OZCOTS 2008

Proceedings of the 6th Australian Conference on Teaching Statistics

Editors:

Helen L. MacGillivray
Queensland University of Technology

Michael A. Martin
Australian National University

Statistical Education Section **Victorian Branch**

Statistical Society of Australia Inc.

ISBN: 978-0-9805950-0-0

PREFACE

OZCOTS 2008

6th Australian Conference on Teaching Statistics

The Australian Learning and Teaching Council, ALTC, http://www.altc.edu.au (formerly The Carrick Institute of Learning and Teaching in Higher Education), was established late 2004 and is an initiative of the Australian Government Department of Education, Science and Training. Its mission is to promote and advance learning and teaching in Australian higher education.

OZCOTS 2008 combined the interests of the Statistics Education Section of the Statistical Society of Australia Inc with activities of Professor Helen MacGillivray's ALTC Senior Fellowship program entitled "The teaching and assessment of statistical thinking within and across disciplines". The Senior Fellowship Scheme supports leading educators to undertake strategic, high profile fellowship activities in areas that support the ALTC mission. Senior Fellows undertake a full time program of highly strategic fellowship activities over one year. Helen was one of the inaugural Senior Fellows.

The six invited speakers at OZCOTS 2008 were members of the Fellowship's international collaborative team, and their participation in OZCOTS 2008 was fully funded by the Fellowship.

OZCOTS 2008 acknowledges the support of the Australian Statistical conference 2008 and the Victorian Branch of the Statistical Society of Australia Inc (SSAI)

OZCOTS 2008 Conference Committee

Helen MacGillivray (joint chair, joint editor), Queensland University of Technology

Brian Phillips (joint chair, local arrangements), Swinburne University

Michael Martin (joint editor), Australian National University

Kay Lipson, Swinburne University

Peter Howley, University of Newcastle

Brief Biographies of Invited Speakers

Professor Adrian Bowman, Department of Statistics, University of Glasgow, Scotland. http://www.stats.gla.ac.uk/~adrian/
Adrian is an internationally respected statistician and researcher who has also been a leading innovator in statistics education for twenty years. He directs one of the principal nodes of the Higher Education Academy MSOR (Maths, Stats and OR) network. Many of the activities of the Glasgow node of this network provide information on, and resources for, the use of technology in the teaching and learning of statistics, but it also provides leadership in ideas and strategies in teaching statistics at the tertiary level. Previously Adrian was the leader of the STEPS project which was created by a consortium of eight UK university departments who developed problem-based learning materials in Statistics, delivered on the computer, which have been used around the world for the past decade. Adrian's most recent interests in learning and teaching have been in developing the rpanel package. This provides a set of simple interactive controls for R functions which are particularly useful in creating dynamic graphics. The package is intended for general use but there is a particularly strong application in a teaching context. The web page for the project gives a variety of examples. Adrian's research interests are nonparametric smoothing techniques, generalised additive models, three-dimensional shape modelling, graphics, statistical computing, and computer based learning.

Dr Robert Gould, Director of the Centre for Teaching Statistics, University of California, Los Angeles. http://www.stat.ucla.edu/~rgould/
Rob is also vice-chair of undergraduate studies for the Department of Statistics at UCLA. His primary interests are in statistics education. He was co-principal investigator (with Roxy Peck), developer and instructor for the INSPIRE project - an on-line content-based course for beginning secondary school statistics teachers. He is a past chair of the American Statistical Association's Advisory Committee on Teacher Enhancement, and on the board of the ASA's Statistics Education Session. He is founding editor of Technology Instruction in Statistics Education, a new on-line journal that focuses on teaching with and about technology in statistics. He has worked as a consultant on the Tinkerplots (statistical graphics software for middle-school students) project with Cliff Konold and is currently the evaluator for Konold's Model Chance project, which is designing simulation software to teach probability to middle-school students. He is working with PI Joan Garfield as an evaluator for the NSF-funded AIMS project (Adapting and Implementing Innovative Material in Statistics).

Professor Michael Martin, School of Finance and Applied Statistics, Australian National University. http://ecocomm.anu.edu.au/martin/
During his time as the Annenberg Distinguished Assistant Professor in Statistics at Stanford University, Michael was awarded the Dean's Award for Distinguished Teaching in the School of Humanities and Sciences. In 1994 he returned to ANU where he convenes the degree program in actuarial studies. He received the ANU Vice-Chancellor's Award for Excellence in Teaching in 2000, and an Australian Award for University Teaching in 2007. He is chair of the Statistical Education Section of the Statistical Society of Australia, and has been recently named as a Fellow of the American Statistical Association for his services to research and teaching in statistics. His research interests include resampling, statistical graphics, statistical education, biostatistics and environmental statistics.

Dr Peter Petocz, Associate Professor, Department of Statistics, Macquarie University. http://www.efs.mq.edu.au/staff/alphabetical_list_of_staff/peter_petocz

Peter's research interests include the biostatistical areas of nutrition and orthodontics, and a long-standing interest in mathematics and statistics pedagogy, both in practical terms and as a research field. On the pedagogical side, Peter is the author of a range of learning materials - textbooks, video packages and computer-based materials - and has been recognised for his teaching by a University of Technology, Sydney, Teaching Award in 2001, as a finalist at the AAUT National Teaching Awards in 2003, and a Citation for Outstanding Contributions to Student Learning in 2006. Peter has been undertaking joint research with Dr Anna Reid (Macquarie University) over a period of several years in pedagogical aspects of statistics and mathematics, and of graduate 'dispositions' such as sustainability, ethics and creativity. His long term strategy in this area is to investigate and publish in the broad area of professional preparation (particularly in the area of statistics and mathematics education) and to explore the connection between students' conceptions of their discipline and study in that discipline, their perceptions of their future profession, and the links between these ideas.

Dr Larry Weldon, retired from Department of Statistics and Actuarial Science, Simon Fraser University, British Columbia, Canada.

Larry has extensive teaching, and consulting experience in Administrative Studies, Medicine, Health Services, Bioscience, Engineering, and General contexts. He has long term teaching and administrative contracts in Indonesia**,** and advanced teaching specializations in data analysis, graphics, multivariate statistics. His research publications are in biostatistics, sampling, statistics education; he has also written two textbooks. Larry's focus since 1983 has been to promote the updating of undergraduate statistics instruction to be more computer-intensive, include more graphics, more simulation, more resampling, more data analysis, and less rote inference. His web page www.stat.sfu.ca/~weldon documents this.

Professor Chris Wild, Department of Statistics, University of Auckland. http://www.stat.auckland.ac.nz/~wild

This department is the largest Statistics Department in Australia and New Zealand and is the birth place of R. Currently, Chris' main research interests are in developing methods for modelling response-selective data (e.g. case-control studies) and missing data problems, and in additive model extensions to multivariate regression techniques. A long term interest in the teaching of statistics has also developed into a research interest with particular emphasis on statistical thinking and reasoning processes. Chris is a Council member of the *International Statistical Institute*, a Fellow of the *Royal Society of New Zealand*, and was President of the *International Association for Statistics Education* (IASE) from 2003 to 2005. He is currently an Associate Editor of the *International Statistical Review*, and has been an Associate Editor of *Biometrics* , the *Statistics Education Research Journal (SERJ)*, and the *Australian and New Zealand Journal of Statistics*. He was Head of *Auckland's Department of Statistics* 2003-2007 and co-led the University of Auckland's first-year statistics teaching team to a national Tertiary Teaching Excellence Award in 2003.

Paper Refereeing Process

Papers referred to in the proceedings as refereed publication were reviewed by at least two referees selected from a panel of peers approved by the OZCOTS 2008 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 a broad interpretation of the accepted norms for reporting of research. It was expected that each accepted paper would represent a significant contribution to knowledge about statistics education or the research processes in statistical education. It was assumed that the refereed published papers for these proceedings would be substantially different from papers that have been previously published elsewhere.

Invited Paper – Chris Wild

The invited paper by Professor Chris Wild was presented live and was based on a paper by Professor Wild produced for ICOTS7. The paper is available as part of the proceedings for ICOTS7:

http://www.stat.auckland.ac.nz/~iase/publications/17/PL7_WILD.pdf

Applied Statistics as a growth engine for statistics programmes

Professor Chris Wild
Department of Statistics
University of Auckland

We explore the tensions between cooperation and competition in the context of improving the content, delivery and penetration of statistics education, and improving the health of statistics groups in universities. University education has many more parallels with business than most of us appreciate. Our environments are increasingly competitive on many, many levels. Competing well is essential for us to prosper, certainly. But it is sometimes necessary even for survival. It is suicidal for us just to be warm, fuzzy, nurturing educators who expect the world to appreciate our essential worth and reward us accordingly. We have also to be entrepreneurs and battlefield strategists. We explore models for increasing the numbers of students studying statistics, improving their educational experiences, and increasing the usefulness of the statistics education they receive. Along the way we develop sets of principles to guide our planning and operations.

