

In this paper we will focus particularly on two statistical software packages, R and Minitab. However, we feel that the points raised reflect deeper considerations which apply to all software packages. R is a software environment for statistical computation. As well as providing functions for statistical calculations, it offers a complete programming language and libraries for many computational tasks such as graphics and data processing. As a statistical package, it is typically used by typing commands into a console window, although add-on packages such as RStudio (https://www.rstudio.com/), R Commander (http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/) and Deducer (http://www.deducer.org/) provide graphical user interfaces for some tasks. R is freely available, open-source, and runs on many computing platforms including Windows and Mac. Minitab is a statistical software package. It is operated primarily through a graphical user interface, including

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

menus, toolbar buttons and dialog windows, though it also provides a command interface. Minitab is commercial software which requires a paid licence for use, and requires Windows (although a version with reduced functionality is available for Mac).

The first of our two subjects, which we will refer to as Introductory Statistics 1 (IS1), is a first-year elective introductory statistics subject for science students. The enrolment is approximately 250 students/year, with students from a range of majors including biological sciences, psychology, and mathematics and statistics. The focus of the subject is data analysis and the concept of modelling variation is a cornerstone of the subject. The modelling approach taken in the subject is described in (Morphett & Gunn, 2016). Conceptual understanding is valued more highly than procedural knowledge, and the primary aims of the subject are the development of statistical literacy, reasoning and thinking. Minitab has been used in IS1 for many years. One ILO is that students should be able to conduct a meaningful analysis using software (Minitab), and this is assessed by a lab test. Typically, students in IS1 have a firm handle on Minitab after just two lab classes, and by that stage are generally capable of performing an exploratory analysis of a simple data set and producing a written report with graphical displays, tables, summary stats and discussion.

The second subject, Introductory Statistics 2 (IS2), is a compulsory first-year subject for biomedicine students. Its enrolment is approximately 200 students/semester and these students, generally speaking, have stronger mathematical backgrounds than the students in IS1. The course was originally designed with a similar focus and underlying philosophy to IS1, and many of the teaching materials were developed in tandem with those of IS1, with adjustments for the different contexts – medical versus life science – and different mathematical backgrounds of the two groups. Both IS1 and IS2 have 3 lectures, 1 tutorial and 1 computer lab per week.

The ILO’s for IS2 are very similar to those of IS1 however, with constant staff changes over recent years the IS2 has lost some of its original focus on conceptual understandings. It was recently decided that IS2 would change from Minitab, which had been used in the subject for many years, to R. This change was largely driven by ‘pragmatic’ considerations – R was more familiar to the lecturers of the subject, was seen as the standard choice by researchers in the field, and, being open-source, was seen as being more readily available to students. In an effort to retain some of the perceived advantages of a GUI, it was decided to use the Deducer GUI for R (www.deducer.org). Deducer is a graphical interface which includes a spreadsheet window for data entry, and a menu and dialog-driven interface for performing statistical tasks, rather than typing R commands at the command line. The change was implemented in semester 1 of 2016. The computer lab activities, which were originally written for Minitab, were modified for Deducer by replacing instructions based on the Minitab interface with R code or instructions based on the Deducer interface. The underlying statistical tasks were changed very little. The use of Deducer was dogged with technical problems (such as the software crashing, or failing to load at all on some lab PCs) so that mid-way through the semester the Deducer interface was largely abandoned in favour of entering R commands at the command line, either in the R console or in the RStudio UI (www.rstudio.com).

SOFTWARE SUPPORTING LEARNING OUTCOMES

In this section we illustrate how Minitab and R may support some typical intended learning outcomes of introductory statistics.

Developing statistical thinking and reasoning. Exploratory data analysis is an important skill developed in IS1 and IS2. It is an essential component of the statistical problem solving cycle and, as such, underpins the development of statistical thinking and reasoning (Chance, 2002).

There are several features of the Minitab user-interface that support exploratory data analysis (and hence statistical thinking and reasoning), the most obvious one being the structure of the Drop down menus. For example, the sequence of selecting a menu (e.g., ‘Graph’; ‘Stat’) then a menu item (eg ‘Dotplot’; ‘Basic Statistics’) and then the necessary options within the resulting dialog window (eg, one Y vs. multiple Y’s, with or without groups; Display Descriptive Statistics, Variables and/or By variables) encourages students to think about what they want the package to do (display or summarize the data), the structure of the data (response variables and grouping variables), and, in the case of summarizing the data, what summary measures are meaningful in the context. The student is thus engaged in analytic thinking from the outset, and the menu structure has helped to scaffold the decision-making process in a manner consistent with a commonly accepted learning trajectory associated with conducting a ‘meaningful analysis of a data set’.

A further example of the user interface structure supporting statistical reasoning is the options to check model assumptions embedded within each statistical analysis procedure. Model assumptions can be checked either graphically and or using formal hypothesis testing, depending upon the type of analysis selected. The checking of assumptions is a critical aspect of statistical reasoning that is often overlooked by beginning students.

We feel that the Minitab menu interface is (potentially) powerful because, to a large extent, the user’s pathway through the Minitab menu system mirrors the learning trajectory of important statistical concepts, such as modelling variability and model building in general.

This is well illustrated by the ‘brush’ tool in Minitab. The ‘brush’ tool allows users to click and drag to select data points in a graphical display. The identifying characteristics of the ‘brushed’ observations can be displayed (see Figure 1 below). This makes it easy for students to look into the data: to seek explanations for unusual observations or for the observed variability in the data and to begin building models to describe the variability in the data. This is an example of representations as dynamic tools for analysis, one of the five affordances of technology identified by Pratt et al (Pratt, Davies, & Connor, 2011).

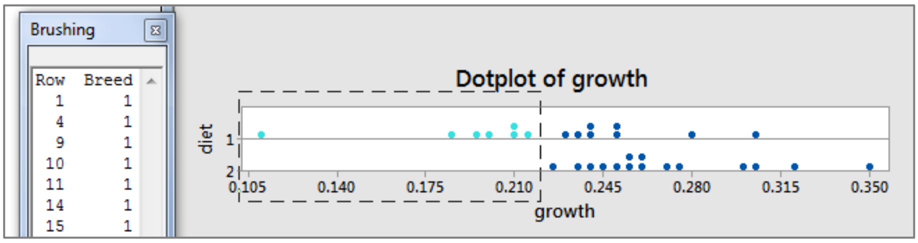


Figure 1. Investigating a possible additional explanatory variable using Minitab’s Brush tool.

Of course, exploratory data analysis is certainly possible in a command-driven environment such as R, but without supports provided by the environment such as those described above, it requires more effort, and places a greater cognitive load on the user, to obtain the same statistical results.

Understanding of concepts underpinning statistical techniques. Technology can support conceptual understanding by being transparent about statistical concepts. We illustrate this with an example. Consider a 2-sample t-test on independent samples. The underlying statistical question is whether the samples provide sufficient evidence that the two population means differ. The parameter under investigation is the difference in means, $\mu_1 - \mu_2$, and we wish to know if 0 is a plausible value for the parameter. Figures 2 and 3 show the output from performing a 2-sample t-test in Minitab and R, respectively. Note how Minitab defines the parameter of interest, namely the difference μ(control) - μ(treatment), and provides a point estimate of the parameter as well as a confidence interval, and states explicitly what is being estimated. R provides point estimates of the separate population means but does not explicitly link this to the parameter being estimated, nor state explicitly the parameter being estimated by the 95% CI. Minitab’s output is thus more transparent about the statistical concept of a parameter. Furthermore, R refers to the alternative hypothesis only whereas Minitab refers to the null and the alternative hypotheses being tested and is thus more transparent about the statistical concept of testing hypotheses. The order in which the output is presented is also worth noting here: in Minitab the movement from sample to inference about a population is explicit, in R it is not so clear but appears to be in reverse order – from assumed population to information contained in the sample (with a bit of to-and-froing). We do not mean to judge Minitab and R on the basis of this one example, and we acknowledge that neither’s output format is immune from criticism and that there will be other examples where R’s output may better fit a desired narrative. Rather, we intend this example to highlight the kind of subtle distinctions which are potentially of pedagogical relevance.

```
Two-Sample T-Test and CI: control, treatment
Two-sample T for control vs treatment

            N   Mean  StDev  SE Mean
control    14  50.07   2.12     0.57
treatment  14  52.50   3.75      1.0


Difference = μ (control) - μ (treatment)
Estimate for difference:  -2.43
95% CI for difference:  (-4.80, -0.06)
T-Test of difference = 0 (vs ≠): T-Value = -2.11  P-Value = 0.044  DF = 26
Both use Pooled StDev = 3.0450
```

Figure 2: Output from independent 2-sample t-test in Minitab (version 17.2.1)

```
> t.test(control, treatment, var.equal=TRUE, paired=FALSE)

        Two Sample t-test

data:  control and treatment
t = -2.1116, df = 26, p-value = 0.04449
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.7959022 -0.0644835
sample estimates:
mean of x mean of y
 50.07060  52.50079
```

Figure 3: Output from independent 2-sample t-test in R (version 3.1.0)

Developing statistical literacy. *Statistical literacy* involves understanding and using the basic language and tools of statistics: knowing what statistical terms mean, understanding the use of statistical symbols, and recognizing and being able to interpret representations of data (Garfield & DelMas, 2010). Statistical software can support students' development of statistical literacy by using terminology which is consistent with statistics discipline. This is generally the case in Minitab, which refers to columns of data as 'variables' and distinguishes between 'text' or 'categorical' and 'numeric' variables. For comparison, R uses terminology such as 'frame' and 'vector' to refer to data sets and variables, and 'double' and 'character' for the type of (non-integer) numerical and textual data, respectively. This terminology is based on the computational structure of the data and reflects R's origin as a programming language. While students with prior programming experience may be comfortable with such terminology, it has the potential to cause increased cognitive load for students without a programming background as they must translate between 'R terminology' and 'statistical terminology'. Moreover, the use of 'text'/'categorical' and 'numeric' directs the user to think about the type of data they are analysing, and consequentially the type of summaries, graphs and analysis procedures that may be appropriate in the context of the data.

An online survey of IS1 and IS2 students was conducted in semester 2, 2016. The purpose of the survey was to explore how students were using their time in the computer labs and how they were managing with the statistics package (Minitab, R). When asked about the terminology used in the relevant statistics package (Minitab/R) 67% of the IS1 respondents either agreed or strongly agreed that the terminology used was familiar compared with 37% of the IS2 respondents. See Figure 4 below. This is, in large part, a reflection of the packages themselves but it also says something about the supporting teaching resources such as the textbook (what package is used there), the lectures (how the package was demonstrated in lectures), the lab materials (how did they deal with terminology).

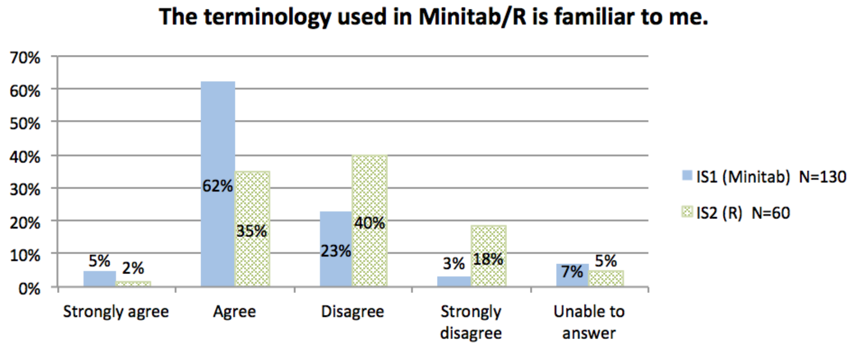


Figure 4: Comparison of Minitab and R –terminology used by package.

In the discussion above we have focussed on the standard R platform, operated from the command-line, without any add-ons. R does support a broad ecosystem of packages, which can extend or modify R's behaviour, as well as graphical front-ends such as RStudio (www.rstudio.com) and Deducer (www.deducer.org). Instructors may be able to find or create some combination of packages and add-ons which provide a software environment more aligned with the ILOs than bare command-line R, but this does come at some cost. In particular, effort by the instructor is needed to find such packages and customise teaching materials accordingly, and additional technical support for staff and students is potentially required. As a general-purpose programming language, with sufficient effort anything is possible in R – the question is, how much effort are you prepared to make?

ADDITIONAL CONSIDERATIONS

As mentioned earlier, pedagogical affordances of technology arise from the interaction between the interface and the user. The user-interface therefore needs to be easy to navigate and the way to access new functions should be discoverable without needing reference to documentation. In the survey mentioned earlier, students were asked about the discoverability of Minitab/R. Forty-five percent (45%) of the IS1 students found it easy to work out how to do new things without being shown how compared with 23% of the IS2 students. See Figure 5 below.

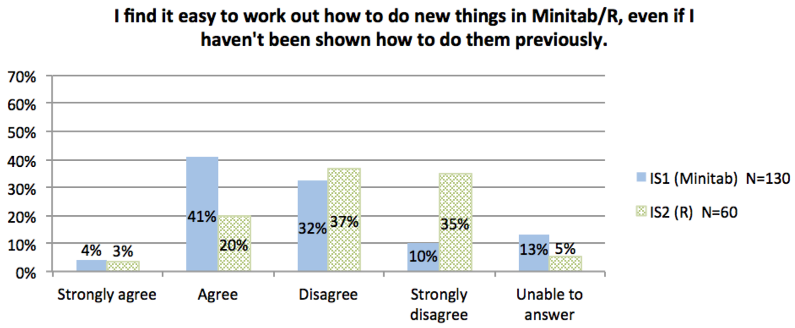


Figure 5: Comparison of Minitab and R -- level of discoverability

In the early stages of IS1 (pre 2000) MTB was used in its command line form. With the advent of drop down menus the pedagogical affordances of Minitab changed and this was reflected in the teaching materials and resources supporting the subject and desired learning outcomes. This interplay between pedagogical affordances and achievable learning outcomes needs to be acknowledged when making decisions about which package to choose… something which became very obvious to us with our experiences in IS2.

The subject IS2 had previously used Minitab for many years, and changed to R at the start of 2016. The lab materials were modified to use R instead of Minitab, mostly by replacing instructions based on Minitab's menus with instructions based on the Deducer GUI or R commands. One thing that we observed was that lab activities which had previously fitted neatly into a 50-minute lab class became much longer and slower with the new software, and students did not seem to get as much out of the activities. In retrospect, we feel that this was because we did not consider the changed affordances of the software. The lab activities had originally been written to take advantage of the affordances offered by Minitab; during the conversion to R, the activities were not re-designed to better align with the changed affordances of the new software package. Hence, what worked well with one software package became harder and slower in another. We feel that the lesson here is that a change in statistical software is not something that can be taken lightly – educators need to think about how the affordances differ, and what impact this may have on the curriculum, and adapt course materials accordingly.

STAISTICAL THINKING VS COMPUTATIONAL THINKING.

The subjects IS1 and IS2 are primarily service courses that prioritise understanding of fundamental statistical concepts and the ability to reason and think statistically. Courses in Data Science, on the other hand, are more likely to prioritise computational thinking. Some consultants hold the view that one must first develop the ability to think statistically before one can become a proficient data scientist. Others suggest that the ISC of the future would include development of both types of thinking as desired learning outcomes.

The two types of thinking, statistical and computational, differ just as mathematical thinking differs from statistical thinking. To encourage these different ways of thinking requires different affordances from the software. The R system is well suited for computational thinking but, in its current evolution, there is a risk of this being at the expense of developing statistical thinking. There is a trade-off. Educators should think about this and ask themselves what type of thinking they wish to prioritise in their course.

FINAL COMMENTS

As is probably clear from the discussion above, we feel that a switch from a graphical package such as Minitab to R comes with some pedagogical cost, as far as the development of conceptual understandings and statistical reasoning, thinking and literacy is concerned. There are valid pragmatic reasons for favouring R, such as it being open-source, its near-ubiquity in research, and its wide ecosystem of add-on packages. We are thus faced with trade-offs between the pedagogical and the pragmatic and between desirable learning outcomes. It is up to the individual educator to decide what trade-offs are acceptable to them, but we feel it important to acknowledge that trade-offs do exist.

Changing curricula may lead to changing requirements. And of course, software also changes over time, and with it, potentially, the pedagogical affordances. Thus, over time the points raised above may be less relevant as R evolves. We hope that the idea of pedagogical affordances will be helpful to guide the evolution of both R and introductory statistics courses over time.

REFERENCES

Chance, B. (2002). Components of Statistical Thinking and Implications for Instruction and assessment. *Journal of Statistics Education, 10*(3).

Garfield, J., & DelMas, R. (2010). A Web Site That Provides Resources for Assessing Students' Statistical Literacy, Reasoning and Thinking. *Teaching Statistics, 32*(1), 2-7. doi: 10.1111/j.1467-9639.2009.00373.x

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Morphett, A., & Gunn, S. (2016). Applets to support reasoning about explained and unexplained variability. *Proceedings of OZCOTS 2016*.

Pratt, D., Davies, N., & Connor, D. (2011). The role of technology in teaching and learning statistics *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (pp. 97-107): Springer.

White, B., & Gibbs, A. (2016). *Changing with technology to facilitate learning: Ideas for using technology to find new ways to help students to achieve learning outcomes*. Paper presented at eCOTS 2016.

Wiggins, G. P., & McTighe, J. (2005). *Understanding by design*. ASCD.

AUSTRALIAN STATISTICS POSTER AND PROJECT COMPETITIONS

RICHARDSON, Alice M.[1] and BARKER, Valerie [2]
[1]Australian National University, Acton ACT
[2]Lyneham High School, Lyneham ACT
alice.richardson@anu.edu.au

Secondary school statistics poster competitions have been running in Australia for over a decade. Students form groups of 2 – 3 students, carry out a data collection and present the results. The results have been traditionally presented in the form of a poster but in recent years Powerpoint presentations and videos have been accepted. State winners were selected by university academics, then judged by the academic team to identify a national winner in two to three categories (initially Years 7 – 10 and 11 – 12, which later changed to Years 7 – 8, 9 – 10 and 11 – 12). In 2008 the Australian Mathematics Trust became involved in the organisation of the competition.
In this paper we will describe the history of the Australian Statistics Competition (ASC) and the National Primary and Secondary Student Poster Competition (NPSSPC), along with the American Statistical Association's poster competition. Summary statistics on the topic of winning posters will be investigated. The distribution of the topics will be examined with reference to the Australian Curriculum. We will also offer suggestions that teachers can implement, in order for them to extract maximum value from participation in a statistics competition.

INTRODUCTION

The history of posters for disseminating scientific advances is sparse, with most attention coming from the medical literature; see for example, Barnicoat (1972), Cooter & Stein (2007), and Rowe & Ilic (2009) and the references therein. The academic statistics conference has a history of over 150 years; Florence Nightingale attended an International Statistical Congress in London in 1860. Since then conferences have grown, and the poster session began to be seen as an efficiency tool, allowing organisers to accept more presenters without having to resort to multiple parallel sessions or lengthy conferences. The presence of poster sessions at Australian Statistical Society conferences dates back to at least 1992. Virtual conference sessions, or entire virtual conferences, with electronic posters, have also become fashionable in the last decade. With the poster therefore being regarded as a legitimate academic output, undergraduate and postgraduate students have also been asked to produce posters of their studies as training for these poster sessions.

Academic competition is also far from new, and as posters have become settled in academic discourse, so have poster competitions (though more for students than researchers). The value of poster competitions to the university academics who frequently volunteer to coordinate and judge them is largely through the service component of their work (recognised by Bernreuter (1995) in the context of nursing students).

School students have for decades been producing posters, to demonstrate the results of an investigation in a succinct way that is convenient for sharing with a broad audience. Subjects with a strong visual element usually work best, and the authors have the strongest memories of producing such posters in what was then known as Social Studies in the 1970s and 1980s. Mathematics classrooms continue to be decorated with posters on topics that contain a strong visual element. For example, posters on the golden ratio, or tessellations, always look good and work well; but there are never as many posters on how to, for instance, solve simultaneous equations. Science Fairs for school students seem to date from the 1920s in the US (Cox 2016), and the authors certainly remember them in New Zealand in the 1980s.

The Australian Statistics Competition (ASC) began in the early 2000s in Queensland, sponsored by the Australian Bureau of Statistics and the Statistical Society of Australia Inc. (SSAI). By 2005 it had "gone national" with a network of volunteer judges in universities in every state. In 2008 the Australian Mathematics Trust (AMT) took over the administrative aspects of the competition, so that the competition could be advertised on an equal footing with the long-running Australian Mathematics Competition.

A trend towards projects began in the ASC in 2009 with the acceptance of Powerpoint presentations of around 8 – 12 slides, or 2 MB in size if another presentation tool were used. The

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

American Statistical Association (ASA) competition has accepted both posters and projects from at least 2000. The project, however, is a pdf of a written report rather than a poster or slide show. Thus the US project does not help students to learn to condense information into the succinct format of a poster or set of slides. On the other hand the ASC encourages creativity in presentation techniques, inviting students to draw on their technological skills which may well have been developed outside of any classroom.

In this paper we address specific address the distribution of content of winning statistics posters across two countries and approximately ten years, and the implications for teachers and students intending to participate. These are important questions to address in order to support teachers' and students' participation in these ever more popular competitions.

METHODS

The data for this study consisted of two parts. The first part of the data consisted of the title of winning posters (and more recently projects) from the ASC from 2008. Attention was restricted to the winners from 2008 onwards whose poster titles and/or links are given on the AMT website. There was no information available for 2009. A total of 84 titles were available for analysis.

The second part of the data consists of titles of winning posters and projects from the American Statistical Association competition from 2000 onwards. From 2011 posters only were named in the *Amstat News*, and links provided to projects. However, there was no systematic difference in topics between posters and projects and so the online information was not pursued. A total of 367 titles were available for analysis.

The first author entered the poster/project names into a database and allocated Australian curriculum areas (http://www.australiancurriculum.edu.au/) to each one. As of 2016, curriculum areas were: The Arts, Civics, Economics, English, Geography, History, Health & Physical Education (HPE), Mathematics, Science, and Technology. Each poster could have up to three curriculum areas allocated, which were given equal weighting. The second author conducted a similar allocation independently.

When the authors met to validate results and resolve disagreements, three principles were applied. First, no poster was to be allocated more than two curriculum areas. Second, when one author allocated a poster to two areas and the other allocated it to one of those, the project was allocated to that single area. There were 25 (Aus, 30%) and 78 (US, 21%) of disagreements resolved this way. Third, all topics involving motor skills, academic performance and social concerns were allocated to HPE. The remaining disagreements were resolved by discussion. There were 34 (Aus, 40%) and 135 (US, 37%) disagreements to resolve, a high number but a nice problem to have in the sense that many titles were quite interdisciplinary, especially at the primary levels in the US.

A limitation of this approach is its focus on titles not content of posters, and on winners not all entrants. Future research could extend the study in both of these directions. In the earlier part of the analysis period *Amstat News* would reproduce posters in their entirety (though very small). Only some reproductions of Australian posters and projects are linked to the AMT website.

RESULTS

The results will be presented in the form of answers to commonly posed questions from teachers making decisions about whether a poster competition is a useful part of their teaching program.

To what extent are the posters interdisciplinary?

In Australia, 70% of the posters were from a single curriculum area. In the US, 85% were from a single curriculum area. This suggests that, taken individually, the posters are not all that interdisciplinary after all. However, bearing in mind that the authors decided that no poster was to be allocated more than two curriculum areas, there is still plenty of scope for a poster to address multiple curriculum areas at once.

What were the double topics?

In Australia, 26 posters were allocated to two areas. Most of them were HPE/Science (8), HPE/Technology (5) and Science/Technology (4). In the US, 52 posters were allocated to two areas.

Most of them were also HPE/Science (12), Science/Technology (12) and for something different, Geography/Science (7). Weather and climate were captured under Geography/Science which is, of course, a topic of perennial interest at all educational levels.

Does the frequency of subject areas vary by country?

In Table 1, the unweighted frequencies count every time a topic was allocated. The weighted frequencies allocate a count of 0.5 to the two topic areas of the double topic posters. Curriculum areas are listed in descending order of unweighted popularity in Australia then the US.

Table 1. Relative frequency of curriculum areas

Curriculum Area	Australia		US	
	Unweighted	Weighted	Unweighted	Weighted
HPE	38	42	41	43
Science	23	22	21	20
Technology	17	16	13	12
Economics	7	7	5	4
Geography	4	4	5	4
English	4	4	2	2
The Arts	4	4	1	<1
Civics	3	1	2	3
Mathematics	0	0	6	7
History	0	0	3	3
Inconclusive	0	0	1	2
Total	100 (n = 110)	100 (n = 84)	100 (n = 419)	100 (n = 367)

The two countries do not differ much in terms of the frequency of the most subject areas. HPE, Science and Technology top the list followed by Economics. Amongst the more rare subjects, the US produced more posters on Geography, Civics, History and Mathematics than Australia.

Do the unweighted percentages of each topic change over time?

Line charts of the unweighted percentages of each topic over time (from 2000 in the case of the US competition and from 2008 in the case of the Australian competitions, not shown) are very hard to interpret. There is some evidence that the number of Science posters is falling, while the number of Technology posters is rising. There is very little difference in trend between the US and Australia.

Do the unweighted percentages of each topic differ by age group?

Again, it is very hard to interpret patterns in tables of counts by topic and age group (not shown). There are more entries at the younger age groups, so adjusting for the number of entries, in the US the number of HPE topics increases as students get older, while Technology falls. In Australia, the trend is almost completely reversed, with older students taking up Technology topics, while HPE falls. In both countries, the number of Science topics falls as students get older. Because of changes in the Australian system of awarding prizes, there are very small numbers involved at certain levels (for example, only two posters at senior level) so the data available provide only weak evidence of patterns.

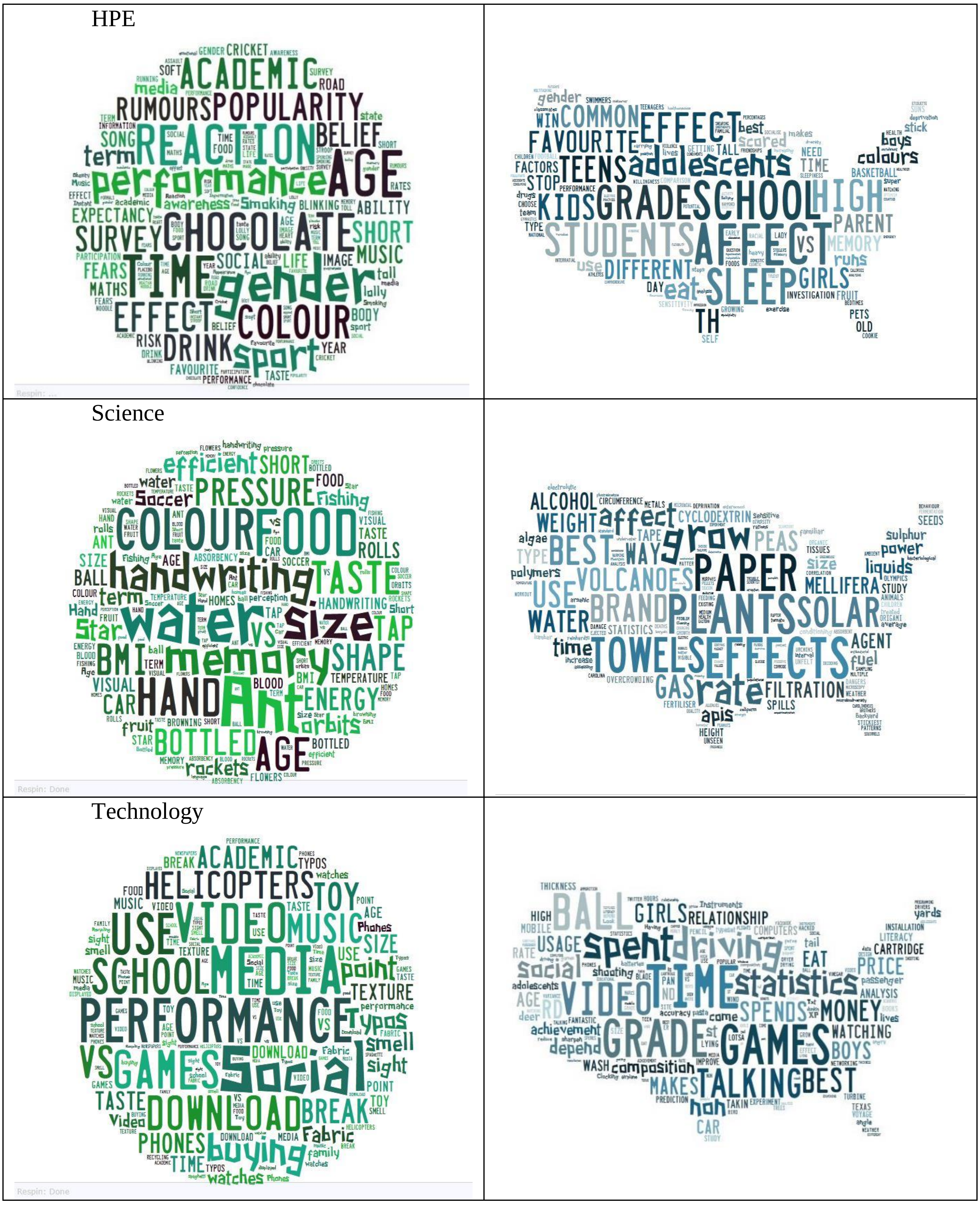
What were the six inconclusive titles?

There will surely be much interest in the six titles which could not be easily categorised. Interestingly, they were all from the US, suggesting that either US students are encouraged to use punning titles or that there are stories of local interest that were picked up by students but of which the authors simply were not aware. The titles were “Which seed wins?”, “This bud’s for you”, “New England’s exploding!”, “Was it a good move?”, “What are you searching for?” and “A shoe for you”.

What visualisation captures the breadth and depth of the titles?

The aim of Figure 1 is to capture the information contained in the titles of posters but to move beyond a list of titles. The word cloud is the most engaging way to present the topics from previous competitions to teachers in those curriculum areas. Viewers can note that the larger words are those that occur most often in titles.

Figure 1. Word cloud (tagxedo.com) of titles of projects that were allocated 100% to HPE, sciences and technology by country (left column = Australia, right column = US).



DISCUSSION

The discussion will focus on the curriculum areas first, then address the impact of the National Students Statistics Poster Competition (NSSPC) on the data that will be available in the future.

Clearly there are a "top three" curriculum areas of HPE, Science and Technology. Along with Figure 1, the titles of posters in these areas suggests that the possibilities of studying "the effect of x on y in z" offers a wide range of possible investigations that will capture the interest of a wide range of students. The remaining discussion of the "other" curriculum areas is in alphabetical order.

The Arts. It is possible to conduct a statistical investigation on an arts topic. Indeed Rosen (2014) made the point was that learners should "be encouraged to think of the arts as including or involving investigation, invention, discovery, play and co-operation and to think that these happen within the actual doing, but also in the talk, commentary and critical dialogue that goes on around the activity itself."

Civics and Citizenship. Most of these posters were from the US and often used data on US Presidents. Most of the posters on legal issues were actually allocated to HPE as they were either about opinions or about health and safety e.g. seatbelt laws, drug use.

English. Zipf's Law (K-3) and target age/sentence length in magazines (9-10) came up more than once and both offer a strong combination of quantitative analysis and texts to capture the interest of language students.

Economics and Business. Some posters were clearly secondary data analysis, so that "Making money: trends of the NASDAQ 2000" probably could have been allocated to Mathematics (see later) but the interest in the stocks was probably enough to support the allocation to Economics. On the other hand, the active data collection investigations in these posters tended to be around contents of piggy banks (K-3) or pricing a "basket" of supermarket goods (4-6). The second author's experience is that a project around the "basket" of goods would have taken place in a school mathematics class a few years ago, but is highly likely to have got pushed out due to time limitations.

Geography. The Australian curriculum is changing over time, and what teachers may remember learning in geography may not be what is being learnt now. The weather, climate change and pollution have all been represented in the posters studied, and there is plenty of opportunity to suggest topics like these to interested students. Some could involve secondary data analysis e.g. temperature records, and some could involve active data collection e.g. comparing forecasts to outcomes over a one-month period.

History. Science is encroaching upon history through shows like the BBC documentary "The King in the Car Park" which brought us the science behind identifying human remains in a Leicester carpark as those of the King Richard III of England. The history topics investigated in the data available were about dinosaurs, or comparisons of data from historical events such as wars or ship sinkings.

Mathematics. This is the subject where Statistics finds its "home" at school, and to some extent that is turned into practice in terms of entries into poster competitions. Some of the posters classified under Mathematics were secondary data analysis where the focus was on the probabilities/frequencies observed rather than the intrinsic nature of the numbers. For example, a poster on "Do the M&M colours you get match what they promise?" would not have been focused on the M&Ms as such (except for consumption at the end of the exercise!), but in the frequency distribution itself. "State rankings vs private schools in the US" looked as if it was probably a secondary analysis of something like MySchool data here in Australia, with (likely) lesser interest in the rankings themselves than in the mechanics of the data analysis. The statistical focus of the Mathematics projects is fairly uniformly split between describing distributions, of everything from jellybean colours to birth rates, languages spoken; and making comparisons, from temperature forecasts to school rankings to waiting times. Most interestingly, all projects classified as 100% mathematics were from the US.

An important recent development in poster competitions in Australia was the arrival of the NSSPC, sponsored by the SSAI. The NSSPC was piloted for schools in the Hunter region in 2014 then launched nationally in 2015. The NSSPC model involves mentors for student groups. The NSSPC only accepts posters, but this has the advantage of allowing winning entries to proceed on to

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

the International Statistical Literacy Project poster competition (http://iase-web.org/islp/Poster_Competition_2016-2017.php). The winning posters are displayed at https://www.ssaipostercomp.info/winners2014.html which shows that the top three categories of HPE, sciences and technology were strongly represented. The outliers were one Economics "Aldi versus Woolworths" and one Geography/Arts "What are the representations of ethnicities in prime-time TV advertisements". It is likely that mentoring will have influenced the level and sophistication of the statistical content, which will be examined in future research.

It may be that this research has pushed a layer of categorisation on top of these posters that was never meant to be there, especially given that the period of time spanned by these posters extends back beyond the changes to the Australian Curriculum in the last five years. Furthermore the US posters have been categorised according to a curriculum scheme different to that under which they were produced. On the other hand, the curriculum areas used to categorise the projects were broad enough that the main aim of this research has been met, which was to provide an evidence base of successful topics in a way that is highly inspirational to Australian teachers now.

CONCLUSION

In this paper we have examined evidence in oral history and published literature regarding the history and pedagogical value of poster competitions. We have collected and analysed novel data from two countries, to provide an evidence base that should encourage subject teachers to raise the uptake of poster competitions at all school levels. Australian schools are increasingly scheduling cross-curriculum activities that can be the catalyst for a conversation about the relationships between curriculum areas, along the lines of conversations already under way between Science, Technology, Engineering and Mathematics. In the words of a referee, these "topical themed investigations … provide the freedom, flexibility and scope that many teachers seek in promoting a typically 'dry' subject to disinterested students". Teachers of Health and Physical Education, Science, and Technology can all draw much inspiration from this research. Economics and Mathematics teachers are also well served. If other teachers can be encouraged to look internationally, or be provided with linked information on Australian sites, there is something for everyone in the end, because every curriculum area was represented at least once.

REFERENCES

Barnicoat, J. (1972). *A Concise History of Posters*. London: Thames and Hudson.

Bernreuter, M. (1995). Poster competitions: another way to increase university-service interchange. *Journal of Nursing Administration* 25, 8 - 9.

Cooter, R. and Stein, C. (2007). Coming into focus: posters, power, and visual culture in the history of medicine. *Medizinhistorisches Journal* 42, 180 – 209.

Cox, J. (2016). A History of Science Fairs. http://www.streetdirectory.com/travel_guide/118515/science/a_history_of_science_fairs.html. Accessed July 5, 2016.

Rosen, M. (2014). How we teach the arts is as important as the fact we're doing it. *The Guardian* https://www.theguardian.com/teacher-network/zurich-school-competition/teach-arts-michael-rosen-education-worthwhile-students?CMP=share_btn_link. Accessed July 5, 2016.

Rowe, N. and Ilic, D. (2009). What impact do posters have on academic knowledge transfer? A pilot survey of author attitudes and experience. *BMC Medical Education* 9, 71.

GUIDANCE FOR TEACHING R PROGRAMMING TO NON-STATISTICIANS

LANGAN, Dean and WADE, Angie
Great Ormond Street Institute of Child Health,
University College London, London
d.langan@ucl.ac.uk

The Centre for Applied Statistics Courses (CASC) at University College London (UCL) provide short courses on statistics and statistical software packages. Popular day-courses include a well-established 'Introduction to R' course and the newly developed 'Further Topics in R'. In the latter, attendees are taught intermediate-level topics such as loops and conditional statements. Attendees range from postgraduate students, academic researchers and data analysts in the private sector without a strong background in statistics or programming. First, we highlight some issues with providing our training course to this demographic, derived from our experience and from anonymous online feedback. Second, we discuss some of our solutions to these issues that have shaped our course over time. For example, one issue is catering to a wide audience from differing fields, different levels of computer literacy and approaches to learning. To address this, we prepare for a high level of flexibility on the day and include intermittent practical exercises to get real time feedback on the abilities of attendees. Finally, we reviewed the experiences of other teachers on similar courses documented online and compared these experiences with our own. We offer guidance to other teachers running or developing courses for intermediate-level R programming.

INTRODUCTION

In a data driven society, it is becoming increasingly important to have both statistical and computer programming literacy. However, students may not have anticipated this when choosing their area of study, nor professionals when choosing a direction in their career pathway. This can lead to a barrier and the student or professional consequently finding themselves in a position where they need to attend a statistics or computer programming course. For example, university researchers may have a firm grounding in their chosen field, but need to acquire knowledge of statistics and the use of appropriate software to perform analyses, in the absence of having a qualified statistician as part of their project team. In particular, researchers may benefit from having computer programming skills in languages such as R (R Core Team 2016).

The UCL Centre for Applied Statistics Courses (CASC) currently run 16 different short courses for statistics and statistical software (http://www.ucl.ac.uk/ich/short-courses-events/about-stats-courses). The courses are primarily aimed at those without a strong background in statistics or programming but who require a basic knowledge for their research. We market to those that often have limited time for learning in their busy work schedule. Attendees are usually diverse, including university students, internal researchers and external working professionals. Generally, our courses use real data examples and little mathematical formula, an approach also effective for adult learners in the workplace (Westbrooke & Rohan 2012).