Invited Paper – Adrian Bowman

Professor Bowman's paper was presented live and is not directly available as part of these proceedings. It is based strongly on his published article

Bowman A., Crawford E., Alexander G., Bowman R. (2007) rpanel: simple interactive controls for R functions using the tlctk package. ***Journal of Statistical Software*****, 17 (9).**

Statistical cartoons: the role of graphics in understanding statistics

Professor Adrian Bowman
Department of Statistics
The University of Glasgow, UK

The term `cartoons' usually suggests humorous, animated drawings, along the lines of Mickey Mouse or Charlie Brown. However, a much older use of the word refers to the prototypes or trial drawings of artistic masters such as Michelangelo, in preparation for the finished work to follow. In a broad sense, graphical insight into statistical ideas connects with both these meanings; the aim is to give students a means of exploring concepts until they are comfortable with their roles, while the ability to animate adds an extra dimension which can often spark additional interest and which can sometimes raise a welcome smile.

This talk will discuss some of the ways on which animated graphics can help in the understanding of statistical ideas at elementary, intermediate and advanced levels. The `rpanel' package for R will be used as a vehicle but other systems will also be mentioned. Over the years there has been considerable focus on illustrations of elementary statistical concepts and there are many good examples at that level. However, the scope for tools addressing more advanced topics, such as likelihood and spatial sampling, will also be discussed.

The advent of R as a standard computing environment in statistics, with increasing connectivity to other systems, makes it entirely feasible for lecturers to construct their own cartoons, rather than simply use those designed by others. The talk will argue for the importance of this mode of use.

Invited Paper (Refereed) – Peter Petocz and Anna Reid

ON BECOMING A STATISTICIAN

PETOCZ, Peter and REID, Anna
Macquarie University, North Ryde, NSW 2109, Australia

Abstract
In this paper, we summarise several components of our recent research into students' conceptions of statistics, their learning of statistics, our teaching of statistics, and their perceptions of their future professional work. We have obtained this information on the basis of phenomenographic analyses of several series of interviews with students studying statistics, both as statistics majors and as service students. In each of these cases, the broadest views relate in some way to personal connection, growth and change – in other words, they contain a strong ontological component above and beyond the standard epistemological component of learning. We discuss the importance of personal change in becoming a statistician – or an informed user of statistics – and investigate the pedagogical conditions under which such change is likely to occur.

PRELUDE: LEARNING STATISTICS – KNOWLEDGE AND SKILLS

Teaching and assessing statistical thinking at tertiary level is a very broad theme, and one, we believe, that can be addressed by looking at information from studies carried out both within the field of statistics education and also from beyond this particular field. In the broad endeavour of understanding and improving statistics pedagogy, different research questions and different research approaches shine the spotlight on different facets – for instance, important technical content, effective teaching techniques, valid and reliable assessment, characteristics of statistical thinking, anxiety about statistics, or conceptions of statistics and learning statistics. We would agree with Ramsden (1992, p.5) that *"the aim of teaching is simple: it is to make student learning possible"* and also with Barnett (2007, p.10) that: *"a part – perhaps the main part – of teaching is that of nurturing in the student a will to learn."* So we find ourselves focusing less on teaching and assessment, and more on learning itself, and particularly on *students'* views of learning. When we have discovered what makes it possible for a student to learn statistics and how to nurture in a student the 'will to learn' statistics, we have gone a long way towards answering any questions about teaching and even assessment of statistical thinking.

In the last decades of the 20th century, a sector-wide interest in improving the quality of learning and teaching in higher education resulted in research that was aimed specifically at identifying those features of learning that could be linked with improvement in learning and teaching. From the phenomenographic tradition, the description of surface and deep approaches to learning, originally identified in the context of reading a text, was already developed (Marton & Säljö, 1976), and the relation between this and conceptions of learning (Marton, Dall'Alba & Beaty, 1993) formed the research basis for Ramsden's (1992) book, *Learning to Teach in Higher Education*, and was summarised in Marton and Booth's (1997) *Learning and Awareness*.

Academics set about investigating the variety of ways that students (and teachers) understood various topic areas in order to design better learning experiences. Some early examples were the mole concept in chemistry (Lybeck *et al*., 1988) and recursion in programming (Booth, 1992): a more recent example is web-based information seeking (Edwards, 2006). This research was supported by studies that looked at students' (and teachers') experience of entire subject areas such as science (Prosser, Walker & Millar, 1995) and mathematics (Crawford *et al*., 1994): a more recent example is computer programming (Bruce *et al*., 2004). Some of these and other studies also investigated the ways that students understood learning in their discipline. Prosser and Trigwell based their text *Understanding Learning and Teaching* (1999) on empirical explorations of aspects of students' learning, incorporating Biggs' presage-process-product model of learning (for details see Biggs, 1999), to develop "a constitutionalist model of student learning" (Prosser & Trigwell, 1999, p.17) that indicated that a student's learning situation simultaneously included his or her prior experience, approaches to learning, perceptions of the situation and learning outcomes.

Although the phenomenographic, "constitutionalist, or relational, research agenda has provided new and important understandings of the character of teaching and learning in higher education" (Bruce *et al*., 2004, p.144), it came relatively late to statistics and we do not believe that it has yet reached its full potential. Within the field of statistics education, the predominant approach during the late 20th century was based on the cognitivist and constructivist tradition, exemplified by Garfield's (1995) discussion of *How students learn statistics*. Garfield discusses ideas from research in the learning of probability and statistics and lists a series of 'principles of learning statistics'. These include the notions that students learn by constructing knowledge, by active involvement in and practice with learning activities, by becoming aware of and confronting their misconceptions, and by using technology to visualise and explore data. They learn to value what they know will be assessed and learn better with consistent and helpful feedback. Many of these ideas are very similar to those espoused by Ramsden (1992) and Prosser and Trigwell (1999), though arrived at from a different philosophical background. The writings of Garfield and her colleagues have represented a strong presence in statistics education, reaching into all aspects of statistics pedagogy, including assessment (Garfield & Gal, 1999), technology (Garfield & Burrill, 1997), statistical thinking (Garfield & Ben-Zvi, 2004) and (most recently) collaborative learning and teaching (Roseth, Garfield & Ben-Zvi, 2008).

We will describe our own studies in the next section. The only other studies using an explicit phenomenographic approach that we can find in the statistics education literature are Gordon's (2004) investigation of service students' conceptions of statistics and a recent PhD thesis by Gardner (2007) investigating secondary school students' experience of learning statistics. Gordon was part of the group that carried out an earlier investigation of students' conceptions of mathematics and mathematics learning (Crawford *et al*., 1994); more recently we have collaborated on a study of teachers' conceptions of teaching in service statistics courses (Gordon, Reid & Petocz, 2007). Other researchers have used different theoretical approaches to describe hierarchies of conceptions; for example, Watson and Callingham (2003) used Rasch modelling to demonstrate a six-level, hierarchical model of students' conceptions of statistical literacy, while Reading and (Jackie) Reid have described hierarchies of reasoning about variation (Reid & Reading, 2008) and distribution (Reading & Reid, 2006). These studies did not use a phenomenographic approach, although they explicated hierarchies of students' conceptions about specific concepts – a basic phenomenographic idea! They illustrate the notion that similar results can be obtained from more than one theoretical base. This leaves us with the impression that much more could be achieved in the field of statistics education using phenomenographic approaches (in comparison, many more phenomenographic studies have been carried out in the area of computer education, including programming and information technology, from Booth, 1992 to Edwards, 2006).

Most of the studies we have discussed, though looking at the discipline and learning in a discipline from a student perspective, focus on *formal institutional* learning. In the last decade, the focus of tertiary curriculum – and also of learning – has shifted towards students' appropriate preparation for the world of work. Rather than concentrating only on what it takes to learn within specific disciplines and formal modes of study, researchers are now integrating those earlier ideas with an exploration of what it takes to *become a professional* (Pollard, 2003; Dahlgren *et al*., 2005; Abrandt Dahlgren *et al*., 2006, 2007). This shift in focus emphasises that learning is constituted by attention to both formal and informal learning activities. Subtly, the orientation of researchers (and consequently teachers and learners) is moving beyond epistemological issues to include a more ontological approach. Previously, epistemological orientations to learning and research in learning have lead to changes in the way curriculum is constructed and teaching/learning are carried out, most particularly in terms of a focus on knowledge and skills. Now, the inclusion of ontological orientations – the *"ontological turn in our thinking about higher education"* proposed by Barnett (2007, p.9) – is leading to increased awareness (by teachers and students themselves) of who the student is, how they think about themselves and how they change during and beyond the course of their studies (Barnett, 2007; Dall'Alba & Barnacle, 2007).