We frequently offer a one-day *Introduction to R* course for those with no prior experience in R and only a basic knowledge of statistics. In addition, a newly developed one-day course named *Further Topics in R* is aimed at those who have already been on our introductory course or learnt the basics of R elsewhere. Both courses are held in a classroom with a projector screen and access to 20 desktop computers, but attendees are welcome to bring their own laptops. Given that our attendees often have limited prior experience of programming or statistical analyses, these R courses are taught in such a way that little prior knowledge of either is required, allowing the participants to concentrate on learning the software. Mascaró et al. (2014) also advocate this approach, stating that computational tools and software "generally not only do not help but rather hamper the learning of the statistics concepts as well as of the use of the tools".

In this paper, we focus on our *Further Topics in R* course, give an account of its development and discuss the lessons learnt along the way. We compare our experience with experiences that other teachers of similar courses have documented online. We provide guidance for teaching R to adult learners without a strong background in statistics or computer programming. Most of the guidance is also applicable to teaching other statistical programming languages.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

COURSE STRUCTURE

Topics covered in this course are considered accessible only to the intermediate level programmer. We cover how to:

- Organise and merge multiple datasets.
- Write conditional commands (where code is only executed subject to a certain statement being true).
- Write loops (a way to repeat a sequence of commands under certain conditions).
- Create new functions (a function contains hidden code that serves some purpose, e.g. the function called *mean*, as you might guess, calculates the mean).

LEARNING OUTCOMES

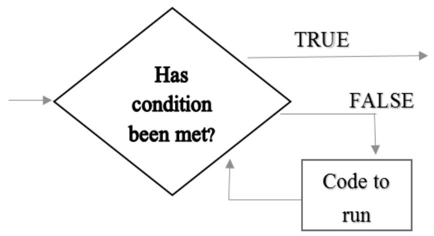
Attendees at the end of the course should have a basic understanding of how R code is structured within these topics above, and have had the opportunity to practice writing code for themselves. However, the primary aim of the course to address any fear that attendees might have of intermediate-level programming and make them more comfortable with further independent learning and the topics outlined above form a framework for this process. Given that a single day is not nearly enough time to cover everything that R has to offer, or even become completely comfortable with all of the topics we cover, it is important that we convey this more general aim. The course also aims to motivate the learner to continue using R for their own project and further develop their skills independently, even if they do not make direct use of the course material. This is achieved by emphasising the breadth of potential that R has. The course reflects the way the first author learnt R programming over the last seven years, which involved formal training only in the beginning but a desire instilled to continue learning through experience.

METHOD OF TEACHING

Each section starts with explaining how the material might be useful for a typical statistical analysis. For example, in the section on loops, we explain that writing a loop in your code can help to produce the same graph for all variables in a dataset. It is important to emphasise 'why this is useful' throughout the course so that attendees feel extrinsically motivated (Coleman 2007); this is known to have a bigger effect on learning in adults than in children (Ormrod 2006). Without relevant examples, there is a danger that the topics covered can seem to be of little practical use.

Next, we introduce the general structure of code and include a diagram to show how this is processed in R. For example, to create a loop in R using the *while* command, the general code structure is:

```
while (conditional statement) {
      ...code to repeat...
}
```



We have found it is more effective to show the general code structure briefly so that any examples that follow can be related to the general context. However, emphasis should be on learning through examples of increasing complexity. We make the first example as simple as possible, even if it does not have any practical application. For instance, we show how a loop can output the numbers 1 to 5:

```
i <- 1 # define a starting value for i
while (i<=5) {
      print (i)
      i <- i + 1 #add 1 to i for every loop
}
```

A loop is not strictly necessary to produce the numbers 1 to 5, but it leads to further examples where loops are required and these are based on real data that attendees find relevant. This approach means that many attendees can more readily transfer the knowledge gained in this course to another context. Example data can be highly tailored when all or most attendees are from the same university department or workplace (Westbrooke & Ellis 2014), but this is not possible for us given our attendees are more heterogeneous. Empirical research has shown students do not naturally generalise what they have learnt to new scenarios (Lovett & Greenhouse 2000, Reed et al. 1985). Using relevant data for R is also advocated by Eglen (2009). For the final example on loops, we go through step-by-step how a loop can be used to read in many datasets from the same folder, explaining that this saves time and prevents errors.

Each section ends with a short exercise to give attendees the opportunity to practice and allows time at the end to go through the solutions as a group. As Lovett and Greenhouse (2000) explain, "students learn best what they practice and perform on their own" and "learning is more efficient when students receive real-time feedback on errors". We give attendees one-on-one help during the exercises and this allows us to identify any common issues and errors in their programming code. Common errors are then discussed as a group, with reasons why the errors occur and how to correct them. This encourages active and collaborative learning, which is known to increase motivation and retention of knowledge (Marriott et al. 2009). An active environment is particularly important for this course, given that the programming language taught has a steep learning curve that can only be tackled by making mistakes and learning from them.

GUIDANCE FOR TEACHING R PROGRAMMING TO NON-STATISTICIANS

In this section, we provide guidance for teaching R programming to non-statisticians, split into nine key points. These are derived from our experience of developing and running the *Further Topics in R* course and from anonymous feedback received from attendees after the course. Feedback is recorded through an online feedback platform (http://opinio.ucl.ac.uk).

Furthermore, we searched online to find where others have documented their own personal experiences from teaching R programming. We reference them here and highlight any similarities and differences from our own experience and guidance. We found four articles (Bååth 2010, Eglen 2009, Mascaró et al. 2014, Westbrooke & Ellis 2014), one blog post (Peruvankal & Muenchen 2014) and one conference presentation (Levy 2016).

1. Recognise variation in learning styles of attendees

We have observed two distinct learning styles from attendees during the taught demonstrations. Some prefer an active approach, which involves writing and executing similar code to that demonstrated on the projector screen. Others prefer a passive approach; they simply observe the teacher and wait for exercises at the end of each section for hands-on practice. Both approaches have their merits and so attendees should opt for the approach which they feel suits them best. However, we received following online feedback "I found the pace was to fast during the taught sections. I struggled to keep up as the teacher typed too fast". The attendees appears to have struggled because they chose the active approach to learning, when they would have been more suited to the passive approach. A similar issue has been observed by Peruvankal & Muenchen (2014) from experience running an introductory course in R. Learning is less efficient when attendees multitask (Ellis et al. 2010), but there is a common misconception that it increases productivity.

To address this issue, we make clear at the start of the course and throughout that, (1) attendees are not required to copy the demonstrations, (2) the code is given to them in the printed course material and (3) there will be time for practice during the exercises. Furthermore, as suggested by Levy (2016), the code on the projector screen can be streamed to the computers of all attendees allowing them to simply copy and paste if they prefer an active approach.

2. Address any timing issues

After our course, attendees are asked in the online feedback questionnaire "Was one day enough time to cover all the material?" We received 25 responses to this question from the first two times the course ran in April and June 2016; 7 (28%) thought that there was more than enough time to properly cover the material, 7 (28%) thought the timing was right and 11 (44%) though there was not enough time. We feel this range of responses to this question reflects the range in abilities of attendees and simply adding or reducing material on the course will not solve this issue. To cater for a wider audience, we split each exercise into several parts that increase in difficulty. Attendees are reminded that completing the whole exercise is not essential as the solutions will be discussed as a group shortly afterwards. We benefit from having at least one other teacher available to offer one-to-one support in a class of size 20. It is important to be flexible on the day, given that the number of attendees that struggle with some of the material cannot be fully anticipated in advance. We have found that having a plan for all scenarios is useful so that topics and exercises can be skipped if there is insufficient time.

3. Address bottlenecks that stop the course progressing

Related to the issue of timing, some who thought the course was too rushed may have thought so because much time was spent loading datasets and setting up the R software ready for use. It is necessary for all attendees to have read in the necessary datasets before continuing with subsequent sections of the course. To minimise the time spent on this task, we give USB drives to the attendees at the start of the course to ensure that everyone is working from the same folder (including those choosing to bring their own laptops). The USB drives contain the datasets and a file that contains all datasets combined (an *Rdata* file). This also ensures that we can promptly help attendees get back on track if there are any computer issues that may result in losing some of their work.

4. Use R output for more effective teaching

From our experience, teaching is always more effective when the emphasis is on the output that R produces, graphical output or otherwise. Attendees are more likely to understand how the code works if they are shown the output first, before it is explained line-by-line. Furthermore, we have found that attendees learn more effectively if the output is visual and interesting. For example, when teaching how to program a loop, we add two additional functions into our examples:

i. *Sys.sleep*: This is a function that sends R to sleep for given amount of time and therefore slows down the speed that the loop is processed. Attendees can then see clearly how the code works.
ii. *txtProgressBar*: Given the loop is slowed down by the function above, it is also useful to output a progress bar so that attendees can visualise how long it may take for the code to fully execute.

Emphasis on graphical output also teaches attendees to explore and visualise their data, which is an important skill in statistical learning (Westbrooke & Ellis 2014).

5. Explain code step-by step

Fortunately, the R programming language allows you to run small sub-sections of code and produce output. This helps to explain how code works. For example, the *if* statement below returns the value 4 because the conditional statement inside the brackets is true.

```
> x <- 1
> if (x==1) 4
[1] 4
```

To decompose how this line of code works, we can extract the conditional statement (`x==1`) and show explicitly this condition is TRUE:

```
> x==1
[1] TRUE
```

This is a teaching style that we found effective from experience. We have since received positive feedback specifically relating to this approach; “A good example is constructing and evaluating code snippets *outside* of the loop and then passing them in”. Deconstructing code in this way is also useful for debugging errors.

6. Avoid copying and pasting pre-written code

We have always programmed live while teaching, as this demonstrates how R code is written in practice. Levy (2016) explains that this style of teaching is more effective because it forces the teacher to slow down to a more suitable pace and ensures they are flexible and responsive to any questions the attendees may have. The difficulty in coding live is that the teacher has an added responsibility to stick to the general principles of programming and set a good example. Guidelines

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

are available (Bengtsson 2009), but have not been formally agreed. R users generally advocate adding comments where the code is not self-explanatory and using consistent indentation (Mächler 2014).

7. Errors are an integral part of the learning process

A consequence of live coding during demonstrations is that accidental errors are likely to occur - counterintuitively, this can be a positive thing. Producing errors are an effective way of learning, but attendees often have a fear them and observing errors during taught demonstrations can help tackle this fear. Mistakes produced during demonstrations can also be intentional. For example, we try to calculate the mean of a string variable to produce the warning message "argument is not numeric or logical: returning NA". Errors promote a more active learning environment as the class can be encouraged to participate in debugging. Similarly, Mascaró (2014) stated "students realised that mistakes were not a problem in R, because feedback of the environment tells them if there is a mistake or they get some illogical answer; this promotes reflection on what went wrong".

8. Teach attendees to teach themselves

The course is designed to encourage independent learning and this can be partly achieved by incorporating errors and teaching debugging skills (as mentioned above). It is also important for attendees to be aware of other material and know how to access the help files that are available for all functions and packages. However, Eglen (2009) highlights a common problem with these help files, stating "students often report that it is hard to discover these functions, as they do not know what to search for". Eglen (2009) advocates the use of *Rseek* (http://www.rseek.org), an internet search tool, for finding relevant functions. We generally find online R forums helpful.

Additionally, to help attendees teach themselves, we demonstrate how code can be developed when consulting a help file is required. This demonstration has the following steps: (1) We explain what we would like to achieve, (2) how to find the help file that we need, (3) how to make sense of the help file and (4) how to incorporate the relevant information in our code.

9. Relate R to other software packages

Learning involves integrating new knowledge with existing knowledge (Lovett & Greenhouse 2000). Therefore, attendees can learn more effectively by relating R to other software packages, explaining their similarities and highlighting their differences (Peruvankal & Muenchen 2014). Many attendees will have experience with at least one other statistical software package. For example, we explain that the R function *cat* works in a similar way to the *concatenate* function in Microsoft Excel. Attendees should also be made aware of the quirks in the R programming language. For example, <− is used to assign a name to an object, where other programming languages would use = (Bååth 2010).

CONCLUSION

In this paper, we described an intermediate-level course in R programming aimed at non-statisticians. The course is one of the more popular courses in the Centre for Applied Statistics Courses, suggesting that many research students and professionals require knowledge of statistical computer programming for their analysis, even those without a strong background in statistics. Its popularity also suggests that there is a big market for face-to-face R courses, despite there being a wide range of free material available online to support independent learning. We aim to address the initial fear of R programming in our course, which is more difficult to achieve through online courses without a teachers support. We have provided guidance for teaching or developing similar courses, derived from our own experience and combined with the experiences of other teachers of similar courses. We included guidance that is most applicable to intermediate R programming; guidance for teaching programming in general (Robins et al. 2003) and statistics (Tishkovskaya & Lancaster 2012) can be sought elsewhere.

Providing a successful training course in R is ultimately about understanding the mind-set of the attendees and anticipating the difficulties with programming that they may have. A post-graduate student, after completing a module in R stated, "I found that teachers can often over-complicate things in an attempt to project their knowledge-base to students. When returning to my notes as an intermediate R user, I realised how important simplicity is when absorbing information. The code could often be simplified." Teachers in a university setting may demonstrate their knowledge of a given subject whilst concentrating less on the intended learning outcomes for those attending the course. The guidance in this paper can help promote simplicity and relevance of short statistical programming courses aimed at non-statisticians and provide motivation for their development.

REFERENCES

Bååth, R. (2010). How Should Programming R be Taught in an Introductory Course in Statistics? Introduction to pedagogy at college level, fall semester 2010, Lund University. Retrieved 15/7/16 from www.sumsar.net/papers/rasmus_baath_teaching_r_programming_01122010.pdf

Bengtsson, H. (2009). R Coding Conventions (RCC) – a draft. Retrieved from www.scribd.com/document/266557294/R-Coding-Conventions-RCC-A-Draft [accessed 19th September 2016]

Coleman, Y. (2007). Tips for Teaching Software – the approach [Web blog post]. Retrieved from metacole.com/2007/12/30/tips-for-teaching-software-the-approach/ [accessed 14th July 2016]

Eglen, S.J. (2009). A quick guide to teaching R programming to computational biology students. *PLoS computational biology*; 5(8), e1000482.

Ellis, Y., Daniels, B., & Jauregui, A. (2010). The effect of multitasking on the grade performance of business students. *Research in Higher Education Journal*, 8, 1.

Levy, M.A. (2016). Teaching R to 200 people in a week. Presented at the International R user conference, Stanford, CA. Retrieved from channel9.msdn.com/Events/useR-international-R-User-conference/useR2016/Teaching-R-to-200-people-in-a-week [accessed 14th July 2016]

Lovett, M.C. & Greenhouse, J.B. (2000). Applying cognitive theory to statistics instruction. *The American Statistician*, 54(3), 196-206.

Mächler, M. (2014). Good Practices in R Programming. Presented at the International R user conference, Vienna, Austria. Retrieved from stat.ethz.ch/Teaching/maechler/R/ useR_2014/ [accessed 14th July 2016]

Marriott, J., Davies, N. & Gibson, L. (2009). Teaching, Learning and Assessing Statistical Problem Solving. *Journal of Statistics Education*, 17(1).

Mascaró, M., Sacristán, A.I. & Rufino, M. (2014). Teaching and learning statistics and experimental analysis for environmental science students, through programming activities in R. In *Constructionism and Creativity-Proceedings 3rd Intl. Constructionism Conf* (pp. 407-416).

Ormrod, J.E. (2006). How Motivation Affects Learning and Behavior. Pearson Allyn Bacon Prentice Hall, www.education.com/reference/article/motivation-affects-learning-behavior.

Peruvankal, J.P. & Muenchen, B. (2014). Secrets of Teaching R [Web blog post]. Retrieved from blog.revolutionanalytics.com/2014/03/secrets-of-teaching-r.html [accessed 14th July 2016]

R Core Team (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. www.R-project.org

Reed, S.K., Dempster, A., Ettinger, M. (1985). Usefulness of Analogous Solutions for Solving Algebra Word Problems. *Journal of Experimental Psychology*, 11, 106–125.

Robins, A., Rountree, J., & Rountree, N. (2003). Learning and teaching programming: A review and discussion. *Computer science education*, 13(2), 137-172.

Tishkovskaya, S., & Lancaster, G. A. (2012). Statistical education in the 21st century: a review of challenges, teaching innovations and strategies for reform. *Jnl of Statistics Education*, 20(2), 1-55.

Westbrooke I., Ellis P. (2014). Training to develop modern statistics in the workplace using R and R commander - experiences from the New Zealand government sector. *Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014).*

Westbrooke I., Rohan M. (2012). Statistical training in the Workplace. In Topics from Australian Conferences on Teaching Statistics 2014 (pp. 311-327). Springer New York.

BIG DATA, DATA SCIENCE, COMPUTER SCIENCE & STATISTICS EDUCATION

KYNG, Timothy[1], BILGIN, Ayse Aysin[2] and PUANG-NGERN, Busayasachee[2]
[1.] Department of Applied Finance and Actuarial Studies, FBE,
Macquarie University, NSW
[2.] Department of Statistics, FSE, Macquarie University, NSW
timothy.kyng@mq.edu.au

Recent advances in Information and Communication Technology (ICT), software and Artificial Intelligence (AI) have facilitated the collection, storage and easy access to vast amounts of data with low cost. The volume of such data exceeds society's capacity to analyse, interpret and use it in decision making. Demand for people with the skills to extract information from the data is growing strongly. Some statisticians are of the view that Data Science is mostly hype and no more than a new name for statistical science. The perceptions of industry and government may be very different. The term "data scientist" appears frequently in recent job advertisements. For most statistical practitioners statistical and other software tools are essential for their work. Software has been developed to filter, access, visualise, and analyse data, handle large datasets, recognise patterns in it, even formulate hypotheses. Computer scientists and statisticians have been working independently of each other in this area. In this paper we examine what statistical skills and software tools are needed by graduates and the differences between the perceptions of practitioners and academics about these issues. The paper also investigates relevant degree programs in Australia regarding the coverage of statistics in data science degrees and of software skills in statistics degrees.

INTRODUCTION

The convergence of computer and communications technology, advances in software and hardware and the widespread use of the internet have led us to the age of Big Data. Massive volumes of data are being collected and this outstrips many organisations' capacity to analyse the data to obtain useful information. The need for people who can analyse the data and obtain useful information from it has led to a new career path: the "data scientist" - a kind of hybrid of a statistician and a computer scientist / IT professional. An expansion of Big Data leads to increasing job opportunities. Gartner (2012) forecasts that the growth in Big Data will produce 4.4 million IT jobs globally. Bradshaw (2013) also says for each new IT job created there will be three new non-IT positions.

The data is obtained from many different sources. Your smart phone and other mobile devices can be used to track your movements and your internet viewing habits. Recently LG and Samsung launched new "smart TVs" onto the market. Recent publicity about these indicates that while you are watching your new smart TV, it is also watching you, and possibly recording your conversations, tracking what programs you watch on TV and transmitting the information back to the manufacturer who can then sell the data to third parties. It perhaps is a case of life imitating art, the art here being George Orwell's famous book "1984" (1949) about a dystopian future. Perhaps in retrospect it should have been titled "2014". Instead of one "big brother" who is watching us we now have lots of "brothers" (both corporations and government agencies) watching us and recording what we are doing (Lewis 2015, McAllister 2013).

The internet age and the massive collection of data have led us to an age of mass surveillance of the population, mainly for the purpose of marketing. Big Data and its analysis is an area in which industry is leading and academia seems to be playing catch up. Universities around the world are responding by introducing new degree programs in data science / big data analysis. We have identified at least 23 different US universities offering such degrees. The University of Washington offers (since 2013) a new PhD in Big Data with collaboration between the Computer Science & Engineering and Statistics departments. In the UK, many universities are offering Big Data Analytics courses and some of these are recently established. Bournemouth University, Sheffield Hallam University, and City University London all commenced offering such degrees in 2014. In Australia, at least 11 Universities have recently established Postgraduate

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

coursework data science courses. In NZ the University of Otago and the University of Auckland have been offering Data Science degrees since 2013.

This paper presents the perceptions of a sample of recent graduates working in big data related jobs and of academics' teaching in related disciplines, about which skills and software tools are considered to be important. In addition the similarities and differences between these groups' perceptions are presented. This study is an exploratory study to identify what could be incorporated into the Statistics and Data Science degrees.

STATISTICS: IS IT SEXY?, WHAT ABOUT DATA SCIENCE, IS IT SEXY?

The answer to this rhetorical question is no and yes respectively! Many fellow statisticians' experience is that telling others at a party you're a statistician produces the same reaction as telling them you have a nasty disease! They steer clear of you. Data Science, however is sexy, or is perceived to be by the media and employers. Here are some recent headlines: "Data scientists are the rock stars of business" according to the Sydney Morning Herald (22/8/2016), "Data Scientist: the sexiest job of the 21st century".

For a recent talk (10th March 2015) to the Statistics Society of NSW, Professor Thomas Lumley of the University of Auckland wrote "Mainstream statistics ignored computing for many years ….. Practical estimation of conditional probabilities and conditional distributions in large data sets was often left to computer science and informatics. Although statistics started behind, we are catching up: many individual statisticians and some statistics departments are taking computing seriously. More importantly, applied statistics has a long tradition of understanding how to formulate questions: large-scale empirical data can tell you a lot of things, but not what your question is. Big Data are not only Big but Complex, Messy, Badly Sampled, and Creepy. These are problems that statistics has thought about for some time, so we have the opportunity to take all the shiny computing technology that other people have developed and use it to re-establish statistics at the centre of data science". Glance (2013) pointed out that a combination of computational, statistical and mathematical skills are required in Big Data analysis and data visualization. So expertise in one discipline is not enough.

METHODOLOGY AND DATA

This study was confined to Australia and New Zealand due to cost and time constraints. Our objective was to identify important areas of generic expertise and particular software skills required for graduates working in the big data analysis and to compare the perceptions of academics and graduates about these issues. We surveyed both: (i) academics in relevant university departments and (ii) graduates in the data analytics work force. We also investigated recent job advertisements for jobs in data analytics, data science, big data analysis. This partly informed the design of the survey questions. University departments and academics are identified via publicly available information on the web. Identification of graduates working in big data analytics was much more difficult as this information is not readily available. Accordingly snowball sampling was used so that a larger number of graduates can be reached. This has the advantage of creating a bigger sample but the possible disadvantage of introducing bias into the sample.

The surveys were conducted through online questionnaires via the Qualtrics Survey software. They were separately designed but included questions covering similar issues to facilitate comparison between the perceptions of academics and graduates. Some questions required selecting a rating from strongly disagree to strongly agree using a five point Likert scale for various statements, others required selecting from among the available options. We also had one open ended question in both surveys for additional comments. Both surveys had two questions about expertise required for employment in the Big Data field, two questions about software tools and skills; and four questions asking for demographic information about the respondents. The academics' survey included questions about their workplace, their experience in the Big Data area, and about degree programs and subjects offered. The graduate survey included questions about participants' education, workplace information and their opinions about the Big Data / Data Analytics roles.

The academics' email addresses were collected from 39 Australian and 8 New Zealand universities' websites in April - May 2014. The target group was the academics in the Statistics, Computer Science, Actuarial Science, Information Systems, Information Technology, Mathematics, and Marketing disciplines. There were 127 Departments in Australian Universities, 31 in New Zealand Universities. From 158 university departments sampled, 62 university departments responded. In addition, 5 responses were received from the United Kingdom, Austria, Canada, and the United States.

The email list of prospective participants for the graduate survey was gathered from two universities: Macquarie University (Australia) and the University of Western Sydney (Australia). The graduates in our list included people who graduated between 2003 and 2014 with a Bachelor degree, PG Certificate, PG Diploma, or Master's degree. Current Master's degree students from Macquarie University were also included. The target disciplines were the same as academics, except Marketing. 497 invitations were sent out which asked graduates to forward the invitation to people who they think would be in the target population. A total of 72 responses were received from graduates.

Many of the questions in the surveys required respondents to give an importance rating on a 5 point Likert scale. The average importance ratings and rankings are calculated to compare the perceptions of academics and graduates. In addition to these exploratory analysis of the academics and graduate, we also investigated the level of statistical content of the new coursework Data Science Postgraduate degrees and the level of software skills covered in the established coursework Statistics and Actuarial Science Postgraduate degrees on offer in Australia to form a view on how Data Science education is currently evolving and the implications of it for statistics education.

RESULTS

Ten areas of generic skills were included in the surveys. These skills were Statistical Analysis and Software Skills, Statistical Learning, Mathematics, Data Mining, Machine Learning, Artificial Intelligence, Programming, Marketing, Business Analysis and Accounting. We have identified the following seventeen software tools as relevant for inclusion in our surveys: Base SAS, SAS Enterprise Miner, SPSS Analytics, SPSS Modeler, WinBUGS, Matlab, R, Java, VBA, Hadoop, MapReduce, noSQL, SQL, JaQL, Hive, Oracle, & Python (Puang-Ngern 2015).

A comparison of the average values of perceived importance of skills showed a high level of agreement between graduates and academics (Table 1). The ranks for ten categories of generic skills were also very similar between graduates and academics.

Table 1: Average importance ratings and rankings for areas of expertise needed in Data Analytics

	Average Ratings		Rankings	
Expertise Area	Graduates	Academics	Graduates	Academics
Statistical Analysis	4.4	4.6	1	1
Data Mining	4.2	4.2	3	2
Statistical Learning	4.3	4.2	2	3
Programming	3.9	4.0	5	4
Mathematics	3.8	4.0	6	5
Machine Learning	3.8	3.7	7	6
Business Analysis	4.0	3.5	4	7
Artificial Intelligence	3.3	3.3	8	8
Marketing	3.3	3.1	9	9
Accounting	2.9	2.6	10	10

Statistical Analysis was ranked the highest by both groups and accounting was ranked the lowest. Artificial intelligence and marketing were ranked next lowest. Statistical Learning, Data Mining, Programming and Mathematics were ranked in the top 5 out of the 10 categories by both groups. Statistical learning was a clear 2nd to statistical analysis but machine learning ranked much lower.

A comparison of the importance ratings by graduates and by academics for the identified *software tools* was also done. There was far less agreement between academics and graduates over the relative importance of the seventeen *software tools* than there was about the ten generic *areas of expertise* (Table 2). R, Python and Matlab are considered more important by academics than by graduates whereas VBA and SQL were considered more important by graduates than by academics. VBA was considered more important by graduates (rank 4) than by academics (rank 16).The importance of R software was ranked more highly by academics (rank 1) than by graduates (rank 6). SQL was ranked 1 by the graduates and ranked 2-3 by the academics. Python was more highly ranked by academics (rank 2) than by graduates (rank 11). For academics, R was ranked as the most important software tool whereas for graduates the highest ranked software was SQL. Winbugs, Hive and JAQL had the lowest rankings with both groups. SPSS analytics and SPSS modeler both ranked reasonably well overall and ahead of SAS enterprise miner. Base SAS ranked ahead of SAS enterprise miner.

Table 2: Average importance ratings and rankings for software skills needed in Data Analytics

	Average Ratings		Rankings	
Software Tool	Graduates	Academics	Graduates	Academics
R	3.8	4.2	4	1
Python	3.5	3.9	11	2
SQL	4.2	3.8	1	3
Hadoop	3.7	3.7	6	4
MapReduce	3.4	3.6	12	5
Matlab	3.4	3.6	13	6
SPSS Analytics	3.8	3.5	3	7
SAS Enterprise Miner	3.6	3.4	7	8
Java	3.6	3.4	9	9
SPSS Modeler	3.8	3.3	2	10
Base SAS	3.8	3.3	5	11
Oracle	3.6	3.3	10	12
NoSQL	3.6	3.1	8	13
JaQL	3.2	3.1	16	14
Hive	3.3	3.0	14	15
VBA	3.3	3.0	15	16
WinBUGS	3.0	2.8	17	17

Postgraduate (PG) coursework degrees in Data Science or related areas offered by 38 Australian Universities. The entry requirements in terms of statistical background, and the level of coverage of statistics in these PG degrees provided by a mathematics or statistics department rather than by computer scientists or others are listed in Table 3. In the majority of cases, the level of statistical coverage in the degrees is as low as 25% or less. Interestingly and significantly, none of these degrees require any statistics background for entry into the degrees. This suggests that many students are graduating with Data Science Master's degrees with very low levels of statistical literacy, let alone statistical thinking and skills.

Six Australian Universities offer coursework postgraduate degrees in Statistics: Australian National University, Macquarie University, Monash University, University of New South Wales, University of Southern Queensland, and University of Sydney. Seven universities offer coursework postgraduate degrees in the closely related discipline of Actuarial Science: Australian National University, Bond University, Curtin University, Macquarie University, Melbourne University, Monash University, and University of New South Wales. None of these degree programs require any computer science background for entry to the degree.

Table 3: Australian Universities currently offering postgraduate degrees in Data Science

Institution	Stats & general entry requirement	Statistical units / total units
Australian National University	Nil, Degree in anything	4/12-6/12
Deakin University	Nil, Degree in anything	4/12-6/12
Macquarie University	Nil, Degree in Computing	3/12
Monash University	Nil, Degree in anything	1/16
University of New South Wales	Nil, Science or engineering degree	0/12 - 6/12
University of South Australia	Nil, UG degree in IT or Maths	3/12
University of Southern Queensland	Nil, UG degree in anything	2/16
University of Sydney	Nil, UG degree with Cr average	1/8
University of Technology Sydney	Nil, UG degree in anything	Up to 3/12
University of Western Sydney	Nil, UG degree in anything	2/12
Victoria University	Nil, UG degree in anything	0/16

These degree programs teach students to use statistical software tools (e.g. R, SAS, excel) which is generally embedded in the statistics or actuarial units in those degrees. Usually, there would be no programming units taught separately by computer science academics. The authors are aware that there is a move in some universities' actuarial degrees to include data science units. This is happening at UNSW and Macquarie University. The actuarial professional societies in Australia, in the UK and in USA have decided to include training in data science in their professional exams or as part of CPD but this is yet to be implemented. The French Actuarial Society has already introduced a yearlong CPD course for their members covering data science.

CONCLUSIONS

In this paper we have investigated the generic areas of expertise required and the software tools used in the new area of data science / data analytics / big data analysis. This is an exploratory study of the situation at the time of writing and it gives a snapshot of the current situation in Australia and New Zealand, especially in Australia. The types of expertise needed and the software tools used in big data analysis are likely to be the same around the world. Our study indicates that many universities around the world are moving or have recently moved to establish degree programs in Data Science.

We obtained a reasonably high response rate from academics. Despite the limitations of our sampling method for graduates (snowball sampling), the results are interesting and illustrate the similarities and differences between the perceptions of academia and industry (graduates working in industry). The spearman rank correlation coefficients for the rankings were computed and these were 92% for the generic skills and much lower at 57% for the software tools.

For the graduate survey the respondents were anonymous, it is possible that the type of respondent has influenced the results, which may be a limitation of the study. However there was diversity in the sample, graduates had degrees in computer science, statistics, business and marketing and worked in jobs categorised as "data analytics" "financial services" and "other". The results may be used to inform the design of university degrees for both statisticians and data scientists.

In particular we find that expertise in Statistical Analysis / Statistical Software, Statistical learning and data mining are very important. Programming and mathematics came next in the overall rankings. This is good news for statisticians and for the future of statistics as a discipline. Artificial intelligence, accounting and marketing are of less importance. The perceptions of the academics and the graduates in industry about the relative importance of the ten *areas of expertise* are very similar. The authors believe that university education for statisticians needs to change to incorporate a lot more content relating to computers and software to enable their graduates to take their place in industry to analyse big data, instead of graduates of other disciplines.

Regarding the *software tools* used in both academia and industry, we find that the perceptions of graduates in industry diverge from the perceptions of academics much more than for the *areas of expertise* we have discussed. There seems to be agreement as to which of the software tools are ranked as least important, but less agreement as to which ones are the most important. Some software tools are regarded as much more important by academics than by graduates in industry and vice versa. Perhaps this is because some software tools are better for teaching and others are better for industrial applications. License fees for some software tools can be expensive. It could be that the academics are used to particular software tools and prefer to use these to the alternative of learning new software tools. This has implications for the software tools used in teaching statisticians at university.

We are of the view that the current set of Data Science degrees have very limited coverage of statistical science, and this is most likely a result of them being established by Departments of Computer Science (CS) and limited collaboration between academics in the CS and Statistics disciplines. In many statistics degree programs there is a lot of emphasis on mathematical theory of statistics and relevant statistical software skills are embedded into various subjects within a statistics degree. In the big data era, consideration should be given to redesigning statistics degrees to incorporate more coverage of the IT and software tools and skills which data science degrees cover. Perhaps the solution is to allow students to specialise more: in either mathematical statistics, or applied statistics with data science, or other specialisations so that the needs of different types of students are catered for. Perhaps some students don't need as much training in mathematics as would be implied by the current structure of statistics degrees. Many new jobs are being created for data scientists that once would have been for statistics graduates. The number of new jobs for those skilled in mathematical statistics is by contrast much more limited.

We recommend that Data Science and Statistics degrees include more software skills as well as statistical skills for big data analytics. Universities should consider giving students exposure to some of the software tools used in industrial application of data analytics. This is consistent with the Curriculum Guidelines for Undergraduate Programs in Statistical Science written by the American Statistical Association (2015) and as recommended by the President of the Statistical Society of Canada, Prof John Petkau (2014).

REFERENCES

American Statistical Association. (2014, November 15). *Curriculum Guidelines for Undergraduate Programs in Statistical Science.* Retrieved from http://www.amstat.org/education/pdfs/guidelines2014-11-15.pdf

Bradshaw, L. (2013). Big Data and What it Means. *Business Horizon Quarterly, 2013* (7):32-35. Retrieved from: http://www.uschamberfoundation.org/bhq/big-data-and-what-it-means

Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally to Support Big Data By 2015. (2012). In *Analysts Discuss Key Issues Facing the IT Industry During Gartner Symposium/ITxpo 2012*. Orlando, FL: Gartner.
Retrieved from http://www.gartner.com/newsroom/id/2207915

Glance, D. (2013, December 2). Solving Big Data's big skills shortage. *The Conversation*. Retrieved from http://theconversation.com/solving-big-datas-big-skills-shortage-20352

Hurwitz, J., Nugent, A., Halper, F., & Kaufman, M. (2013). *Big data for dummies*. Hoboken, NJ: Wiley.

Lewis, D. (2015). Is Your TV Spying On You? Accessed on 21 August 2016 at http://www.forbes.com/sites/davelewis/2015/02/10/is-your-tv-spying-on-you/#51fa296747f5

McAllister, N. (2013) You THINK you're watching your LG smart TV - but IT's WATCHING YOU, baby. Accessed on 21 August 2016 at http://www.theregister.co.uk/2013/11/20/lg_smart_tv_data_collection

Orwell, G. (1949). Nineteen Eighty-Four. New York: Harcourt, Brace & Co.

Petkau, J. (2014). President's Message. *SSC Liason, 28*(4), 3-7. Retrieved from http://www.ssc.ca/sites/ssc/files/liaison/liaison-28-4.pdf

Puang-Ngern, Busayasachee. (2015). Big data and its implications for the statistics profession and statistics education. (Unpublished Master of Research Thesis). Macquarie University, Sydney, Australia.

MEASURING LEARNING WITHIN A LARGE DESIGN RESEARCH PROJECT

HARRAWAY, John[1], FORBES, Sharleen[2], and DRYSDALE, Megan [1]
[1] University of Otago, [2] StatEd Consultancy
jharraway@maths.otago.ac.nz

Conceptual learning of students from universities, schools and the workplace taking part in a research project to develop new material on bootstrapping and randomisation is investigated. The aim was to develop teaching strategies using dynamic visualisation software. Before and after instruction, students sat tests including questions on sampling, confidence intervals, bootstrapping and randomisation. Multiple-choice, true/false, yes/no and open-ended questions were used. Student performance is analysed in terms of the percentage of correct answers; open-ended questions are only analysed to a limited extent. For questions that are common to both the pre-test and post-tests, increases in correct answers and changes in responses are recorded. The percentage correct in the pre-test and post-tests varied widely and most responses (65%) to the common questions were unchanged. Over a third of the changes were from a correct to an incorrect answer. Pre- and post-tests can measure student understanding and prior skills, but multiple-choice and true/false questions may not be adequate for this purpose. The difficulty of teaching new material while introducing new software, the relevance of questions to the taught material, appropriateness of multiple-choice and true/false questions in assessment, and levels of learning or unlearning acceptable to teachers, are discussed.
KEY WORDS: Statistics Education; Bootstrapping; Randomisation; Measuring Changes in Performance

1. PURPOSE OF THE RESEARCH PROJECT

The purpose of this project was to investigate the conceptual statistics learning gained by students taking part in a large design research project (Cobb et al. 2003). The overall purpose of this design research was to develop new learning trajectories, resource materials, and dynamic visualisation software for new material on bootstrapping and randomisation being introduced into New Zealand's senior secondary schools (http://seniorsecondary.tki.org.nz/Mathematics-and-statistics). This was followed by implementation with students and retrospective analysis which resulted in modification of teaching materials (Pfannkuch et al. 2011, 2013). The principles which drove the design of the visualisation software were to avoid cognitive overload, direct attention to salient features, build familiarity before introducing new concepts and combine pictorial, verbal and movement elements in key actions.

Visual simulation software, iNZight (now freely available on line https://www.stat.auckland.ac.nz/~wild/iNZight/), was developed by Professor Chris Wild of the University of Auckland for use in teaching bootstrapping and randomisation as part of the research design. While some students may have been exposed to simulation software in other subjects, it is likely that this type of software would be a new experience for the majority of the students. Both randomisation and bootstrapping would have been unfamiliar to almost all the students. Informal inferential language, as applied in the work of Arnold and Pfannkuch (2010) was used throughout the instructions, for example, '*making a call*', '*a fairly safe bet'* and, to explain the null hypothesis, '*chance acting alone or together with other factors*'.