An interesting feature of the phenomenographic studies that we have looked at, both from areas other than statistics education and also those within statistics education, is that they seem to include at the broadest and most holistic levels students' views of their discipline and their learning that contain explicit recognition of this ontological aspect of learning. So, for instance, the early studies of conceptions of learning describe the broadest conceptions as 'seeing something in a different way' and 'changing as a person'. This is exemplified in these student quotes about learning: *"Opening your mind a little bit more so you see things (in the world) in different ways"* and *"I think any type of learning is going to have to change you … you learn to understand about people and the world about you and why things happen and therefore when you understand more of why they happen, it changes you."* (Marton, Dall'Alba & Beaty, 1993, pp.291–2).

From one of our own (non-statistics) studies, Sharmaree (all names are pseudonyms) expresses the same view in a quite passionate way. Although she was studying law, her words could have related to any discipline – even statistics:

It's just completely opened my eyes, like just completely made me look differently at the world, and it's only really now that I can see that, I think, I want to go somewhere where I can use everything that I have learnt, but I think that I have learnt so much from it that it doesn't really matter which career I choose to take, I'm always going to look at the world differently because I've studied it, I'm sure I think I'll take it with me. … And I think I just maybe aim to not ever shut my eyes again, just realise that there's always so much more than what you see going on . (Sharmaree, quoted in Reid, Nagarajan & Dortins, 2006, p.92)

Before we explore this idea in more detail, we will give a summary of our own studies and findings in the area of statistics pedagogy.

STUDENTS' CONCEPTIONS OF STATISTICS, LEARNING AND TEACHING

We have carried out a number of studies by interviewing students about their conceptions of statistics and its learning and teaching. We started with the aim of developing pedagogy that supports students' learning of statistics from their own perspectives: the most obvious way of doing this seemed to be actually asking students how they understand statistics, how they go about learning statistics and how they view our teaching of statistics. We did this by carrying out several series of in-depth interviews with undergraduate students and recent graduates, majors in the mathematical sciences including statistics, as well as students who were studying statistics as part of another discipline. In these interviews, we asked students open-ended questions such as *"What do you understand statistics to be about?"*, *"How do you know when you have learned something in statistics?"*, *"What part do you think statistics will play in your future professional work?"* and *"How does your lecturer's teaching affect your learning?"* Students' responses were investigated with further probing questions; general questions such as *"Can you give me an example of that?"* and specific questions such as *"Do you find you learn differently when you study for a test to when you're doing an assignment?"* Altogether, we carried out interviews with over 80 students during 1999–2003, each lasting between half and one hour, resulting in over 250,000 words of transcripts. The results of our investigations have been published in a series of papers (Petocz & Reid, 2001, 2003, 2005; Reid & Petocz, 2002, 2003). We have also supported this with investigations of *teachers'* views of teaching statistics and of the characteristics that make 'good' students (Gordon *et al*., 2007).

The theoretical basis for our approach was a methodology known as phenomenography: this looks at how people experience, understand and ascribe meaning to a specific situation or phenomenon (Marton & Booth, 1997; Bowden & Green, 2005). It is a qualitative orientation to research often used to describe the experience of learning and teaching, seen as a relation between the person and the situation that they are experiencing. Phenomenography defines aspects that are critically *different* within a group involved in the same situation, and its emphasis on (qualitative) variation parallels the emphasis that statistics itself places on (quantitative) variation. It is this variation that makes one way of seeing statistics or learning statistics qualitatively different from another, and allows definition of qualitatively different categories. The outcome of a phenomenographic study consists of the set of categories and the relationships between them: this is referred to as the *outcome space* for the phenomenon. Often, such categories show a hierarchical and inclusive relationship, in terms of the logical definition of the categories themselves and/or in terms of an empirical hierarchy. In the latter case, people who seem to hold the 'broadest' conceptions also show an awareness of the 'narrower' categories, while those who seem to hold the narrowest conceptions do not seem to be aware of any broader ones. This is, indeed, the reason why we, as educators, favour the broader, more inclusive categories over the narrower, more limited ones.

Here we summarise our findings and give some specific quotations pertaining to the broadest conceptions: the details, together with more quotations, are given in the papers referred to earlier. We identified six qualitatively distinct *conceptions of statistics*, which can be grouped into three levels from the most limiting (1) to the most expansive (6):

- *Focus on techniques:*(1) statistics is individual numerical activities, (2) statistics is using individual statistical techniques, (3) statistics is a collection of statistical techniques.
- *Focus on using data:* (4) statistics is the analysis and interpretation of data, (5) statistics is a way of understanding real life using different statistical models.
- *Focus on meaning:* (6) statistics is an inclusive tool used to make sense of the world and develop personal meanings.

A representative quote from Jessica shows her view of a strong connection between statistics and life in general, and illustrates the broadest conception:

[What do you find interesting or important in statistics for you?]
It's pretty relevant in lots of things. Like, they might compare cultures or something like that, and just the statistics involved in …, for example, in our exam there was a question about drugs, and it's just interesting just what they get out of statistics and how they analyse people and things and life in general from statistics. I find that interesting. (Jessica, quoted in Reid & Petocz, 2002)

Additionally, we identified six qualitatively distinct *conceptions of learning in statistics*, which can also be grouped into three levels, from the most limiting (A) to the most expansive (F):

- *Focus on techniques:* (A) learning in statistics is doing required activities in order to pass or do well in assessments or exams, (B) learning in statistics is collecting methods and information for later use.
- *Focus on subject:* (C) learning in statistics is about applying statistical methods in order to understand statistics, (D) learning in statistics is linking statistical theory and practice in order to understand statistics, (E) learning in statistics is using statistical concepts in order to understand areas beyond statistics.
- *Focus on student:* (F) learning in statistics is about using statistical concepts in order to change your views.

Quotes from Julie and Lily illustrate the broadest conception of learning as personal change. Julie almost appears to be formulating her insight into learning during the interview itself, while Lily uses an example from a different area of learning to illustrate the same point:

It changes, like, sometimes, like, like I thought always studying was about just studying to get a job, and so forth, but it changes the way you view things, and so forth; I think of things differently! (Julie, quoted in Petocz & Reid, 2003)
I guess, you look at things differently when you have learned something. Like you know, this is totally non statistically based but if you learn about photography or light or stuff like that and how light focuses, I guess you will always look at light differently. So whenever you see data and whenever you see graphs and things like that then you can look at them a little more critically. (Lily, quoted in Petocz & Reid, 2001)

When we looked at students' conceptions of teaching statistics, we identified five qualitatively different conceptions of *expectations of their lecturers*, which again can be grouped into three levels, from the most limiting (a) to the most expansive (e):

- *Focus on provision:* (a) to provide students with quality learning materials, motivation for learning, and appropriate structure for doing so.
- *Focus on subject:* (b) to explain material coherently, help students with their work and review material at appropriate stages, (c) to link statistical concepts by clarifying and elaborating on ideas and making connections between areas of the course.
- *Focus on student:* (d) to anticipate students' individual learning needs and to know the best methods of dealing with their problems, (e) to be a catalyst for students' learning by showing them the importance of statistics, helping them change their views and opening their minds.

Natasha's description of her teacher's role in changing her viewpoint sums up the broadest category:

Well, OK, different ways of looking at… well you are given data, different ways of looking at it and also helping you understand concepts and just opening your mind to… sometimes I have a one track mind so I wouldn't see a different scenario with some of the labs, different viewpoint, expanding my knowledge. (Natasha, quoted in Petocz & Reid, 2003)

The outcome spaces that we have identified and described are empirically hierarchical and inclusive. Students who described the more limiting views of statistics or learning in statistics or expectations of their lecturers seemed unable to appreciate features of the more expansive views. However, students who described the more expansive views seemed to be aware of the narrower views, and were able to incorporate characteristics of the whole range of conceptions in their understanding of statistics, and in learning and teaching in statistics (transcripts of the interviews show this clearly).