The students were all given a hands-on task before moving to the software. Use of hands-on instruction is reported positively in other studies (Forbes 2014, Holmes and Jemes 2008; Payne and Dawson 2014), and especially with regards to statistical inference (Rossman 2008). The bootstrapping hands-on exercise in this project involved groups of students taking a sample of size 9 from a bag of paper slips, each representing the weight of a student in a university class of 521 students (population), then resampling with replacement from their sample many times calculating the median on each occasion and building an interval of values for the sample median weight. Randomisation was restricted to its use within experimental design. The hands-on exercise involved students physically tearing apart pieces of paper on which were recorded babies' walking ages and allocated group (treatment/exercise or control), then randomly reassigning the babies and recalculating the difference in the mean walking age.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

2. IMPLEMENTATION OF THE DESIGN RESEARCH

The design research project was conducted over 2 years and the main study involved 2765 students from throughout New Zealand; 14 year-13 secondary school classes, seven introductory first-year classes at two universities and one group of students from a statistics workplace with the learning format differing according to level (schools learning in short teaching sequences outside normal classes); university students in normal lectures with tutorials; and workplace students in a full-day workshop. The data analysed here were pre- and post-tests. There were two versions of the post-test, one focussing on bootstrapping (post-test A) and the other on randomisation (post-test B), to which students were randomly allocated. Only one of the common questions (question 6) was specifically related to the new material (randomisation). Questions are denoted by *A* if they only occur in post-test A (bootstrapping) and by *B* if they only occur in post-test B (randomisation). The actual questions used are available on the following website: http://www.maths.otago.ac.nz/files/TLRI, and are also summarised below:

Questions common to the pre-test and both post-tests:

- *Question 3* required interpretation of sampling variation. This question had an open response in addition to a *yes/no* answer but only the latter was analysed here.
- *Question 4* explored the impact of *sample size* on the width of a *confidence interval*.
- *Question 5* also explored the impact of *sample size* on the width of a confidence interval.
- *Question 8* posed three statements on interpretation of a *confidence interval* and asked students whether each was TRUE or FALSE.

Common to the pre-test and post-test B (randomisation):

- *Question 6* investigated the reason for random assignment.

Post-test A only (bootstrapping):

- *Question 7A* investigated the meaning of a bootstrap distribution from 1000 bootstraps.
- *Question 8A* asked students to interpret the bootstrap confidence interval with space provided for a short sentence.

Post-test B only (randomisation):

- *Question 8B* asked students to give the tail proportion in a given diagram of the re-randomisation distribution as a fraction.
- *Question 9B* involved interpreting a bootstrap distribution for 1000 re-randomisations.
- *Question 10B* asked students to select which of a set of possible reasons contributed to researchers being able to conclude that a *causal* relationship existed. (could circle more than one).
- *Question 12B* investigated understanding of the tail proportion.

Confidential unique identifiers were used to match an individual student's responses between common questions in the pre- and post-tests. This paper reports on the quantitative results from common questions in the pre- and post-tests, together with multiple-choice questions and those requiring simple written answers in the two post-tests. Students did not receive feedback after the pre-test. Students' answers were categorised as correct, incorrect or non-response. Increase (or decrease) in the proportion of correct answers between the two tests was analysed together with the proportion of students whose responses remained unchanged (status quo), changed from incorrect to correct, or changed from correct to incorrect (providing a simple measure of the learning gained by students).

3. RESULTS

After data cleansing (removal of duplicates, etc.), a total of 2757 students completed pre-tests: 2544 from Universities, 198 from secondary schools and 15 from the statistics workplace. However, not all the students who sat pre-tests completed a post-test. Just over 50% of females completed both the pre-test and one of the two post-tests. This varied by location with one university and the schools having a statistically significant larger proportion of responses from females than males.

Table 1: Number of students by type and post-test completion

Test type	University	School	Workplace	Total
Post-test A	647	66	6	719
Post-test B	658	70	7	735
Total	1305	136	13	1454

A total of 2695 (98%) of students answered a question on previous statistics experience. Most responding students (83%) had some previous statistics learning at secondary school level, with just over half of all the students (56%) having studied statistics in their final year (year 13). As expected, there was a statistically significant difference in reported experience between the university and school groups. One question asked the number of lectures/hours of instruction received by students as part of the project. As expected with students randomly assigned to one of two post-tests, there was no significant difference in the number of lectures attended between these groups. The majority (72%) of university and workplace students received at least 7 hours of lectures and/or tutorials, but the school students were more evenly spread between 3-4 hours (33%), 5-6 hours (28%) and 7 hours (29%). Although the following results reflect the University experience, responses for the three types of student have been aggregated given initial understanding of new statistics concepts and effort on presenting material are similar.

3.1 LEARNING GAINED IN STATISTICS QUESTIONS

There were no questions about bootstrapping in the pre-test. Of the five questions related to bootstrapping in the post-test, only two were analysed, the other three being long open-ended answers. Of the 1030 students doing post-test A, 44% correctly disagreed that the bootstrap distribution was the distribution of the variable in question, most had learnt that the bootstrap distribution was created by resampling from the original sample and that it was used to find an interval of plausible values (88% and 73% respectively) but only 49% knew that it could be used to estimate the variation in the sample mean (*question 7A, parts A- D*). Just over half of the students (56.1%) gave a correct interpretation of the bootstrap confidence interval *(question 8A),* but over 20% simply gave the numerical values for the confidence interval.

Five questions related to randomisation in post-test B, but only one of these (*question 6*) was also in the pre-test. Of the 1040 students doing this post-test, of the 90% who initially gave an incorrect answer in the pre-test, only one in five (21%) changed this to a correct answer in the post-test. Only 10% of students gave a correct answer to *question 6* (that random assignment to the two groups was done to produce groups with similar characteristics) before the teaching took place and this increased to 24% in the post-test. Almost 60% of students knew that the difference in the sample means was not sufficient evidence of a difference between the treatment groups and almost three-quarters knew that the group sizes did not prevent the researchers from being able to conclude that there was a difference between the treatment groups. Roughly half of the students understood that randomisation contributed to the researchers being able to conclude that a causal relationship existed.

In the questions about confidence intervals only just over a quarter of the students correctly read the tail proportion from the given graph, just over half thought that a small tail proportion gave evidence of a difference between the treatment groups and two-thirds thought that high plausibility of chance alone gave evidence of no difference. In *question 12* although any interpretation of the tail proportion of 0.3 that contained the words '*chance acting alone*' or a statement about not rejecting the null hypothesis was accepted as correct, only 6.9% of the students gave the full correct answer in the language taught (*chance acting alone or together with other factors*.) About half of the students with incorrect answers thought that 0.3 was less than 10%.

The seven questions asked in both the pre-test and at least one of the two post-tests were analysed to assess learning, lack of learning, or un-learning resulting from the teaching. Non-responses include students who did not answer the pre-test or at least one of the two post-tests and are not included in the measures of learning, lack of learning or unlearning.

Table 2: Number (and %) of students getting correct and incorrect answers

Question	Responding students				Non-responses	Total
	Pre-test		Post-test			
	Correct	Incorrect	Correct	Incorrect		
3	420 (32)	887 (68)	516 (39)	791 (61)	147 (10)	1454
4	461 (33)	946 (67)	608 (43)	799 (57)	41 (3)	1448
5	425 (31)	938 (69)	535 (39)	828 (61)	87 (6)	1450
6	74 (10)	641 (90)	172 (24)	543 (76)	18 (3)	733
8, part A	355 (31)	807 (69)	296 (25)	866 (75)	290 (2)	1452
8, part B	628 (52)	575 (48)	761 (63)	442 (37)	249 (17)	1452
8, part C	906 (72)	356 (28)	1094 (87)	168 (13)	190 (13)	1452

Overall the percentages getting correct answers on the pre-test were low. The question *(6)* on randomisation, had a markedly lower percentage correct on the pre-test as expected with new material. As shown in Table 2, of the seven questions or parts of questions that were in both Pre- and Post-tests, only one showed a statistically significant decrease (of 6 percentage points) in the proportion of students giving correct answers after the learning. This question involved interpreting a confidence interval. For the five questions with gains these ranged from 7 to 15 percentage points. The most substantial gains were in *question 6* and *question 8, part C*. The first is not surprising as this question was related to the randomisation process in experimental design - one of the two key elements of the teaching. *Question 8, part C* was the only one of the three confidence interval true/false questions where the answer was true rather than false. Looking at data matched on ID, of the students who gave incorrect answers in the pre-test, just over 20% changed to correct answers in the post-test on the sampling and sample size questions (*3, 4* and *5*). For *Questions 8B* and *8C,* on interpreting confidence intervals, similar proportions of students changing from incorrect to correct were found on matched data.

Table 3: Number (and %) of responding students by the type of change in answers

Question	Status quo		Changed		Total responses
	Correct	Incorrect	Correct to incorrect	Incorrect to correct	
3	211 (16)	582 (45)	209 (16)	305 (23)	1307
4	302 (21)	640 (46)	159 (11)	306 (22)	1407
5	237 (17)	640 (47)	188 (14)	298 (22)	1363
6	38 (5)	507 (71)	36 (5)	134 (19)	715
8, part A	144 (13)	655 (56)	211 (18)	152 (13)	1162
8, part B	438 (36)	252 (21)	190 (16)	323 (27)	1203
8, part C	799 (63)	61 (5)	107 (9)	295 (23)	1262

Table 3 gives the proportions of students who gave answers on both tests by whether their answer remained the same (status quo) or changed from either incorrect to correct or vice versa. For each question, over half (57% - 76%) of the students gave the same sort of answer (status quo) in the post-test as in the pre-test. However, those giving an incorrect answer may have given a different incorrect answer than in the previous test. These students can be described as *non-learners.* Across all the questions, 5% -18% of students who had previously given correct answers changed them to incorrect ones (*unlearners?)* and 13%-27% changed answers from incorrect to correct. Of the students who made changes to their answers between the pre- and post-test, an average of 62% over the seven questions changed from an incorrect to a correct answer and an average of 38% over the seven questions from a correct to an incorrect answer.

4. COMMENTS

It is clear that learning did take place in the teaching sessions. The majority of students understood the process of bootstrapping (resampling many times from the original sample and constructing an interval of plausible values). There was also a small increase between the pre- and post-tests in the number of students understanding the purpose of randomisation. After the instruction over two-thirds of students understood what 'high plausibility of the chance alone explanation' meant, that not all relationships are causal in nature and that small group sizes don't prevent causality being investigated when random assignment has been used.

There were a number of true/false questions in which each alternative response was given by roughly equal numbers (within 10 percentage points) of students and the adequacy of multiple-choice questions as an assessment tool has been questioned (summarised in Liu, 2014) and also its fairness across different gender and ethnic groups (summarised in Forbes, 2000). However, here the students were all sitting identical questions in identical settings in the two tests so these biases should not affect measures of change in performance. Liu (2014), analysing an almost identical data set, used Chi-squared tests on these questions to determine that there was strong evidence that student answers were not simply the result of random guessing. It also should be noted that an incorrect answer on a multiple-choice question does not necessarily mean that the student doesn't understand an underlying concept, and vice versa for a correct answer.

With respect to questions common to both the pre- and post-tests, those students who changed from correct to incorrect answers in the common questions may have been guessing or *unlearning* previously learnt material. Another possibility given the tight teaching time (constrained by overall school and university course requirements) is that students were distracted by learning the visual tool and had not been given enough time to familiarise themselves with it before learning or revising the statistical concepts. It may also be that some students were not familiar with the language being used in the questions, especially as a number would have English as their second language. Some students also didn't use the informal inferential language used in the teaching and expected in the answers. Where some students have had prior statistics learning with more formal statistical language, new language may cause confusion.

From a teacher perspective, motivation is provided by the *learning* group of students. This was less than a quarter of students in most of the questions analysed here. This raises a number of questions. Is this sufficiently high? What levels of learning do teachers expect? How acceptable is it to have the students *unlearning* old material when new ideas are introduced?

5. IMPLICATIONS FOR RESEARCH AND PRACTICE:

There were two aspects to the design research in this study; one was the use of dynamic visualisation software to aid students to understand underlying statistical concepts and the other was the introduction of new statistics techniques that would have been unfamiliar to almost all the students. With respect to the first aspect, Budgett and Wild (2014) suggested that experience with the dynamic visual simulation tools used in the teaching appeared to have consolidated abstract inferential concepts. However, they also acknowledged the importance of preceding experience with the simulation tool by hands-on activities.

Overall, these results raise the possibility that the new learning may have been gained at the expense of some 'unlearning' of concepts that had been introduced earlier. Teachers need to ensure that new teaching methods being introduced into classrooms do not introduce new language or methods that conflict with past learning, and that sufficient time is given for students to familiarise themselves with new teaching tools before new concepts are introduced. Teacher motivation is related to the learning gained by their students. Pre- and post-tests can be used to measure student's understanding as well as to determine the skills that they come into the class with. While simple multiple-choice and true/false questions may not be totally adequate for this purpose, they do provide a snapshot of the whole student group, particularly where change in performance is being measured. This analysis suggests that teachers face four types of students; those who already know the material, learners, non-learners and potential un-learners who get confused about material that they previously knew. What is not known is what levels in each of these groups are acceptable to teachers. Without the sort of pre- and post-tests used here teachers may never really know the value added by their teaching.

Further potential issues for discussion include the relevance of assessment questions to the taught material and the appropriateness of multiple-choice and/or true/false questions in assessing students' understanding of conceptual concepts. What is clear is that language matters. Questions should be in as simple English as possible. There needs to be a careful process for moving students from the informal language used here in the teaching to more formal statistical language.

ACKNOWLEDGEMENT

This paper is based upon, but extends, an earlier version, "Measuring Learning within a large design research project" (Forbes et al.) presented to the IASE Satellite Conference, Rio de Janeiro, 2015.

REFERENCES

Arnold P. and Pfannkuch, M. (2010), "Enhancing students' inferential reasoning: from hands on to 'movie snapshots'", *Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS 8)*, 6 pages [online]. *Available at* http://www.stat.auckland.ac.nz/~iase/publications.php?show=icots8

Budgett, S. and Wild, C. (2014), "Students' visual reasoning and the Randomization test." In *Proceedings of the 9th International Conference on Teaching Statistics*, ICOTS 9. 6 pages, [online]. Available at http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_8A1_BUDGETT.pdf

Cobb, P., Confrey, J., diSessa, A.A., Lehrer, R. and Schauble, L. (2003). "Design Experiments in Education Research." *Educational Research*. 32,9-13.

Forbes, S. (2000), Measuring students' education outcomes: Sex and ethnic difference in mathematics, (*Doctoral dissertation*), Curtin University of Technology, Perth, Australia.

Forbes, S. (2014), "Using action research to develop a course in statistical inference for workplace-based adults." *Journal of Statistics Education.* 22(3), 29 pages, [online]. Available at www.amstat.org/publications/jse/v22n3/forbes.pdf

Forbes, S., Harraway, J. and Drysdale M. (2015).Measuring Learning within a Large Design Research Project. *Proceedings IASE Satellite Conference. Rio de Janiero*. (2015)

Holmes, K. and Jemes, A. (2008), "Teaching Statistics and Research Methods: A Collection of Hands-on Activities and Demonstrations", OTRP Online, *Society for the Teaching of Psychology: Office of Teaching Resources in Psychology*. [online], pp17. Available at http://teachpsych.org/resources/Documents/otrp/resources/holmes08.pdf

Liu, J. (2014), "Multi-choice and true/false assessments in introductory statistics: What can they tell us about student understanding?" Unpublished student BSc Honours Project, University of Auckland, New Zealand.

Payne, B. and Dawson, T. (2013). "Hands-on data activities in the classroom - enthusing teachers and students." In S. Forbes and B. Phillips (Eds.) *Proceedings of the Joint IASE/IAOS Satellite Conference, Macao, China.* 5 pages, [online]. Available at http://iase-web.org/documents/papers/sat2013/IASE_IAOS_2013_Paper_1.1.3_Payne_Dawson.pdf

Pfannkuch, M., Forbes, S., Harraway, J., Budgett, S. & Wild, C. (2013). "Bootstrapping" students' understanding of statistical inference. Summary research report for the Teaching and Learning Research Initiative. 18 pages, [online]. Available at http://www.tlri.org.nz/sites/default/files/projects/9295_summary%20report.pdf.

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

Rossman, A. (2008). "Reasoning about informal statistical inference: A statistician's view." *Statistics Education Research Journal.* 7(2), 5-19. International Association for Statistical Education. Available at https://www.stat.auckland.ac.nz/~iase/serj/SERJ7%282%29.pdf

STATISTICAL CAPABILITY BUILDING IN A GOVERNMENT DEPARTMENT - THE NEW ZEALAND MINISTRY OF BUSINESS, INNOVATION AND EMPLOYMENT

ELLIS, Peter
New Zealand Ministry of Business, Innovation and Employment
peter.ellis@mbie.govt.nz

The Sector Trends team in the New Zealand Ministry of Business, Innovation and Employment and its predecessor teams have implemented an intensive program of statistical capability building since early 2012. After a rocky start the program achieved substantial success and by early 2016 was rolling out a standardised set of in-house quantitative analysis training modules providing benefits to around 80 staff so far, with staff from other Ministries asking to join. Training is in R, SQL and Git fundamentals before moving into elements of data visualisation, statistical inference, regression modelling tools and strategies, and specialist areas such as time series modelling and forecasting. The keys to success have included: workplace-specific environment, data and problems; use of more advanced staff rather than external trainers whenever possible; an atmosphere of continuous learning and improvement; and well defined, relatively short modules (three to eight weeks of one 60-90 minute session per week supplemented by "homework") that between them constitute a roadmap of clearly progressing skills. The program - and the results made possible by the high level of capability - show that it is possible in a workplace to foster and apply relatively advanced skills including coding and modern statistical methods.

THE DEMAND FOR DATA SKILLS EXCEEDS SUPPLY

Statistical skills and others relating to the new skill domain loosely called "data science" are at a premium in today's job market. For example, one study by McKinsey Global (Manyika et al 2011) suggests "by 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions." The pressure is felt in other countries such as New Zealand.

In the Ministry of Business, Innovation and Employment (MBIE) and one of its pre-merger predecessor Ministry of Economic Development, I have managed a succession of teams (Tourism Research and Evaluation; Sector Performance; Sector Trends) that have felt the pressure for more and better use of data first hand. The pressure comes directly from Ministers, from internal customers, and external stakeholders, all of whom have heard of the data science revolution. Expectations are raised by high profile developments ranging from the high quality interactive data visualisations in the New York Times, to widespread prediction of economic disruption by big data technologies of industries ranging from transport to retail.

In the case of the teams I have managed, pressures have emerged in four particular forms. These are demands for:

- New, improved, quicker, highly quality-controlled and cheaper *official statistics*, making use of new administrative data sources such as credit card transactions. The New Zealand Tourism Data Improvement Program, which has seen the rebuild online of major surveys and the creation of new datasets such as Monthly Regional Tourism Estimates, is the prime example of our response.
- Better, more frequently updated, more comprehensive *interactive web access* for non-specialists to existing data, meaningfully merged and classified. Our responses have included products such as the New Zealand Tourism Dashboard (MBIE 2016a) and the Regional Economic Activity Report web-tool and mobile-app (MBIE 2016b).

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

- Better management and access for analytical *re-use of internal data*, such as administrative data on science funding and output.
- More insightful use of *statistical modelling and inference* for forecasting, scenario exploration, and predictive analytics. Areas of interest at MBIE have included the need more reliable and granular tourism forecasts, better forecasting of revenues associated with patents, and identification of regional impacts of specific skills shortage.

Building the skills and infrastructure to meet these demands requires more than selecting the right people, or training up existing staff in particular skillsets. There is a need for a new approach to data infrastructure and asset management, a reconsideration of the respective roles of IT and analytics teams with relation to internal and external clients, and a broadening of what "statistical skills" means in the modern workplace.