The latest interviews that we carried out were with students studying 'service' statistics as part of courses in engineering and sports science. The number of students studying statistics as part of another degree is much higher than those majoring in the discipline, so we felt that it was important to also investigate their views. The details of our analyses are given in Petocz and Reid (2005): however, our results indicate that students in professional disciplines that use statistics have essentially the same range of conceptions of statistics and learning in statistics (we did not ask them about teaching) as do students majoring in statistics. We were surprised by this finding, as we had expected differences, based on the generally accepted wisdom that service statistics students are different from statistics majors. However, these results are broadly consistent with views about statistics shown by psychology students (Gordon, 2004) and college students reflecting on their secondary school experience (Gardner, 2007).

This quote from Joe illustrates the broadest conceptions of statistics, as a way of making personal sense of the world, and of learning, as personal change through statistics. It seems to be very close in spirit to the previous quotes, despite the fact that Joe is an honours student in sports science rather than a student majoring in statistics:

[What do you think the main things are from what you have learned here in stats that you take with you when you leave?]
The whole way of thinking about things differently, you know, ideas of formulation that I would have never had come up with before, that maybe I could lay things out a little bit differently so it works, that I may not have thought of previously. It will just make my life easier basically. /…/ I think differently now because I can see now that it's much wider and can be used for a much wider range of things, as previously I may have been a bit more closed minded, thinking it was just nerdy stuff that we don't need to know. But now it's like, oh this really applies to everything. You know I can work out this, and whack these things together. Just thinking differently, thinking more advanced. (Joe, quoted in Petocz & Reid, 2005)

University students generally look beyond their classes and curriculum towards their future professional life. Their perceptions of their future profession influence their approach to their learning at university (as indeed their lecturers' perceptions of their professional world influence their teaching approach) and this link is important pedagogically. The idea of the *Professional Entity* (Reid & Petocz, 2004) developed from a recognition that views of professional work and learning, and the relationship between them, had similarities across disparate disciplines – initially in music education (Reid, 1997), then design, law, mathematics and statistics – as well as some disciplinary variation.

The Professional Entity is a way of thinking about students' (and teachers') understanding of professional work using three levels of conceptions. The narrowest is the *Extrinsic Technical* level, in which people describe a perception that professional work consists of technical components that can be used when the work situation demands it. In statistics, this is shown by a view that statistical work is concerned with gathering statistical techniques for use in different situations. At the broader *Extrinsic Meaning* level, people hold that professional work is about developing the meaning inherent in discipline objects. In statistics, this is shown by the view that statistical work is focused on finding meaning in sets of data. The broadest view is the *Intrinsic Meaning* level, in which people perceive that their professional work is related to their own personal and professional being. In statistics, this corresponds to a view of statistical work as creating and modifying views of the world based on numerical evidence.

The Professional Entity is an important idea since each of its levels corresponds with a particular approach to the discipline and to learning (and teaching) in that discipline. We have made this explicit by describing our conceptions (of statistics, learning and teaching) in three groups, each lining up with one of the levels of the Professional Entity. For example, a limiting 'technical' view of the profession of statistician corresponds with a learning focus on development of atomistic and technical statistical skills – the 'focus on techniques/provision' conceptions. By contrast, an expansive 'personal' view of the statistical profession enables students to focus their learning on the meaningfulness of statistics – the 'focus on meaning/ student' conceptions. If students are encouraged to broaden their conception of statistics and the statistical profession, they will tend to develop correspondingly broader approaches to learning (see Reid & Petocz, 2002).

POSTSCRIPT: BECOMING A STATISTICIAN – BEYOND KNOWLEDGE AND SKILLS

The theme of this OZCOTS conference is 'teaching and assessing statistical thinking at tertiary level', and as we have indicated earlier, we translate this to a focus on learning, and in particular on students' own views of their learning. The phenomenographic approach implies a theory of learning that is focused on the variation in students' conceptions of any particular phenomenon, and sees learning as the process of broadening views from narrower or more limited to broader or more inclusive conceptions. Indeed, the notion of 'deep' and 'surface' approaches to learning (essentially the same idea discussed in Garfield & Ben-Zvi, 2005) comes from early studies carried out by phenomenographers (e.g., Marton & Säljö, 1976). A whole range of phenomenographic research, in statistics and other discipline areas, indicates that the broadest conceptions of discipline, of learning and of teaching include the important aspect of personal change and connection with personal and professional life. That is, there is an acknowledgement of the importance of becoming a statistician (or a person who uses statistics and thinks statistically), as opposed to simply learning about statistics. Barnett (2007, p.18) points out that *"a 'deep' orientation towards her studies is a personal stance on the part the student in which she invests something of herself as a person; in a 'surface' orientation, by contrast, the student lacks such a will and subjects herself passively to her experiences. That is, underlying the apparently cognitive level on which the 'deep'/'surface' distinction works, is an ontological substratum"* (emphasis in the original). The hierarchical and inclusive nature of the conceptions we have identified implies that we are considering the ontological *in addition to* (rather than instead of) the epistemological aspects of learning statistics.

The connection between students' tertiary studies and their future professional work, summarised in our model of the Professional Entity, is also important in reinforcing the ontological aspects of a course of statistics study. Dall'Alba and Barnacle argue: *"what it means to be(come) a teacher, artist, physicist, historian, engineer, architect, [statistician] and so on, must be a central and ongoing question that continues to be addressed explicitly throughout (and* *beyond) higher education programmes"* (2007, p.687, our addition). Our students are becoming statisticians (or professionals who will use statistics as a component skill), and we should ensure that this remains a focus of their studies. The broadest level of the Professional Entity, the *Intrinsic Meaning* level, makes the explicit connection between personal life and professional work – a connection that can be a powerful motivation for engagement with tertiary studies (see, for example, in the context of statistics study for future engineers, MacGillivray, 1998).

For statistics major students, the focus is on 'becoming a statistician': but for service students of statistics the focus could be better described as 'becoming a competent and confident user of statistics'. Such students can be introduced to statistics as a higher-level professional component. In this context, we have argued (Petocz & Reid, 2007) that the inclusion of a range of professional 'dispositions' such as creativity, ethics, sustainability and cross-cultural sensitivity into the curriculum can increase the relevance and interest of a statistics course for statistics majors or service students alike. This can increase students' engagement with their studies, while allowing them to develop and practice such professional dispositions. In a sense, this is a broader version of Ramsden's (1992) principle that assessment drives learning. A relevant and important aspect of such professional dispositions is that they too contain a strong ontological aspect – a student aims (or can be pedagogically 'provoked') to *become* an ethical practitioner or a culturally sensitive manager rather than simply learning about ethics and internationalisation.

So, finally, how do we play our part as teachers in this process of "making student learning possible"? We would claim that we should do this by focusing on students themselves – on their being and their becoming. We do our job best by helping students to become aware of variation (that is, qualitative variation in ways of seeing statistics and learning) and encouraging them towards the broadest conceptions of their discipline and their learning, and the broadest views of their professional role as statisticians (or users of statistics). The importance of this is indicated by (a small paraphrase of) Barnett (2007, p.50): *"The degree in [statistics] confirms that the student has become, if only embryonically, a [statistician]."*

In this endeavour, there are a number of steps that we can take. First, we can actually talk to students about the range of conceptions of statistics and learning, use opportunities such as online questioning to display this variation in students' views, and encourage them to work with each other in groups in order to give them opportunities to understand each others' views. Such techniques can be very powerful: it is surprising how many students believe that all their colleagues think in the same way that they do! And of course an important step in encouraging people to change their views is to make them explicitly aware of their and others' views. Secondly, we can develop, select and utilise learning materials and teaching approaches that go beyond a focus on techniques, and even beyond a focus on the discipline of statistics itself, to include a focus on students – on the personal meanings that statistics can have for them, and on the future role of statistics in their professional life (Reid & Petocz, 2003). In this way, we can encourage their engagement with their learning and the "will to learn" that seems to be a necessary pre-requisite for successful learning. And finally we can use assessment methods that acknowledge the variability in views of discipline and learning, that encourage students to construct meaning from their learning and to make links with their personal and professional lives: we have given some specific examples at a recent conference (Petocz & Reid, 2007), including such approaches as group, peer and self-assessment. The aim is to encourage the broadest, 'partnership' conception of assessment (Shreeve, Baldwin & Farraday, 2004), where students see themselves as equal partners with their teachers in the process of evaluating and judging their own work (as opposed to the narrower 'developmental' or narrowest 'correction' conceptions of assessment). Note again that many of these recommendations are not dramatically different from those arising from other research in statistics education, though of course they have a different genesis and rationale.

In summary, we have discussed the ontological aspects of learning statistics and their connection with the broadest conceptions of the discipline and learning, and perceptions of professional work. This connection provides a way in which we can integrate ideas of student engagement and identity formation in the context of preparation for work as a professional (Abrandt Dahlgren *et al*., 2007). Dall'Alba and Barnacle (2007, p.689) conclude that learners need to *"transform as people"* in order to become professionals – statisticians or informed users of statistics – and that this requires *"educational approaches that engage the whole person: what they know, how they act, and who they are."* We have investigated the conditions under which such transformation is most likely to occur, specifically in the discipline of statistics.

REFERENCES

Abrandt Dahlgren, M., Hult, H., Dahlgren, L.O., Hård af Segerstad, H., Johansson, K. (2006). From senior student to novice worker: learning trajectories in political science, psychology and mechanical engineering. *Studies in Higher Education*, 31 (5), 569–586.

Abrandt Dahlgren, M., Petocz, P., Reid, A. and Dahlgren, L.O. (2007) Preparation for professional work? A meta-analysis of two international research projects on the transition from higher education to work life. Online at http://rwl5.uwc.ac.za/usrfiles/users/99062813/documents/Abrandt_Dahlgren_Madeleine_368.doc .

Barnett, R. (2007). *A Will to Learn: Being a student in an age of uncertainty*. Society for Research in Higher Education, Open University Press, Great Britain.

Biggs, J. (1999). *Teaching for Quality Learning at University*. Society for Research in Higher Education, Open University Press, Great Britain.

Booth, S. (1992). *Learning to Program. A phenomenographic perspective*. Acta Universitatis Gothoburgensis, Göteborg.

Bowden, J. & Green, P. (eds.), (2005). *Doing Developmental Phenomenography*. RMIT University Press, Melbourne.

Bruce, C., Buckingham, L., Hynd, J., McMahon, C., Roggenkamp, M. & Stoodley, I. (2004). Ways of experiencing the act of learning to program: a phenomenographic study of introductory programming students at university. *Journal of Information Technology Education*, 3, 143–160, [Online] http://jite.org/documents/Vol3/v3p143-160-121.pdf .

Crawford, K., Gordon, S., Nicholas, J. & Prosser, M. (1994) Conceptions of mathematics and how it is learned: the perspectives of students entering university, *Learning and Instruction*, 4, 331–345.

Dahlgren, L.O. *et al*. (2005). *Students as Journeymen between Communities of Higher Education and Work*. [Online] http://www.hewl.net/HPSE_CT-2001-0068__final_ju.pdf .

Dall'Alba, G. & Barnacle, R. (2007). An ontological turn for higher education. *Studies in Higher Education*, 32(6), 679 – 691.

Edwards, S. (2006). *Panning for Gold: Information literacy and the net lenses model.* Auslib Press, Adelaide, South Australia.

Gardner, K. (2007). *Investigating Secondary School Students' Experience of Learning Statistics*. PhD Dissertation, Georgia State University, United States.

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, 63(1), 25–34.

Garfield, J. & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking: issues, challenges and implications. In D. Ben-Zvi & J. Garfield (eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*. Kluwer, Dordrecht, The Netherlands.

Garfield, J. & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, 4(1), 92–99, [Online] http://www.stat.auckland.ac.nz/~iase/serj/SERJ4(1)_Garfield_BenZvi.pdf .

Garfield, J. & Burrill, G. (1997). *Research on the Role of Technology in Teaching and Learning Statistics*. International Statistical Institute, Voorburg, The Netherlands.

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

Gordon, S. (2004). Understanding students' experiences of statistics in a service course. *Statistics Education Research Journal*, 3(1), 40–59.

Gordon, S., Reid, A. and Petocz, P. (2007). Teachers' conceptions of teaching service statistics courses. *International Journal for the Scholarship of Teaching and Learning*, 1(1), [Online] http://www.georgiasouthern.edu/ijsotl/v1n1/gordon_et_al/IJ_Gordon_et_all.pdf .

Lybeck, L., Marton, F., Stromdahl, H. & Thalberg, A. (1988). The phenomenography of the mole concept in chemistry. In P. Ramsden (ed.) *Improving Learning: New perspectives*. Kogan Page, London.

MacGillivray, H. (1998). Statistically empowering students. *Proceedings of the Third Engineering Mathematics and Applications Conference*, 335–338. Institute of Engineers Australia.

Marton, F. & Booth, S. (1997). *Learning and Awareness*. Lawrence Erlbaum, New Jersey.

Marton, F., Dall'Alba, G. & Beaty, E. (1993). Conceptions of learning. *International Journal of Educational Research*, 19, 277–300.

Marton, F. & Säljö, R. (1976). On Qualitative Differences in Learning: I. Outcome and Process. *British Journal of Educational Psychology*, 46, 4–11.

Petocz, P. & Reid, A. (2001). Students' experience of learning in statistics. *Quaestiones Mathematicae*, *Supplement 1*, 37–45.

Petocz, P. & Reid, A. (2003). Relationships between students' experience of learning statistics and teaching statistics. *Statistics Education Research Journal*, 2(1), 39–53, [Online] http://www.stat.auckland.ac.nz/~iase/serj/SERJ2(1).pdf .

Petocz, P. & Reid, A. (2005). Something strange and useless: service students' conceptions of statistics, learning statistics and using statistics in their future profession. *International Journal of Mathematical Education in Science and Technology*, 36(7), 789–800.

Petocz, P. & Reid, A. (2007). Learning and assessment in statistics. *Proceedings of the ISI/IASE Satellite Conference on Assessing Student Learning in Statistics*. International Statistical Institute, Voorburg, The Netherlands. [Online] http://www.stat.auckland.ac.nz/~iase/publications/sat07/Petocz_Reid.pdf .

Pollard, A. (2003). Learning through life – higher education and the lifecourse of individuals. In M. Slowey & D. Watson (eds.), *Higher Education and the Lifecourse*, Society for Research in Higher Education, Open University Press, Great Britain, 167–185.

Prosser, M. & Trigwell, K. (1999). *Understanding Learning and Teaching*. Society for Research in Higher Education, Open University Press, Great Britain.

Prosser, M., Walker, P. & Millar, R. (1995). Different student perceptions of learning physics. *Physics Education*, 19, 670–675.

Ramsden, P. (1992). *Learning to Teach in Higher Education.* Routledge, London

Reading, C. & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: from a variation perspective. *Statistics Education Research Journal*, 5(2), 46–68, [Online] http://www.stat.auckland.ac.nz/~iase/serj/SERJ5(2)_Reading_Reid.pdf .

Reid, A. (1997). The hierarchical nature of meaning in music and the understanding of teaching and learning. *Advancing International Perspectives*, 20, 626–31.

Reid, A., Nagarajan, V. & Dortins, E. (2006). The experience of becoming a legal professional. *Higher Education Research and Development*, 25(1), 85–99.

Reid, A. & Petocz, P. (2002). Students' conceptions of statistics: a phenomenographic study. *Journal of Statistics Education*, 10(2), [Online] http://www.amstat.org/publications/jse/v10n2/reid.html .

Reid, A. & Petocz, P. (2003). Completing the circle: researchers of practice in statistics education. *Mathematics Education Research Journal*, 15(3), 288–300.

Reid, A. & Petocz, P. (2004). The Professional Entity: researching the relationship between students' conceptions of learning and their future profession. In C. Rust (Ed.), *Improving Student Learning: Theory, research and scholarship*, Oxford Brookes, 145–157.

Reid, J. & Reading, C. (2008). Measuring the development of students' consideration of variation. *Statistics Education Research Journal*, 7(1), 40–59, [Online] http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(1)_Reid_Reading.pdf .

Roseth, C., Garfield, J. & Ben-Zvi, D. (2008). Collaboration in learning and teaching statistics. *Journal of Statistics Education*, 16(1), [Online] http://www.amstat.org/publications/jse/v16n1/roseth.html .

Shreeve, A., Baldwin, J. & Farraday, G. (2004). Variation in students' conceptions of assessment. In C. Rust (Ed.), *Improving Student Learning: Theory, research and scholarship*, Oxford Brookes, 223–234.

Watson, J. & Callingham, R. (2003). Statistical literacy: a complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3–46, [Online] http://www.stat.auckland.ac.nz/~iase/SERJ2(2)_Watson_Callingham.pdf .