THE HORIZON NEEDS TO EXPAND WELL BEYOND 'STATISTICS'

An earlier paper (Westbrooke and Ellis 2014) briefly reviewed the limited literature on statistical training the workplace and discussed the experience of MBIE and of the Department of Conservation. We identified the following key needs:

- Managing and manipulating data in ways that are efficient, reproducible and that facilitate peer review and quality control.
- Graphing and data exploration and visualisation.
- Modelling, with emphasis on effect size estimation rather than tests
- Dealing with increasingly large data sets and automated data collection.
- Designing and analysing observational studies and complex surveys.
- Statistical software/computing to support these

This is still a good list but with the benefit of three years more experience I now give much greater emphasis to the parts of that skillset that is not related to statistical modelling and inference. I do not mean the so-called "soft skills" such as communication, stakeholder engagement, project management and so forth; important as these are, they are out of scope for my current discussion. The experience in my teams at MBIE from 2011 to 2016 and my observations from the outside of other analytical teams suggests that statistical modelling and inference is just one of five key dimensions in which even the *technical* skills need to be developed. Those five dimensions are:

- *Statistical modelling and inference* – raising the game from simple cross tabs, to incorporating estimates of uncertainty, more routine use of confidence and prediction intervals, appropriate methods for surveys and time series data in particular, increasingly complex linear and non-linear models, and increasing use of machine learning developments.
- *Statistical computing tools*. Our first efforts to train people in statistical modelling and inference were thwarted by the poor toolset available. It was necessary to fully embrace a code-based environment – we chose R (R Core Team 2016) in combination with extensive use of SQL and (more recently) JavaScript – and accustom people in its use for simple everyday tasks (such as graphics for policy briefings) before resource could be spent on improving statistical modelling and inference.
- *Data storage*. Storage of the data for my team in 2011 was frankly amateurish (a combination of Excel, SPSS, some data files that had never been received from consultants except as aggregate tables in portable document format) and a big investment in IT projects was necessary to establish database infrastructure, and build sustainable extract-transform-load processes and systems. As the teams' mandates widened beyond tourism and many more datasets needed to be at our fingertips, we negotiated with our IT

department a model where IT provided basic infrastructure – “a big, safe swimming pool for you to swim in” as described by the Chief Information Officer at the time – that could be easily scaled up to allow additional data collections to be warehoused without going through gates of change control and project management. A high quality data warehousing team in IT to support the analytics team was only available fairly late in our journey and forms a key component of successful delivery of overall statistical capability, but the existence of such a team does not excuse the statisticians from an obligation to learn to use SQL, the basics of good practice data modelling and data architecture, and the basics of the workings of the infrastructure on which data warehouses sit.

- *Data manipulation*. As the number of datasets we had to simultaneously deal with grew and there was increasing subject-matter scope (growing from simply tourism to also include science and innovation data, sectors and regions, and a range of strategic micro-economic issues), there was growing importance of skills in general data sourcing merging, tidying, reshaping – and in particular the ability to do this efficiently, in a way that scales up, and can be fully reproducible. Hadley Wickham’s “tidyverse” – (Wickham 2016) and (Wickham and Francois 2016) – is a vital component, but so are many elements of the R eco-system that are too diverse to list, particularly including the numerous packages devoted to efficiently obtaining data from the Internet via screen-scraping or Application Programming Interfaces. The ability to efficiently source diverse datasets from the web and combine them in ways meaningful for analysis is a core skill, and not an easy one to maintain because the environment is changing so fast
- *Workflow*. As the team grew in size we had to borrow methods from the software development world. The writings of software development practitioners such as (Spolsky 2000) were highly influential on our approach, as was the Extreme Programming paradigm. Concepts which are now basic to our day to day work include source version control, unit testing, validation testing, continuous (or frequent) integration, automation, standardised style guides, standardised approaches to peer review and peer support. These tools are vital for projects to scale up from “belonging to individuals” to belonging to a team or an institution. I now see adoption of version control as a key indicator of mature professionalization of an analytics team allowing proper quality control, efficiency, scaling up and reproducibility (see (Ellis 2016) for more reflections).

THE KEYS TO OUR SUCCESS MIGHT BE USABLE ELSEWHERE

I wish to highlight four key themes of the actual capability building program that I believe contributed to its success. Underlying these is a meta-theme of the importance of overwhelming management commitment – not just from the team manager, but from a succession of senior managers who, although largely unfamiliar with the technical details, appreciated the size and scale of the transformation needed and provided support. Crucially, in a transformation of this sort output is likely to get worse before it gets better, so convincing stakeholders to wear this cost and risk was a critical part of the program’s success.

Workplace-specific environment, data and problems

One of the biggest challenges for applying external learning (from university or elsewhere) to the workplace is the hurdle of “I don’t know whether and how to apply method X in this situation”, regardless of how well X was learned in a more defined context. A second challenge is the practical question of “how do I get hold of the right data using our tools and data sources”. Our experience with external training suggests that these problems will often prevent take-up in day to day work. These challenges are addressed by using our own data and environment in training, particularly for base level skills. This meant creating our own courses in R and SQL basics even though this appeared to duplicate courses available elsewhere; what is distinctive about our courses is that they apply the skills to the actual problems of the workplace.

Use staff rather than external trainers where possible

We noticed early on that staff preferred to learn from one of their peers. This is particularly helpful for understanding pragmatic choice of techniques and knowing that "yes, this method works for us because I've used it myself". For example, early in the journey several team members had to teach themselves SQL, and their success in subsequently translating their basic skills for others at a point when database experience was light to non-existent was more than was achieved in other experiments with external, better-qualified trainers. We make use of external trainers where off-the-shelf courses are available for advanced topics (like complex survey analysis) and for a smattering of mid-level topics that have a particularly active supply of trainers (such as JavaScript). It is in use of tools like R, data management, and fundamental statistical methods where using peers as trainers has proven much more effective than external provision.

Create an atmosphere of continuous learning and improvement

We identified a danger of training being seen as an affront to existing professional skills. For this and other reasons we deliberately created an atmosphere of continuous creative destruction of the skillset – repeatedly emphasising that everyone from the Manager down should expect current techniques to be outmoded in a year. With rapid developments in the toolkit in the R universe this is easy to justify, but more important than the constantly innovating and ever-more-powerful tools is the expectation of all staff that they will be always learning. To help create this atmosphere, we implemented a weekly 30 minute "Instant Seminar" series as the second half of team meetings, delivered usually by a team member on a topic or technique about which they know relatively more than the others. Each team meeting also begins with a "round the table" where everyone must report "one thing I learned in the last week". Training is promoted to the degree that people see pride rather than shame in being sent to learn skills in a course with "Fundamentals" in its title. The team manual is built as a wiki, expected to be constantly updated; and management allocates priority time and resource to its updating and expansion.

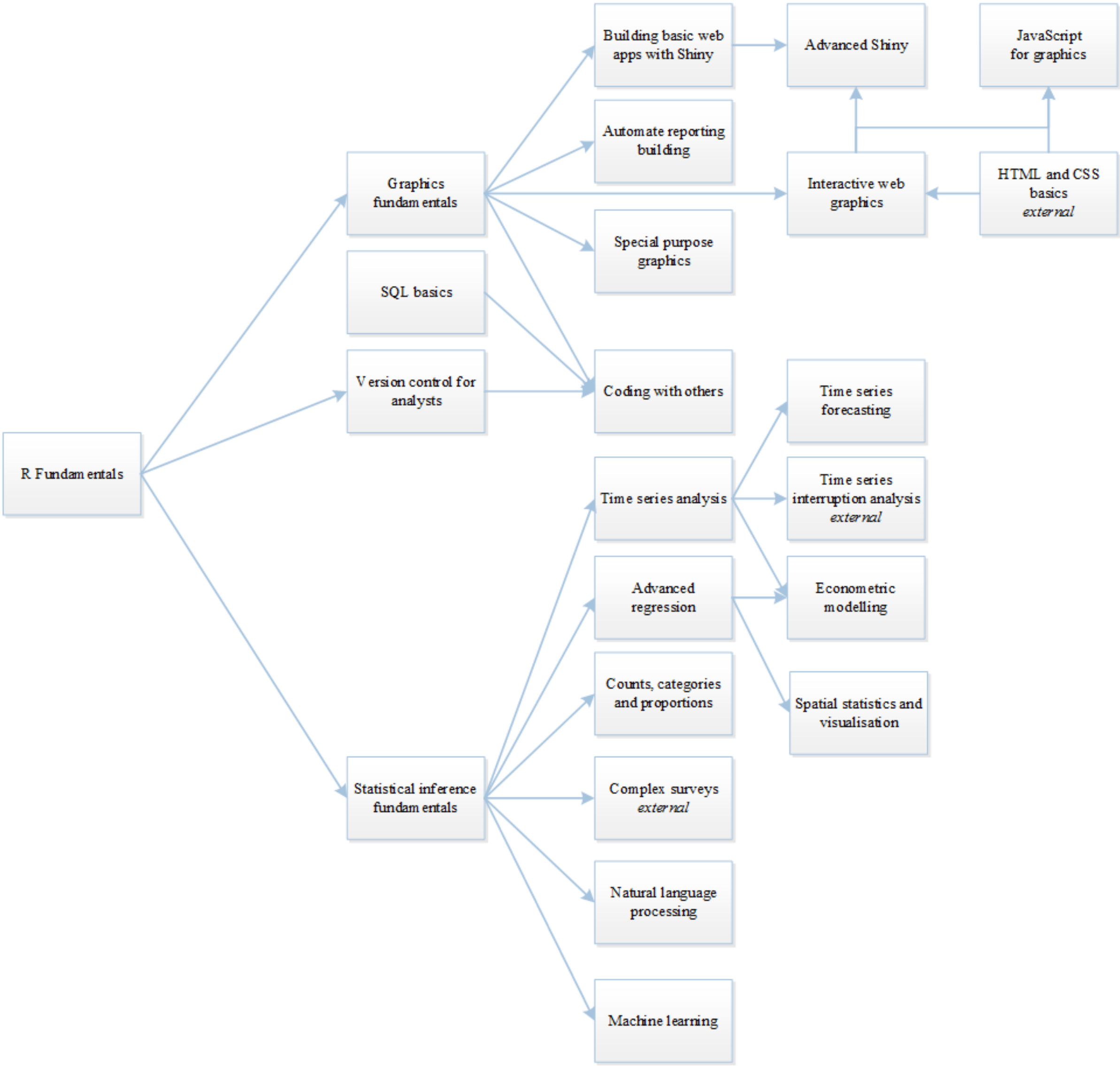
Well defined, relatively short modules

Relatively late in the journey it became apparent that we were paying a price in efficiency for the lack of clearly defined, repeatable training objectives and assets. This was an issue both for new recruits and for scaling up to assist other teams' development. The solution was to standardise and modularise the training, and create a clear development progression. Figure 1 shows the current version of this program (although not all courses indicated in the diagram are complete). Each box is a course of two to six sessions (usually one per week and 60 to 90 minutes long), a combination of seminar-style and in a computer training lab.

This tree makes it possible to identify the "next course" for building individuals' skills and gives a meaningful basis for conversations about prior learning. For example, the "statistical inference fundamentals" course is heavily focused on the bootstrap, cross-validation and robust estimation and real data and methods from our work. Even staff with a strong prior quantitative background (eg PhD level) generally attend this course, to see "how things are done here" and be sure of a common understanding of concepts and terms, ready for the more advanced applications downstream. The modularized approach is also important for sustainability. With the courses comprising a standard combination of presentation slides and lab scripts, it is easier to pass on material to new trainers to deliver and improve on.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

Figure 1 – progression of short courses and skillsets for MBIE analytics teams



THE RESULTS HAVE JUSTIFIED THE PROGRAM

The capability building program has demonstrated success by its achievements. The following is a (very partial) list of tangible outputs that would simply not have been possible without the revolution in workflow and tools and statistical capability of the team over the past five years:

- The various forms of Regional Tourism Estimates published from 2012 (MBIE 2016a) each combined new data sources and management with time series forecasting and complex survey analysis for world-first insight into regional tourism spend.
- The International Visitor Survey redevelopment project (MBIE 2013) faced numerous complex statistical and data management challenges during its design, pilot and back-casting. While specialist consultants handled the design and questionnaire, the project was only a success because of the rapidly building statistical skills in MBIE for analysis, testing and trouble-shooting.
- The Modelled Territorial Authority Gross Domestic Product estimates (MBIE 2015) combine management of many data sources with small area estimation techniques.
- The New Zealand Tourism Dashboard (MBIE 2016b) is one of the larger and more complex web applications developed with the `shiny` framework (Chang, W. 2016).

- The Regional Economic Activity Report (MBIE 2016c), built by consultants, is now maintained and managed by staff in a big boost for sustainability.
- The `mbiemaps` R package (MBIE 2016d) allows easy creation of choropleth maps with various New Zealand boundaries and other statistical overlays.

The capability building program continues to expand. In addition to the Sector Trends team, a further 60 MBIE analytical staff have learned from at least one of the courses. There is also growing interest from other Ministries, six external trainees having completed courses so far.

ACKNOWLEDGEMENTS

I wish to thank past and present staff and contractors of the Sector Trends and its predecessor teams; past and present senior managers Liz MacPherson, Adrienne Meikle and Michael Bird; and externally sourced trainers from the early days David Lillis and Ian Westbrooke.

REFERENCES

Chang, W., Cheng, J., Allaire, J., Xie, Y. and McPherson, J. (2016). *shiny: Web Application Framework for R. R package version 0.14.1*. https://CRAN.R-project.org/package=shiny

Ellis, P. (2016). Why you need version control. Blog post in *Peter's stats stuff*. http://ellisp.github.io/blog/2016/09/16/version-control

Manyika, J., Chui, M., Brown, B. Bughin, J., Dobbs, R., Rocburgh, C., Byers, A. (2011). *Big data: the next frontier for innovation, competition, and productivity*. McKinsey Global Institute. http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation

MBIE (2013). *International Visitor Survey revision 2013*. http://www.mbie.govt.nz/info-services/sectors-industries/tourism/tourism-research-data/ivs/about-the-ivs/ivs-revision-2013

MBIE (2015). *Modelled Territorial Authority Gross Domestic Product: experimental estimates to help research leverage New Zealand's official statistics – summary document.* http://www.mbie.govt.nz/info-services/sectors-industries/regions-cities/research/modelled-territorial-authority-gross-domestic-product/document-and-image-library/mtagdp-summary-document-optimised.pdf

MBIE (2016c). *Monthly Regional Tourism Estimates.* http://www.mbie.govt.nz/info-services/sectors-industries/tourism/tourism-research-data/monthly-regional-tourism-estimates/about-the-mrtes

MBIE (2016b). *New Zealand Tourism Dashboard.* http://tourismdashboard.mbie.govt.nz/

MBIE (2016c). *Regional Economic Activity Webtool*. http://webrear.mbie.govt.nz/

MBIE (2016d). *mbiemaps: boundaries in R for New Zealand Regional Council, Territorial Authorities, and others.* https://github.com/nz-mbie/mbiemaps-public

R Core Team (2013). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/

Spolsky, J. (2000). The Joel test: 12 steps to better code. Blog post in *Joel on software.* http://www.joelonsoftware.com/articles/fog0000000043.html

Westbrooke, I. and Ellis, P. (2014). Training to develop modern statistics in the workplace using R and R Commander – experiences from the New Zealand Government Sector. In K.Makar, B. de Sousa and R. Gould (Ed.), *Proceedings of the Ninth International Conference on Teaching of Statistics*, Vol. II, (Paper 5F3). Voorburg, The Netherlands: International Statistical Institute. http://icots.info/9/proceedings/pdfs/ICOTS9_5F3_WESTBROOKE.pdf

Wickham, H. (2016). *tidyr: easily tidy data with `spread()` and `gather()` functions. R package version 0.6.0.* https://CRAN.R-project.org/package=tidyr

Wickham, H. and Francois, R. (2016). *dplyr: a grammar of data manipulation. R package version 0.5.0.* https://CRAN.R-project.org/package=dplyr

DEVELOPMENT OF A TRAINING PROGRAM FOR NON-STATISTICIANS

WADE, Angie, ALDRIDGE, Vicki, KOUTOUMANOU, Eirini, LEE, Sophie, LANGAN, Dean
Centre for Applied Statistics Courses (CASC), Great Ormond Street Institute of Child Health, University College London (UCL), UK.
awade@ucl.ac.uk

In 2008, with a single Teaching Fellow appointment at our Institute of Child Health, we began developing training courses for non-statistical researchers and workers. Today, 6 staff members including an administrative assistant comprise the UCL Centre for Applied Statistics Courses (CASC) annually providing 2,349 non-examined person training days for Continuing Professional Development. We run 16 courses ranging from a 5-day introduction to Statistics and Research Methods, providing a thorough grounding in p-values and confidence intervals, through to software training (R and SPSS) and 1-day courses in topics such as Bayesian analyses and Dealing with Missing Data. Courses are commonly over-subscribed and include diverse audiences in terms of seniority (junior staff through to company directors), professional discipline, and application (charities, transport providers, advertising agencies and government officials are examples of the regular attendees in addition to our primary audience of paediatricians). This enterprise is unique within the UK and we know of none similar elsewhere. The template used has been more effective than initially anticipated, in part due to the general applicability of our subject area. Here we share the story of our journey to this point, what we have learnt and may have done differently, potential limitations and our future plans.

INTRODUCTION

The concept of lifelong learning has long been established, particularly in Denmark where there is a tradition of taking part in education through all phases in life (http://eng.uvm.dk/Education/General/Lifelong-learning). It has been defined as voluntary and self-motivated, enhancing competitiveness and employability (Commission of the EC, 2006). There are various reasons why individuals whose primary focus is not statistics may wish to learn statistical techniques beyond any formal training that they may already have received. There are many ways in which the desired training may be obtained. For example, study may take place independently from books or online material, from personal tutoring or as part of a group within a large lecture hall or smaller classroom.

Within our academic institute and attached paediatric hospital we identified many staff, mainly via our one-to-one consultancy sessions, with such a need and who would perhaps benefit from classroom focused teaching. In response, an introductory course was developed covering the usual diet of a basic introductory statistics course (research design, summarising data, confidence intervals and p-values, t-tests, anova) and material not always included in a course of this length (sample size estimation, reliability and validity). This course ran twice yearly and was consistently over-subscribed. A one-day ‘Introduction to Regression’ was added and this also proved to be popular.

We decided that expanding the statistics training program would be beneficial for local staff and that we could also move beyond the local audience and extend provision to external, fee-paying, participants. We reasoned that this move to external provision would have many advantages. These included (1) generating capital that could be used for further local growth (2) allowing a wider range of courses to be presented locally (ie. more specialised topics required by some local researchers could be covered and still delivered to full classrooms) (3) allowing courses to be run more frequently (4) bringing external students into UCL who may then engage in other taught courses and/or form other research collaborations. On the downside, there was the possibility of diluting the applicability of courses for our own staff by catering for a more diverse audience.

There has been an increasing body of literature concerning statistics training at tertiary level (Tishkovskaya & Lancaster, 2012), but this generally relates to the development of single statistics courses, usually to a homogenous group. We could find no literature specifically aimed at the development of a suite of pick-and-mix type courses for heterogeneous groups of participants, who may be many years post any formal training and whose primary interest would be to gain the skills required to work effectively, rather than qualification focussed. In this paper we will describe the process via

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

which we developed such an entity. By discussion of what we have learnt on route and where we are now, we provide information that may be of use for others embarking on similar initiatives.