Invited Paper (Refereed) – Larry Weldon

EXPERIENCE EARLY, LOGIC LATER

K. L. Weldon
Simon Fraser University, Canada

Abstract

The motivational value for students of problem-based immersion in the process of data collection, data analysis and interpretation, is accepted by many. However, the culture of instruction through technique-based courses is still used at the tertiary level in many universities. The coverage of topics seems to trump guidance through the process of data analysis. In this paper, I suggest how to complement a problem-based experiential presentation of statistical methods with a presentation of the abstract structures necessary for future applications. A series of problem-based courses might fail to highlight the general and transferable concepts and principles that help to bring coherence to the toolbox of statistical techniques. To overcome this shortcoming one can present the logical structure – that is definitions, strategies, theoretical frameworks and justifications - to unify the collection of problem-specific methods, but only after extensive immersion in practical problems. Once students have experienced the effectiveness of the practical statistical approach, they may be better prepared to absorb the abstract generalizations.

INTRODUCTION

Statistics educators have been trying to improve undergraduate statistics instruction for decades. Some progress has been made but the forces of the status quo are formidable. One of the most frustrating constraints relates to the economics of textbook publication: few publishers will accept a script that is much different from the current market. Another constraint is the human effort required by both instructors and students to blaze a new path. Moreover, the disincentives to teaching effort, especially at influential universities, are well known. In spite of these impediments to reform, a small group of reformers is motivated to keep trying.

If students as a group had a keen interest in statistics, both teaching and learning would be more successful. In this paper, I want to encourage course designers and instructors to focus on the student motivation for the subject, even at the expense of shortchanging the student with the usual list of inferential tools. I will argue that guided immersion in real data-based problems, in contexts of interest to students, is a more effective way to produce useful learning of statistics basics than to present a logical sequence of techniques, even if the techniques are illustrated with applications as they are introduced.

The organization of the paper is as follows: The first section considers an overview of the progress of statistics education over the last quarter century. Next, the style and content of textbooks is used as a proxy for the teaching style and course content of many undergraduate courses in statistics – that the advances in textbooks have not solved the pedagogical problems. The important role of context-based motivation, "experience-based instruction", is then discussed. Next, some suggestions are presented concerning the year-levels at which context-based instruction is appropriate, and the related issue of class size is considered. Three examples of context-based teaching of statistics theory are then outlined. The final sections of the paper discuss the implications of context-based instruction for both undergraduate and graduate statistics courses.

REFORM IN STATISTICS EDUCATION

The ICOTS conferences that began in 1982 initiated a continuing international focus on the issues of teaching statistics. OZCOTS, USCOTS, ICME, and the ISI/IASE Satellite Conferences have also been a part of this activity. An unofficial theme of all the early conferences seems to be that instruction in the subject had not adapted appropriately to the expansion of statistics audiences from math majors to all majors. An additional theme of the more recent conferences seems to be that the changes associated with statistical software availability have not been adequately absorbed into undergraduate curriculum and pedagogy. In fact, an overarching theme is the lack of adaptation to changes in statistics instruction to reflect the changing practice in the discipline. As a participant in ICOTS 2, I joined the rising voices asking for change, and there were many good ideas being proposed in 1986. Consider the following quotes from ICOTS 2 in Victoria, BC, 1986:

"The development of statistical skills needs what is no longer feasible, and that is a great deal of one-to-one student-faculty interaction ..." (Zidek 1986)

"The interplay between questions, answers and statistics seems to me to be something which should interest teachers of statistics, for if students have a good appreciation of this interplay, they will have learned some statistical thinking, not just some statistical methods." (Speed 1986)

"Using the practical model [of teaching statistics] means aiming to teach statistics by addressing such problems in contexts in which they arise. At present this model is not widely used." (Taffe 1986)

"To take advantage of these developments, one must recognize that, while most statistics professors like statistics for its own sake, most students become interested in statistics mainly if the subject promises to do useful things for them. I believe that even the seemingly limited goal of developing "intelligent consumers of statistics" is best attained if students try to produce statistics on a modest scale. Only then do most students seem to become sufficiently intrigued with statistics to want to learn about statistical theory." (Roberts 1986)

These ideas span different parts of the problem: the need for interaction of students with experts in statistics, the need for students to learn the whole process of statistics from verbal questions to verbal answers, the need to incorporate context into students' experiences in statistical analysis, and the need to excite students about applications before presenting the theory. As an index of the extent to which these suggestions have been adopted, consider their impact on current textbooks. I suggest that the impact has been very slight, partly because the suggestions all relate to the process of teaching rather than the techniques to be learned. In fact, it is hard to imagine how a textbook would be written that would help the instructor with the above recommendations. One might conclude that a good start to implementing the recommendations is to contemplate abandoning the dependence on a textbook, at least for the sequencing of course topics. For undergraduate courses, the current textbooks could be used as a reference resource for students, rather than as a course outline. The text assigned to the course could remain as in a traditional course, but the instructor could change completely the role of the text.

I cannot summarize all the recommendations of the series of ICOTS, ISI, and ICME conferences any better than to quote from Brian Phillips report of David Moore's Invited address to the ICME 1996 conference (Phillips(1996)):

In discussing what helps students learn, [David Moore] listed the following:

Hands-on activities	Explaining reasoning
Working in small groups	Computer simulations
Frequent and rapid feedback	Open questions real settings
Communicating results	Learning to work co-operatively

Even though many of these ideas were discussed in earlier statistics education conferences, they were still "new" in 1996 as they are today. The question I wish to ask readers to consider is how well modern textbooks incorporate these strategies, and also whether it is possible for a textbook to provide for all these strategies. I suggest experiential learning of the kind proposed in this paper is one way to incorporate all these strategies, and that textbooks should pursue an important but limited role as reference agents for students, and not as lecture guides for instructors and students.

MORE THAN TEXTBOOK REFORM

Lovett and Greenhouse (2000) review and update the psychological research on course design in statistics to make recommendations for curriculum reform. The reforms they highlight are listed in their paper as "Collaborative Learning, Active Learning, Target Misconceptions, and Use of Technology". "Collaborative learning" refers to learning in teams and peer discussion. "Active learning" includes exploratory investigation and data collection, the "target misconceptions" item is described more fully as "confront students with their misconceptions", and "use of technology" means allowing students to use statistical software for both calculations and exploration. Few instructors would find these suggestions startling. However, it should be noted that textbooks do not help much with any of these reforms. What do textbooks say about teamwork, learning based on exploration, or confronting students with their misunderstandings? More and more textbooks do encourage "use of technology" although even in this reform area, exploratory use of software is less often proposed than are demonstrations or prescribed calculations with software. The implementation of the recommended reforms requires much more than reforming the textbook.

Moving away from the textbook, or a sequenced curriculum based on a list of techniques, raises many questions for the course designer. Some of the things we want students to experience are unobservable, and the final examination that tests the outcome may not be, by itself, the perfect instrument for guiding the learning. One suggestion to improve the situation has been proposed by Wessa (2008): providing an archive that captures students calculations as well as their calculation outcomes. In fact, the facility provided by Wessa facilitates communication between instructor and student, and among students as well, about the actual calculations under discussion. This removes the concern about arithmetic, and replaces it with a focus on method and interpretation. A pedagogical benefit is that the instructor does not have to force the student into one particular mode of calculation, and this recognizes that there are often a variety of ways to extract information from data. Wessa's facility is based on R but the user does not need to know R to use it. There is no charge to use the facility. To fully understand the potential of the Wessa facility, it is necessary to explore the website www.wessa.net. However, Wessa(2008a) has reported informally some quantitative evidence that involving students in discussion of quantitative strategies does actually improve their score on objective examinations, in which the examinations aim to assess conceptual understanding.

Another way to assist the instructor in moving away from the textbook as a lecture outline, is to provide a reference-friendly electronic textbook for student access. Students depend increasingly on clickable sources. One excellent example of this is the freely downloadable text called CAST (Sterling (2006)). With searchable key words and a detailed table of contents and index, this provides easy access to text material. Because it is electronic, optional links for further information are available. Another benefit of the electronic source is the multitude of java-based animations and parameter sliders. It is just more fun to use than a paper text!

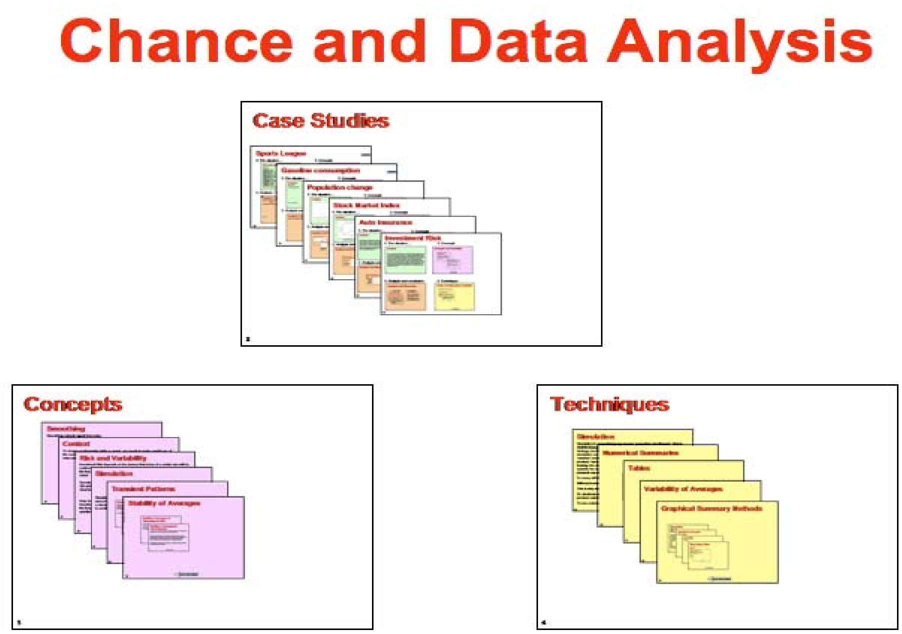


Fig. 1 Display from Roadmap Tools version of STAT 100 (Weldon and Carr (2008)). Each icon provides links to a more detailed display, and cross links are facilitated.

An interesting development making an experience-based course feasible is the Excel-based Roadmap Tools (Carr 2008). This software allows the instructor to prepare notes outlining the "experiences" and at the same time capturing the techniques and concepts illustrated as they are met in the explorations of the experiences. Fig. 1 gives a hint of the display: imbedded slides are brought forward by clicking, and communication with the instructor or other links are also enabled through this medium. In reviewing a technique, it is possible to return with a click to the case study in which the technique was introduced, or more than one case study if appropriate. Similarly for the concepts arising in the case studies. Or , if the student wants to know what techniques and concepts were supposed to have been learned from a case study, the linkage is there to provide the information. Students can use the display initially to access the case studies, and subsequently to ensure that they have mastered the intended techniques and concepts. This is one way technology can help students stay organized within a case-study or experiential approach.

An example of an electronic textbook designed for reference rather than as a lecture guide is the handbook provided jointly by NIST and SEMATECH at

http://www.itl.nist.gov/div898/handbook/

With such a helpful electronic resource, students should be less dependent on a traditional textbook.

The benefits of experiential learning have been recommended for many years, notably as a common theme of ICOTS 4 (1994) sessions in Marrakesh. Don Bentley's "Hands-On and Project-Based Teaching" (Bentley, 1994) was so popular that it had to be broken into three sessions. However, the adaptation of this strategy into the current curriculum context seems to be problematic, since it is still not a common strategy at the tertiary level. An example of the creative atmosphere surrounding these sessions is the following quote from an abstract of Allan Rossman (1994): "In this presentation, I describe a project which takes this approach to the extreme of abandoning lectures completely". His abstract of the paper titled "Learning statistics through self-discovery" concludes with "The goals of these activities are to create a more enjoyable and productive learning environment as well as to deepen student's understanding of fundamental statistical ideas." However, university traditions seem to require lecture schedules and student-faculty interaction, and so a practical problem is how to incorporate the widely-recommended project device into a traditional lecture course. The proposals of this paper suggest a way of incorporating experiential learning into the undergraduate statistics curriculum.

EXPERIENCE-BASED INSTRUCTION

To a mathophile, logic is beautiful, but most practitioners of statistics are not mathophiles. We need to keep that in mind when we are directing our pedagogical efforts toward students of statistics. We want to attract future practitioners of statistics to our statistics courses. What device can we use to show the charm of statistics without losing the underlying logical structure? I will argue that the logic of statistics can be instilled subversively by seducing students through immersion in the process of context-specific, data-based "discovery", and only later providing the logical framework that is more generally applicable.

Of course, this approach is not new. Not only the ICOTS 4 concentration on the idea in 1994, but an intriguing compilation of student-conceived projects is recorded in the locally published volume in 1997 by MacGillivray and Hayes "Practical Development of Statistical Understanding". The report records the results of student-selected projects which satisfied the criteria for an problem-based statistics course. Although each student would have primary responsibility for only one project, and not a sequence of projects, the resource does suggest the richness of student-selected problems for motivating learning of the entire process of statistical analysis.

Even earlier, Tukey (1977) emphasised the importance of involving students in data analysis unencumbered by assumptions of parametric models. His emphasis on visualization and an exploratory approach was revolutionary at the time. He felt that students needed experience with data more than knowledge of formal methods of parametric inference. But the project-based approach that would provide this experience tends not to be used – most undergraduate courses still follow closely textbooks organized by parametric technique. When the project-based approach is used in a text-dependent course, it is thought to be an add-on rather than the main driver of exposure to statistics. The suggestion here, as it was in Rossman(1994), is that with adaptation, it can be the main driver, and that there are compelling reasons to consider doing it this way.

The mathematical culture of statistics instruction is pervasive. In this culture we think it is obvious that, to teach statistics, we need to start with basic definitions, follow up with basic tools, and build on these basics to construct the commonly used strategies of statistics practice. However, if we apply this seemingly obvious approach to other disciplines, it does not seem so obvious. For example, to teach conversational Spanish, we would start with vocabulary, grammar and pronunciation, and after a long period of becoming familiar with these skills, encourage students to converse. But immersion programs show that the formal phase of instruction works best after a lengthy immersion in motivated oral practice. Likewise, English grammar is best taught to children after they can speak English! Or, as another example, consider the math approach to teaching social geography: start with definitions of urban, climate, transportation corridor, enumeration area, etc. and get to the human impacts much later. To engage students' interest, it might work better to talk about human impacts first, and get to the formal definitions later. This same approach might not work for mathematics instruction, but it may well work for statistics instruction.

Whether a statistics course is designed as a service course or a mainstream course, the content tends to be technique-based rather than problem-based. Textbooks encourage this approach, and both students and instructors find textbooks a useful guide. Within the style of technique-based courses, many strategies have been devised to increase interest in the presentation of the techniques: data-collection projects, personal-data comparisons, in-class presentations, computer-based games, computer-quizzes with feedback, and simulation by applets or statistical software programs. These strategies certainly improve the likelihood that students will learn the techniques, and in some cases will increase interest in the techniques. But absent from these many strategies is the thrill of discovery: unexpected findings or anomalies that may have a bearing on the information gained from the data. What is often missing from traditional courses is the opportunity to use general intelligence, in combination with techniques learned from past experiences, to uncover information from data. How many students of statistics are aware of the fact that most applications do not use the "standard" techniques without adaptation?

An alternative to the technique-based course is the project-based course. The obvious argument against a project-based course is that the students will find the collection of techniques associated to be a jumble of unrelated tricks, rather than a logical sequence of strategies. But just as with geography or language, the formalities are a lot more useful to students after the students have an in-depth exposure to some examples. If students have been motivated to wonder what it takes to decide if a differential group effect is consistently reproducible, or transient, then they will be interested in the concept of a hypothesis test to tidy up the confusion. But the understanding of the information dilemma really needs to be internalized before this tidying is appreciated. Most instructors would agree that both theory and application need to be covered in a statistics course for the course to be useful to students. However, the mathematical approach of theory-first, application-after may not work as well as application-first, theory-after. The reason is that for the vast majority of students, it is easier to arouse interest in an application than in a theoretical concept.

Some instructors routinely use application examples to introduce statistical techniques, which is a reasonable way to motivate theory. But if the guiding theme of such a course is a series of application-technique pairs, this has not really deviated from the technique-oriented course. For application immersion, one needs an application that engages students for several days or weeks, with data to analyze, external background information to track down, feedback on analytical techniques suggested by students, opportunities to verbalize what the question is and what the findings are, and time to read up on similar studies that have been done. Classes are an opportunity for the instructor to suggest multiple approaches, invite criticisms, provide feedback from past suggestions of students, and ask questions for students to work on outside class. Only after two or three such projects are completed (maybe over 15 hours of instruction) would the instructor have enough technique material to outline the logical links among the techniques: types and numbers of variables, formal and informal analyses, estimation vs testing, prediction vs modeling, experiments vs experiences. The theory would be presented as a *simplifying* device rather than as an perplexing abstraction.