THE FRAMEWORK

The process of creating CASC (Centre for Applied Statistics Courses) began in 2008 with the appointment of a single teaching fellow and the idea that we would develop short courses in statistics, for non-statisticians, ranging in duration from one to several days. It has long been recognised that students learn concepts such as statistical thinking and application more effectively if they are engaged in the subject matter and are motivated to learn (Garfield and Ben-Zvi, 2007). Although this may be less of an issue for our adult learners than for taught-course, examination- focussed students, we believed (from experience) that what had been observed by others that workshop based courses help to facilitate true understanding and building a good relationship with the students is important (Bradstreet, 1996), was likely also true for our intended audience.

Courses would be designed to develop statistical reasoning skills primarily and an understanding of the methodological processes involved utilising real data. This approach has been shown to be most appropriate for our intended audience (Bradstreet, 1996). The placement of training of statistical methods within the broader framework of the context of how and why data were collected has long been recognised and acknowledged (for example, Gal 2002). The courses we produced would be suitable for those with little or no prior knowledge of that specific area, although there may be a requirement for a basic statistical knowledge (p-values, confidence intervals and simple descriptive statistics) for some courses as appropriate.

Our participants would be drawn from a wide range of areas and experience. It was unlikely that all would be at a comparable level of understanding and so we felt it important to provide comprehensive notes with all courses. These notes, which were produced in book format with a proper introduction and index, contained material delivered in the course, practical exercises and, where applicable, code used in computer based examples. Powerpoints were also provided, but the existence of the comprehensive course notes meant that participants could fully concentrate on the lecture rather than needing to write their own notes on those powerpoints, as is generally the case.

One tenet that we decided to make was that there would be commonality between all courses. We did not want a disparate set of individual courses, with different presentation styles and no coherence between them, perhaps delivered by lecturers who were unaware of the content of other CASC courses. Each CASC member would be involved at some point in delivering each course (either being the lead person or helping with practicals), ensuring that all courses ‘fit’ together.

Participant Interaction

All short courses would have practicals to be done by the students in the classroom with support from at least 2 lecturers. These practicals would be of varying lengths and interspersed during the lecture time as this was our preferred interactive teaching style, aiming to catch anyone falling behind at the earliest opportunity. Having a large amount of interaction also means that the lectures may diverge along paths more suitable to a particular group than having a standard set pattern regardless. Where sessions are repeated many times to many groups it may be difficult for the lecturer to remain engaged and not become staid, varying the practicals and discussions means that this is less likely to happen. We also developed data generation exercises to help with understanding and facilitate group participation (Burwalls UK Annual Statistics Teachers meeting 2010, 2013). Where appropriate, technology such as Personal Response Systems (PRS) would be used to both encourage student participation but also to give realtime feedback to the lecturer. These PRS allow the lecturer to intersperse their powerpoint with multiple choice questions that the audience can vote on, the results of the vote showing to what extent the class has understood the principles taught. What we aimed to do was to replace the ‘flat’ lecture format often used within training of this type, perhaps followed by a practical, with sessions that are broken into segments with interactivity from the audience using a variety of media.

Class size

On deciding on the number of course participants to enrol, several factors were considered. A minimum class size of 10 was determined by our desire to use resources effectively, smaller than this would probably be inefficient in terms of income generated and individuals taught in relation to CASC staff time and use of rooms. For class based courses, we required that individuals had a desk for practical work and for the software based courses they would also require a computer. Given our local situation, this inferred classes of more than 20 individuals for software based and 35 for classroom based courses were infeasible. To properly facilitate classes of the proposed size (10-35) we decided that each course should have a minimum of 2 CASC teaching fellows in attendance. For each course, one or more teaching fellows would take the lead, delivering the lecture, and all would help participants during practical sessions.

AFTER THE COURSE

It is important with any training to ensure that this is effective in catering to the needs of the participants. This is especially difficult to determine with the framework we had proposed whereby individuals would perhaps attend for a single standalone course. Even when statistics education is aimed at students with an ongoing relationship to the host institution (for example, 3 year degree students), it is difficult to determine the impact of a given teaching method on improved student outcomes, even when exams are undertaken (Garfield and Ben-Zvi, 2007).

Several processes were put into place with all courses to determine whether the aims of effective education were being attained. As stated above, interspersing practicals throughout the day and having at least 2 teaching fellows on hand to assist would enable us to identify quickly where the teaching was not being understood, as would the use of PRS.

Feedback at the end of the course initially took the form of paper questionnaires and later electronic surveys emailed to the participants. There has been much work on the comparison of these two types of feedback from university based students. We are currently investigating the differences (via a randomised trial) of the extent to which feedback varies according to media with students attending CASC courses, where there is perhaps less incentive to give any feedback as there is less chance of ongoing interaction and less of a feeling of ‘belonging’ to the university.

One question we ask on the end of course e-survey is how they heard of the course. Two of these options are ‘word of mouth’ and ‘attended previous course’. Either of these options suggests satisfaction with CASC courses and is a form of positive feedback. However, there are problems with interpreting this data since stating other reasons is not necessarily negative. In recent years, we have initiated a loyalty card system whereby obtaining 6 course stamps yields a free course. The fact that so many individuals have completed these is also a positive endorsement.

BUSINESS CASES

To initiate and extend the CASC proposed framework of courses required employment of teaching fellows and for this we needed to convince the University of the economic viability of such a move. The financial case for a recent teaching fellow post is outlined in this section.

Using the academic and research non-clinical scales including on-costs the annual cost to the University of employing a teaching fellow was calculated to be between £33,758 and £39,755. We do not often run courses with less than 15 attendees, usually there are between 20 and 30. Each full day course has been charged at £125. Catering and printing costs per day are usually below £15 and never above £25. Figure 1 shows the net benefits if we assume a net income per participant per day of £100 and the upper salary range of £39,755, which are conservative estimates.

Note that this graph is also based on the maximum appointment rate (grade 7), for appointment at grade 6 then the lines (net benefit) will rise by £5,997. The graph shows that it is extremely unlikely that there would not be a net benefit attributed to employment of a further teaching fellow to the statistics courses. With an additional 2 days teaching for 44 weeks of the year, which is a likely scenario, a healthy profit is certain to be returned. A loss only occurs when there are less than 39 extra day courses per year and these attract on average 10 participants or less. Such a scenario is extremely unlikely to occur and hence the chance that net benefit will be less than the costs involved can reasonably be discounted.

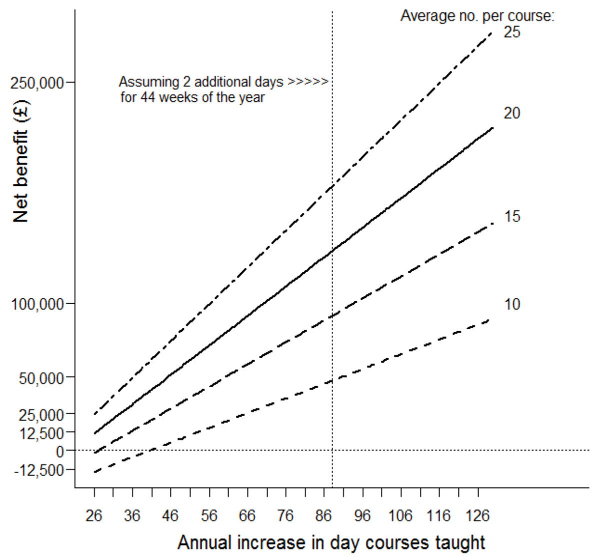


Figure 1: Likely net benefit from employment of an additional teaching fellow

GROWTH

In 2008 the first teaching fellow was employed, one more followed in 2012, one in 2015 (70% time, 30% research) and a further appointed in 2016. An administrator to take on the burden of course bookings etc. was employed in 2014. The flagship 5-day introduction Statistics and Research Methods which was originally developed in 1991, and Introduction to Regression in 2007. These 2 courses formed the basis on which CASC developed. Subsequently a further 18 courses were added as shown in Table 1. In the last academic year courses ran between 2 (2x2 tables) and 8 (Intro to SPSS) times.

Table 1: Timescale of initiation of courses run by CASC. All courses are 1 day 10am-4.30 unless otherwise stated.

Year of introduction:	Course title(s):
1991	(i)Statistics and Research Methods (5 days 10-4.30pm)
2007	(ii) Introduction to regression analysis (2 days 10am-3pm)
2008	(iii) Analysing time to event data (1 day 10am-3pm) (iv) Logistic regression analysis (v) Sample size estimation (incorporating CD for calculations) (vi) Introduction to Epidata
2009	(vii) Analysing a 2x2 table (viii) Introduction to missing data (1 day 10am-4.30pm, optional software 2nd day 10-1)
2010	(ix) Critical appraisal (x) Introduction to Bayesian Analysis (xi) Introduction to R (xii) Dealing with Statistics with R (3 days)
2011	(xiii) Introduction to Quantile Regression
2012	(xiv) ANOVA with SPSS (xv) Introduction to Data Linkage (xvi) Multilevel models using R
2013	(xvii) Introduction to SPSS (xviii) Regressions with R
2014	(xix) Reliability and Validity
2016	(xx) Further topics in R

Numbers of student training days provided from 2009 until the end of the last full academic year (Sept 2014–Aug 2015) are shown in Figure 2.

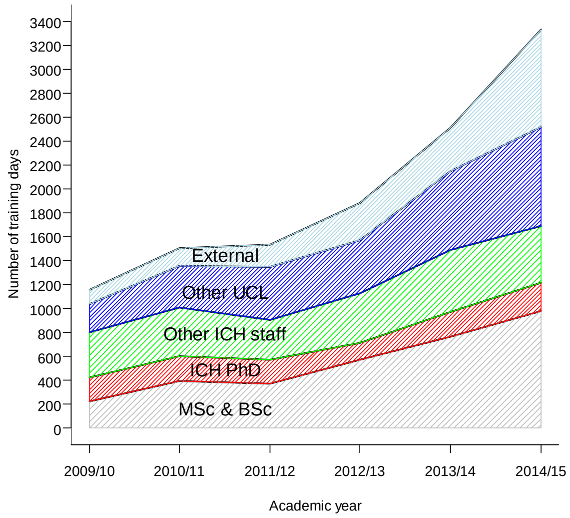


Figure 2: Training days provided by CASC for academic years 2009/10 until 2014/15 according to location of course participant. (ICH: Institute of Child Health; UCL: University College London)

The MSc and BSc student places have grown over the time period to accommodate modules for 9 separate taught courses. The numbers of days training provided for PhD students and other local staff (ICH – our host institution for whom we originally ran training) has remained relatively static. The number of days taken by other UCL (our wider host institution for whom we do not have a primary obligation and who buy into our services) has increased about 4-fold. The greatest percentage increase has been in external training days provided, which has been around 8–fold and continues to grow.

WHAT HAVE WE LEARNT

The basic framework selected for the courses has been successful. Feedback from participants suggests that hard copy notes are invaluable. In recent years, we have also made echocast recordings of the lectures for the participants to watch at their leisure if they wish to. Incorporating lots of interaction during the courses has really helped us to tailor for and better accommodate participants. The more informal moments such as talking to individuals during coffee breaks or lunch often has provided insightful information. We have found that incorporating lots of practicals per se is not necessarily enough to identify those who are flailing but there is a need to be pro-active. No matter how many times you tell people to ask if they are struggling, some just will not make that move. It is necessary to actually target individuals and approach them to see if they need help or have misunderstood anything.

Commonality of teaching fellows across courses has worked well. It means that a back-up is always available, that we can easily advise participants about the other course contents in detail and has generated a centre that functions as an entity with many course attendees coming to the whole range of CASC courses, having found a teaching mode that suits their needs.

Whilst we have found it necessary to have at least two teaching fellows per course to ensure all course participants are following, in retrospect we could have moved to that position earlier. Initially, we all attended all courses, particularly during the period when there were 3 of us, and it was in some respects difficult to ‘let go’ of a course you may have helped create and run in the early stages. With the software courses in particular it is useful to have more than 2 teaching fellows on hand at the start of the day in case of issues with computers or to identify individuals who perhaps don’t have as many of the pre-requisite skills for that course. It is important that even if all teaching fellows are not likely required for the entire day, there are enough that have the time blocked in their diary, so that they are available if needed.

Timing has been one of the biggest problems. It is very difficult to determine how long a course, incorporating practicals and encouraging interaction in the form of questions if things are unclear, will take to deliver. This is particularly an issue for one day one-off attendees, as there is no chance to recoup time 'later in the term' for instance. Whilst as teachers we tend to prefer to produce more rather than less in our written notes, with the idea that even if we don't cover everything it can serve as a reference for later, this is not the best approach as far as participants are concerned. Even if the lecturer states at the start that we may not cover all in the timescale, the course participants tend to feel 'cheated' or that they have 'missed out' when there are items left uncovered. It is preferable to have additional material prepared in case there is time at the end, but not to show this to the participants unless that time is available. For the lecturer, this gives the security that they will not run out of material, without giving participants a feeling that they are pressured to complete all. Alternatively, any time at the end can be used for additional questions. We sometimes worry that if people have paid for a day course they may be disappointed to not get all the hours teaching they signed up for, but experience shows us that most prefer to finish early rather than feel they haven't completed all of the material. The first run of any course is now marketed at half the intended price.

One positive aspect of our courses was that they were short and not restricted to an academic timetable. We were therefore able to schedule our courses to suit room availability. However, it was pointed out to us that this meant many courses ran during school holidays, which were not so easy for some to attend. As a consequence, we now ensure that we do not consistently run the same courses within the school holiday, so that any particular course is more likely accessible to more people.

Marketing of courses has not been a problem and most have been full or over-subscribed. Describing the courses to the market has been something of a challenge though. Many of our courses are labelled as 'Introduction to …' and we do try to emphasise this in an 'intended for..' section, but still have some participants who expect a higher entry level and feel the pace is too slow. To address these issues we continue to work on advertising material and are considering pre-course 'tests' to clarify the level of understanding expected or required.

Overall the venture has worked better than originally anticipated. Over 8 years we have increased from 1 to 6 staff members and now provide up to 20 different short courses. Income has been generated via inclusion of external fee-paying participants and this does not seem to be to the detriment of internal staff, who benefit from a wider range of courses that run more frequently.

REFERENCES

Bradstreet, T. E. (1996). Teaching introductory statistics courses so that nonstatisticians experience statistical reasoning. *The American Statistician*, 50(1), 69-78.

Gal, I. (2002). Adults' statistical literacy: meanings, components, responsibilities. *International Statistical Review*, 70 (1), 1-25.

Garfield, J. & Ben-Zvi, D. (2007). How students learn statistics revisited: a current review of research on teaching and learning statistics. *International Statistical Review*, 75 (3), 372- 396.

Tishkovskaya, S. & Lancaster, G.A. (2012) Statistical education in the 21st century: a review of challenges, teaching innovations and strategies for reform. *J of Stat Education*, 20 (2), 1-56.

Commission of the European Communities, Brussels (2006). Adult learning: It is never too late to learn.

WERE WE AT THE SAME CONSULTATION?

GORDON, Ian and FINCH, Sue
Statistical Consulting Centre
The University of Melbourne
irg@unimelb.edu.au

Statistical consulting is a context in which much statistical education can and should take place. However, this is rarely formalised or evaluated, and is generally ad hoc in its structure. Apart from anecdotal comments, the effectiveness of the statistical education involved is usually not well measured.

We carried out surveys of clients and consultants based on actual consultations between experienced applied statisticians and clients they were assisting. Clients and consultants responded independently. We adopted the five stages of the "problem, plan, data, analysis, conclusions" (PPDAC) framework in designing the surveys, and asked clients and consultants common high level questions in terms of these stages. The consultant also described details of the consultation within each of these stages.

The results of these surveys allow us to interrogate the consistency between the consultant's and the client's expectations and perceptions of the content of the consultation, in relation to the PPDAC stages. We also describe the statistical education needs of clients based on the detailed information provided by the consultants.

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

TEACHING VISUALISATION IN THE AGE OF BIG DATA: LESSONS FROM THE PAST

CHIERA, Belinda and KOROLKIEWICZ, Malgorzata
University of South Australia
Belinda.chiera@unisa.edu.au

Recent technological advances have led to increasingly more data becoming available than ever before, a phenomenon known as Big Data. The volume of Big Data runs into zetabytes, offering the promise of valuable insights. Visualisation is key to unlocking these insights, however the size and variety of Big Data pose significant challenges. Traditional approaches to visualising data, typically used in statistics courses, often fail to produce meaningful results. The fundamental principles behind tried-and-tested methods for visualising data, irrespective of its size, are still as relevant as ever. However, the emphasis necessarily shifts to why we are attempting visualisation.

In this paper, we discuss graph semiotics to build data visualisations for exploration and decision-making, with emphasis on formulating elementary-, intermediate- and overall-level analytical questions. We illustrate our ideas using the public scanner database Dominick's Finer Foods, containing approximately 98 million observations. We employ commonly used analytic tools for Big Data (SAS, R and Python) to produce visualisations and present exemplars of student work based on the suggested approach to visualisation. Our target student audience are working professionals undertaking postgraduate studies in statistics, or more broadly 'data analytics'.

KEYWORDS: Visualisation, Graph Semiotics, Big Data, Dominick's Finer Foods (DFF)

INTRODUCTION

Recent technological advances have led to data collection at a rate never seen before, from sources such as climate sensors, transaction records, social media and videos, to name a few. With the advent of *Big Data* come insights at a previously unseen depth and breadth of detail. Operational decisions are increasingly based on data rather than experience or intuition (McAfee & Brynjolfsson, 2012) and more broadly, a shift in perspective is under way on the relationship between data and knowledge generation (Ekbia et al, 2015).

Big Data is typically defined in terms of its *Variety*, *Velocity* and *Volume*. Variety refers to expanding the concept of data to include unstructured sources such as text, audio, video or click streams. Velocity is the speed at which data arrives and how frequently it changes. Volume is the size of the data, which for Big Data runs to the order of petabytes through to zetabytes.

Visualisation is a potentially valuable way to make sense of Big Data, to uncover features, trends or patterns to produce actionable analysis and provide deeper insight (SAS, 2014). There is an increased focus on visualisation over formal data analysis, partly due to the proliferation of easy-to-use web-based visualisation tools (e.g. Tableau Online), and partly due to an added emphasis on the power of well-designed visualisations and the demand for a new type of Big Data 'dataviz' analyst (McCosker & Wilken, 2014). However, the opaque character of large data sets makes it difficult to describe in a systematic way how to effectively translate Big Data into visual or other kinds of knowledge. Furthermore, a 'black box' approach to generating data visualisations puts data analysts at risk of producing ornate and visually pleasing graphics that are otherwise useless.

In this paper, we propose a portable framework to arm aspiring visual analysts with tools to decide what is/is not useful when visualising Big Data. The target audience are working professionals seeking to expand or formalise their data analytics skills through postgraduate university study. Based on our own teaching experience in this context, we advocate introducing a framework for visualising Big Data through real-life case studies, instead of presenting students with an encyclopaedia of methods and graphical displays.

Despite the new challenges introduced by the emergence of Big Data, visualisation techniques themselves need not be new – *lessons from the past* in the form of tried-and-tested graphs can still be effective – however with a new way to perceive the data. In this paper we adopt the seminal work of (Bertin 1967) and extensions in Tufte (1983) in the field of *graph semiotics*, to explore and harness the scope of Big Data to formulate meaningful questions and drive analysis. To illustrate graph semiotics, we use the Dominick's Finer Foods (DFF) scanner database, publicly

In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

available at https://research.chicagobooth.edu/kilts/marketing-databases/dominicks courtesy of James M. Kilts Center, University of Chicago Booth School of Business. We also demonstrate the use of graph semiotics to formulate three levels of questions from the supermarket database: elementary-, intermediate- and overall-level, before presenting our conclusions.

CASE STUDY DATA PREPARATION: DOMINICK'S FINER FOODS

The DFF supermarket database contains data recorded weekly from September 1989 - May 1997, yielding approximately 98 million observations that are non-homogeneous in time with some missing records. The database is split into 60 files which are either: *general files*, capturing sales and store-level demographics; and *category-specific*, for the different grocery items stocked, including weekly sales data such as *store*, *item price*, *units sold*, *total dollar sales* and so forth.

Prior to visualisation the data was checked and merged in a meaningful way. Given the database size, viewing the raw data to check validity was not possible thus pre-processing Big Data takes much forethought coupled with computer language skills. We used the programming language *Python* and its in-built *Pandas* (PANel DAta Structures) library for the data cleaning and merging process. We chose Python as it provides data structures that are amenable to large volumes of data and can be used for the fast removal of missing values, fast merging of database entries and slicing the data into smaller subsets, ready for visualisation and question formulation.

BIG DATA: GRAPH SEMIOTICS FOR VISUALISATION

Data visualisation embodies at least two distinct purposes: (1) to communicate information meaningfully; and (2) to 'solve a problem' (Bertin, 1967). It is suggested the latter purpose can be achieved by answering and/or postulating questions from data visualisation, as was the approach in the originating work (Bertin, 1967). It is thus in the same spirit we adopt graph semiotics and reference the fundamental data display principles, in the visualisation that follows.

At the crux of graph semiotics are *retinal variables*, a selection of which are given in Table 1, with application to data type as indicated (Bertin, 1967). The perceived accuracy of each retinal variable is also provided, which can be used to make informed choices about data visualisation, rather than blindly rotating through a selection of visualisations.

Table 1 *A selection of retinal variables applied to data type for visualisation.*

Perceived Accuracy	*Quantitative Variables*	*Qualitative Variables*	
	Discrete, Continuous	*Ordinal*	*Nominal*
Most Accurate	Position	Position	Position
	Orientation	Density	Colour Hue
	Size	Colour Saturation	Density
	Density	Colour Hue	Colour Saturation
	Colour Saturation	Size	Shape
Least Accurate	Colour Hue	Orientation	Orientation

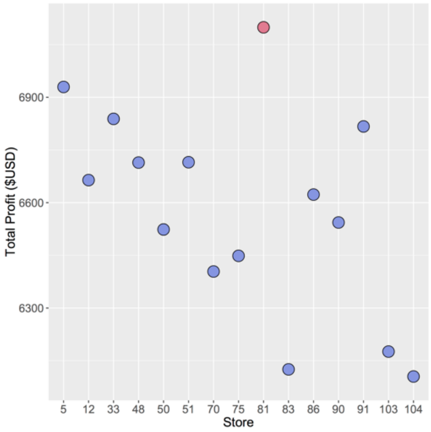
The retinal variables are: *Position* – the position of graphing symbol relative to the axes; *Orientation* – the direction of the plotted data; *Size* – the space occupied by the graphing symbol; *Density* – the distance between plotted observations; *Colour Hue* and *Colour Saturation* – shades and transparency levels of colour to highlight differences; and *Shape* – the graphic symbol used to represent the data. We will now investigate these variables further.

BIG DATA: VISUALISATION FOR INSIGHT AND EXPLORATION

A motivation behind visualisation is the construction and/or answering of questions about the data. It was suggested in Bertin (1967) that a question about temporal data can be defined by its *type* and *level*, where there are at least as many types of questions as physical dimensions used to construct the graphic. This work was extended to spatio-temporal data in Koussoulakou & Kraak (1995), where the distinction between question types can be independently applied to both the temporal and spatial data dimensions. The *level* of a question encapsulates the concept of *elementary*-, *intermediate*- or *overall-level* insights, which we now demonstrate with the aid of the R and SAS programming languages and their respective visualisation libraries.

Elementary-level questions typically involve a single observation, providing focused insights, e.g. *the store with the maximum volume of cheese sales in Week 103* (Figure 1(a)) which in this case is store 81, based on the use of the retinal variables *Position* (profit levels) and *Colour* (maximum profit). A rainfall plot (Figure 1(b)) uses the retinal variable *Position*, while *Colour* is used solely for aesthetic purposes. Adding the qualitative variable *Price Tier*, indicating the level of a store, allows for elementary-level questions such as *Which Price Tier has the largest variability in cheese sales?* or *Which Price Tier sells the most cheese?*

Figure 1 *Dot plot of cheese profit by store* *(a) Rainfall plot of cheese by Price Tier*



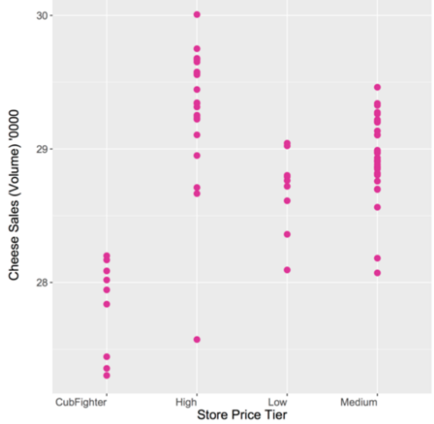


Figure 2 depicts a 100% stacked bar chart of the cheese profit in each store over time, displaying two types of categorical information: *year* and *Price Tier*. The retinal variables *Position* and *Colour* are used with the inherent temporal and spatial dimensions of the data (weekly profit over geographically disparate stores) to represent a quantitative variable (cheese profit) and two types of qualitative variables (ordinal and nominal). Thus questions can be formulated over either or both of these dimensions, e.g. *Which Price Tier had the highest cheese profit? In 1992, in which Price Tier do stores attain the highest profit for cheese? In which year did stores in the "High" Price Tier see the lowest profits for cheese?*

Intermediate-level questions involve several items, e.g. *product sales over a week* (temporal) or *the five stores with the highest weekly sales* (spatial). Figures 3 (a) and (b) depict bubble plots which use the retinal variables *Position, Colour, Shape, Density* and *Size* to represent profit and price (quantitative) by Price Tier (qualitative), where bubble size is directly proportional to product price. Questions can thus focus on quantitative variables, e.g. *Is there a relationship between cheese price and profit?* or on qualitative variables, e.g. *Is there a relationship between front-end candies prices and profits across price tiers?*

Not all plots are useful. Comparisons of Cheese and Cracker sales across store Price Tiers using a boxplot (Figure 4(a)) are cumbersome, with the data dwarfed by the variability of sales in the medium Price Tier. A butterfly plot of the same data (Figure 4(b)), in which the retinal variable *Orientation* is used, provides a more informative comparison.

Overall-level questions focus on general trends (Bertin, 1967). Some time-based plots can render very little information (Figure 5), however, through the addition of the retinal variable *Colour Hue* the same information can be captured for informative insight (Figure 6). Questions such as *What is the general trend of cheese sales over time?* can be asked and answered. Adding a qualitative variable extends the visualisation to a treemap (Figure 7(a)). The quantitative variables Price and Movement of front-end candies set the size and colour of the rectangular tiles; the qualitative variable Price Tier is used to create groupings in the plot. Thus while the price of front-end candies appears to be relatively uniform across all stores in each price tier, the colour variation indicates variability in buying patterns, e.g. stores in the Low and Cub Fighter price tiers.

Figure 2
Stacked 100% bar chart of cheese profit by Price Tier between 1991-1997.

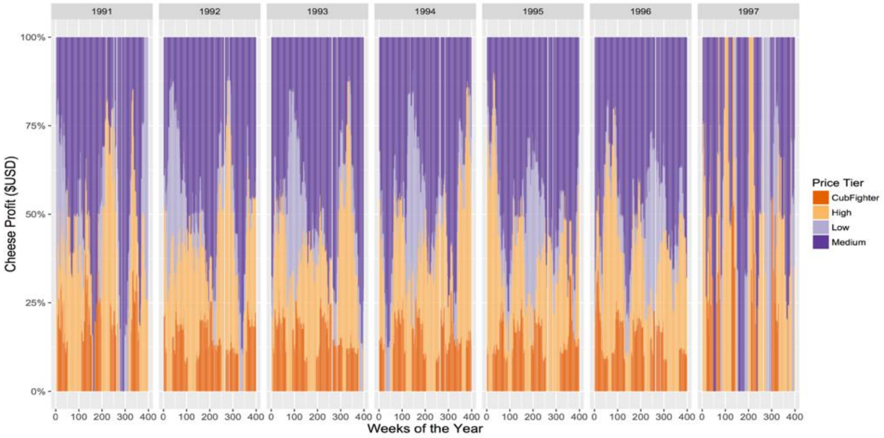
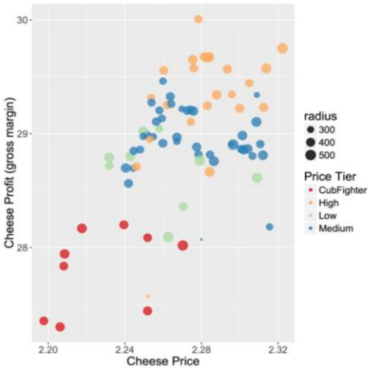


Figure 3
(a) *Bubble plot of Cheese*



(b) *Bubble plot of Front-End Candies*

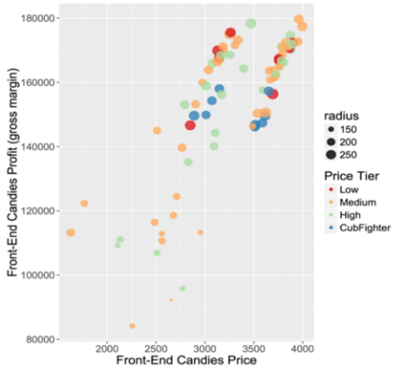
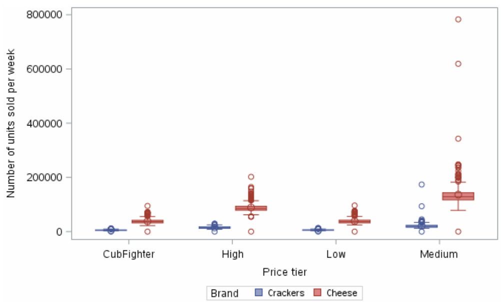


Figure 4
(a) *Box plot of Cheese and Cracker Sales by Price Tier*



(b) *Butterfly plot of Cheese and Cracker Sales by Price Tier (mean, in thousands)*

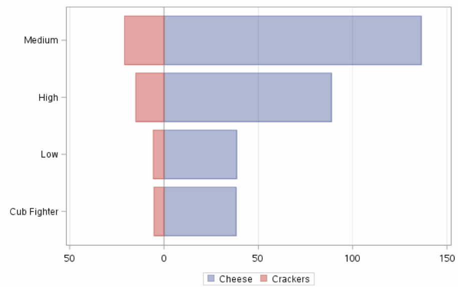
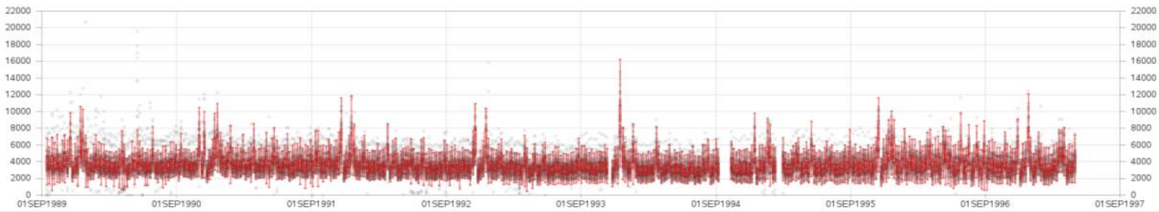


Figure 5 *A plot of weekly cheese sales (Number of units sold by store)*



In H. MacGillivray, M. Martin and B. Phillips (Eds.). Proceedings of the Ninth Australian Conference on Teaching Statistics, December 2016, Canberra, Australia.

Figure 6 *Calendar Heat Map of maximum weekly cheese profit.*

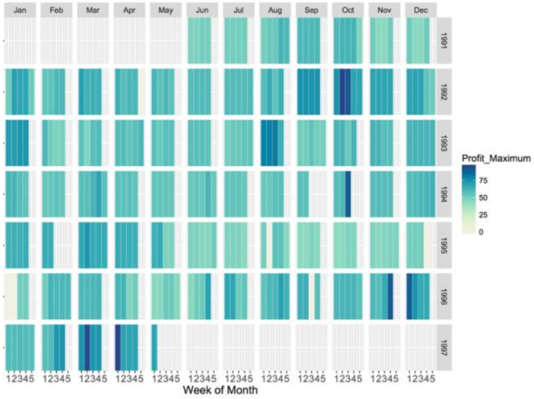
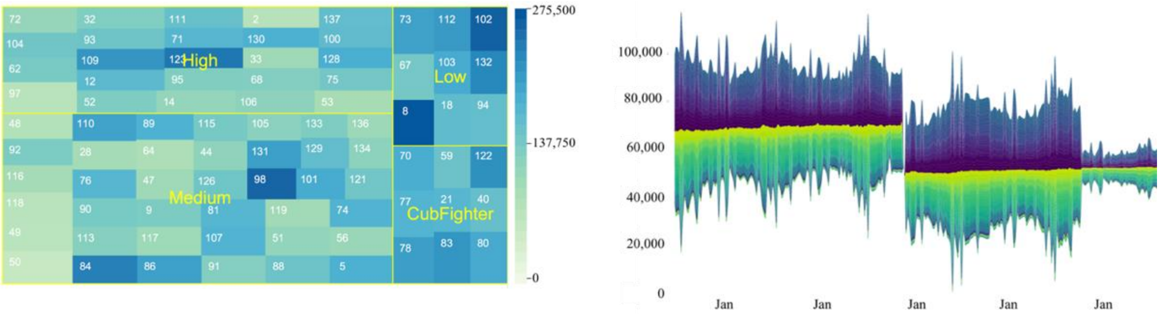


Figure 7(b) is another variation of a time series plot – the streamgraph. The retinal variables *Length, Orientation* and *Colour Hue* combine quantitative information (beer sales) with qualitative groups (stores) to allow for questions such as *What is the overall trend of beer prices? Do the prices change over time?* Then, coupled with Colour Saturation to focus on a store of interest, *What is the overall trend of beer prices at a particular store?*

Figure 7

(a) Tree Map of front-end candies by Price, Movement and Price Tier *(b) Streamgraph of Beer Sales by Store*

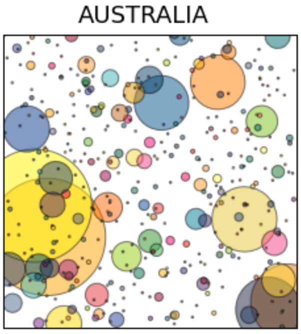


STUDENT EXEMPLARS - LESSONS FROM TEACHING PRACTICE

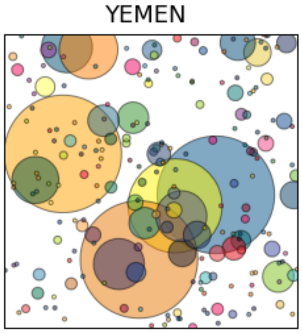
In our teaching practice at postgraduate level we frequently encounter working professionals with no or limited prior background in programming and/or Statistics. Through coursework in which they are taught Python, R, SAS and Statistics, our students are encouraged to apply the principles presented here to visualise real-world Big Data they collected and pre-processed themselves (Figures 9(a)-(d)). In Figures 9(a) and (b) a student used the retinal variables *Size* and *Colour Hue* to visualise a country's overall mood based on data mined from blog posts expressing a multitude of feelings, with brighter colours reflecting brighter moods and vice versa. The contrast between the frequency and mood of Australian blog posts versus those in Yemen were demonstrated. In Figure 8(c), another student used *Size*, *Position and Colour* to visualise billionaire net worth and age (2014 figures) across the world. Figure 8 (d) shows the distribution of numbers 1 to 9 in the aggregation of 204,432 four-digit iPhone passcodes using *Size* and *Shape*. These exemplars illustrate the outcome of using newly developed computing skills combined with graph semiotics in support of deriving insights from Big Data without sacrificing too much of clarity and accuracy.

Figure 8

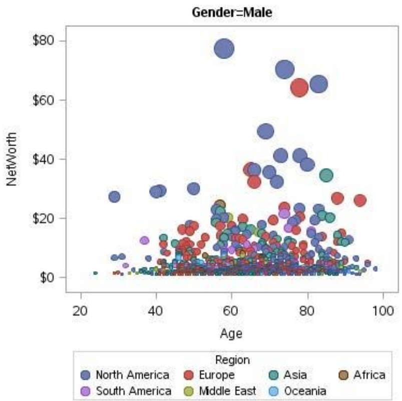
(a) Python Plot: Australia's Mood by Blog



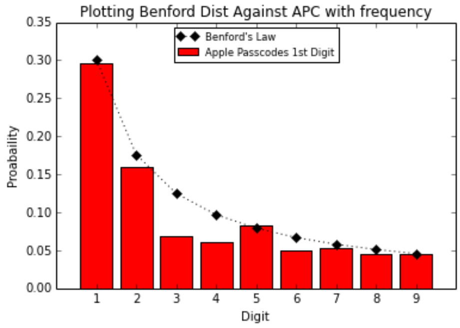
(b) Python Plot: Yemen's Mood by Blog



(c) SAS Plot: Male billionaires by region



(d) R Plot: first digit of iPhone passcodes



CONCLUSION

Recent advances in the new age of Big Data have challenged the concept of data visualisation. In this paper we suggested that tried-and-tested techniques – *lessons from the past* – can be adopted for Big Data visualisation to drive the formulation of meaningful questions. We used graph semiotics to depict multiple characteristics of data through retinal variables, which drove the formulation of elementary-, intermediate- and overall-level questions. These techniques were applied using a case study based on Dominick's Finer Foods. We demonstrated insights from Big Data with commonly-used visualisations and presented exemplars of student work. By applying modern graphics to Big Data, with foundations still traceable to retinal variables (Bertin, 1967), lessons from the past were shown to support data visualisation in the future.

REFERENCES

Bertin, J. (1967). *Semiology of Graphics: Diagrams, Networks, Maps*. Madison, Wisconsin: The University of Wisconsin Press, 712 pages.

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V. R., Tsou, A., Weingart, S. and Sugimoto, C. R. (2015), Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*, **66**, 1523–1545.

Koussoulakou, A., Kraak, M.J. (1995). Spatio-temporal maps and cartographic communication. *The Cartographic Journal*, **29**, 101-108.

McAfee, A., Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*, **90**(10), 60-68.

McCosker, A., Wilken, R. (2014). Rethinking 'big data' as visual knowledge: the sublime and the diagrammatic in data visualisation. *Visual Studies*, **29**(2), 155-164.

SAS (2014). *Data Visualization Techniques: From Basics to Big Data with SAS Visual Analytics.* SAS: White Paper.

Tufte, E.R. (1983). *The visual display of quantitative information*. Graphics Press, Cheshire, Connecticut, 197 pages.