Students in "mainstream" courses of statistics (courses that would not be called service courses) generally have some tolerance for a mathematical approach to statistics. In particular, students of engineering, natural science, business, or psychology are required to have some math basics in their programs. But do these students actually like the mathematical, logical approach to their subjects? Is the chemistry student more interested in the periodic table or the relationship of sugar and alcohol? Is the business major more interested in the definitions of credits and debits or the earnings growth rate of Google? Mathematically-oriented students might actually be more interested in the periodic table and definitions of debits and credits, than the more applied topics, but such students are a small minority of the ones we want to introduce to statistics. As Cleveland (1993) says:

"A very limited view of statistics is that it is practiced by statisticians. … The wide view has far greater promise of a widespread influence of the intellectual content of the field of data science. "

If we accept that the target audience, even in mainstream courses, is users of statistics, rather than the statisticians that are a subset of this group, then we need to think about how to make statistics courses attractive to future users. While some advantage can be had from making the course process a pleasant social experience, with team assignments, and convenient venues, an important draw is to have very interesting content. To say that the logical, sequential approach to the introduction of statistical techniques is the only viable one is to deny the importance of motivation in learning. It is important to make the introduction of statistical ideas fall out of stimulating investigations into real data-based questions. The logical structure of statistical strategies can come later.

Improved learning as a result of heightened motivation is just one reason to use in-depth exposure to applications. Another reason is that students need to learn the *process* of statistical investigation, and not only the strategies and tools. This process involves many attitudes that need to be learned: the presumption that more subject-matter knowledge may help; the realization that modeling and analysis is to some extent a trial-and-error process, the appreciation of the dangers of overfitting or overanalysis, the importance of graphical methods for identifying anomalies and summarizing results, the availability of resampling methods when standard methods fail, and the appreciation of power in summarizing findings. These things are hard to reduce to single lessons – demonstration and repetition are required. Guiding students through immersion in real data analysis exercises, right from the problem formulation stage to the ultimate report of findings, is one way to "indoctrinate" students in this process.

The efficacy of experiential learning has support from the psychological literature. The "constructivist" approach to learning involves social collaboration and communication and individual responsibility and experimentation. Although the concept of constructivism is quite old, course designers are still working toward courses which incorporate all the aspects identified by recent researchers. (Moreno, Gonzalez et al (2007)). Another recent example of support for this approach comes from Konold (2007). He argues for "bottom-up" instructional design rather than "top-down". The idea is that for effective teaching, we need to start with the student's current context, and use this as a starting point for introducing new ideas. Experiential learning, including exploratory data analysis, student-choice projects, and verbal reports, does seem suited to bottom-up instruction.

A recent and extensive bibliography of the contribution of educational psychology to pedagogy in statistics is given by Garfield and Ben-Zvi (2007). In reviewing this bibliography, one is struck by the long list of difficulties experienced by students in learning from traditional technique-based courses. It does seem that experience-based pedagogy deserves a greater emphasis than has been usually practiced so far.

An important contribution to the role of experiential learning in the computer age is explored extensively in the series of articles recently published in the International Statistical Review. A lead-in to this series is provided by Seneta and Wild (2007). Although the emphasis is on computer-based learning environments, the important role of experiential learning is highlighted.

INTEGRATING EXPERIENCE-BASED COURSES INTO UNDERGRADUATE EDUCATION

Where does such an "experience" course fit in the undergraduate curriculum? I think the approach can be used at all levels. In recent years I have initiated a few new courses into the offerings at my university. Although they were proposed for various levels of student, they all have the feature that they are a series of examples rather than a logical sequence of techniques. I'll give a brief overview of these courses, since their style is close to the style I am recommending.

STAT 100: "Chance and Data Analysis". At the first year level, data relating to accidents by young drivers, sports leagues, blue whales, and the stock market, and others introduced ideas of causality, time series smoothing, simulation, sampling surveys and survival analysis. Of course, the basic definitions of means and standard deviations, frequency distributions, sampling variation, and scatter diagrams are introduced and repeated many times, in the context of the examples discussed. By the end of the course, the student has had all the techniques normally included in a first course although perhaps less drill than usual.

STAT 300: "Statistics Communication". After two years of basics, students can begin to comment verbally on what they have learned. More specifically, they should be learning how to explain why certain techniques are appropriate in a particular data analysis, and what the analysis really shows. This course asks students to critique or defend criteria like "unbiasedness" or "minimum variance", to comment on the use of hypothesis tests when the sampling frame is uncertain, to discuss how to report anomalous data, and how to present orally or in writing what the real findings are in an instance of data analysis.

STAT 400: Data Analysis. The approach in this course is to ask students to suppose that the information in the data is more important than the techniques normally used to analyze the data. This helps the student understand that statistical practice is problem-based, and students are expected to use all their knowledge and intelligence to get at the information in the data. Of course, it helps if the content of the examples is of intrinsic interest to students. I have used a badly designed internationally funded agricultural study to draw attention to the value of good design as well as to provide an opportunity for students to try to rescue some information from the study in spite of its bad design. Another example uses a profit maximization strategy for a wholesale distribution network, in a situation where demand data is censored by inventory. Other examples use the data sets from Cleveland's *Visualizing Data* Book for which multivariate graphing provides a visual approach to information retrieval. Creativity in analysis is encouraged. These experiences lead to the use of resampling techniques, simulation to assess trial and error solutions, graphical smoothing in one or more dimensions, context-guided strategies to avoid the pitfalls of stepwise regression, and more.

Students find these courses both challenging and rewarding, judging from the feedback provided as part of the department's routine evaluation procedures. They are challenging because they require the student to move away from mere textbook knowledge, and they are rewarding because they confirm that a student can integrate their intelligence with the techniques they have learned to produce useful information from data.

Of course, the big question concerning this type of course is: When do the students find out about the logical structure of the discipline? There are different approaches for students with different needs. For students who take only one course, it may be futile to try to convey the logical structure. Perhaps for this group it is better to convey an appreciation for the utility of statistical strategies, rather than the basic tools and concepts themselves. For students who take more than one course, the subsequent courses can supply the logical structure – if the students appreciate the utility of the subject from the problem-based course, they will be both motivated and receptive to the more formal approach. However, an alternative is to include this step within a problem-based course. For example, each module of say, ten contact hours, can be followed by a "what tools have we learned" session with the logical structure emphasized. As mentioned earlier, this phase can be described as a simplification of the apparently chaotic collection of tools introduced for a particular problem. The website for STAT 100 at www.stat.sfu.ca/~weldon includes some examples of this approach. STAT 300 and STAT 400 are imbedded in course sequences that include more formal courses, and so the logical structure is, to some extent, left to these other courses.

THE CLASS SIZE CONSTRAINT

The ideal of 1-1 instruction is clearly impractical at the undergraduate level. But small classes that allow discussion can sometimes be afforded. Over the last few decades, with the ubiquitous spread of data-based research into most disciplines, undergraduate class sizes have grown to one hundred and more, making discussion during a class meeting a rare event. How can students be exposed to the whole process of data analysis in a setting of large-class lectures? The efficiency of large classes may be an illusion in the case of statistics.

Various strategies have been used to try to solve the lack of student-faculty interaction that occurs with large classes. Small group tutorials is one common approach. Tutor-assisted group projects such as the ones documented by MacGillivray (1997) is another. Group assignments in which students help each other and thus reduce the need for faculty help is a third approach. But the guidance in the complete process of data analysis is most effective if an instructor experienced in both tools and applications has frequent interaction with students. Instructor-to-student lectures, as are common with large classes, do not provide this interaction. Strategies involving group work outside of lectures (as just mentioned) do provide some benefit. But the small class ideal would be best at allowing the instructor to balance the motivation of student exploration with the provision of guidance in the most effective tools and strategies.

One recent report (Carnell (2008)) which attempted to gauge the impact on learning of a single project in an introductory stats course, found that this project addition did not make an appreciable difference in learning outcomes. Perhaps several projects are necessary. If a project is seen as an extra, it may not be treated the same way as if projects are the main drivers of the course.

In a world of large classes in statistics, how does one move in the direction of experience immersion as a teaching device? One way is to have whole courses that are taught to smaller classes at the advanced stage (like STAT 400). Another is to try to create the experience in the large class by abandoning the "technique-coverage" approach and instead describe for students the process of development from idea to report. In such an approach, the assumption is that enough tools and strategies will be covered incidentally to the case-studies described to satisfy the programs that require one course in statistics. (like STAT 100). The justification for this shotgun approach is that it is better to have an understanding of a few common tools, than little understanding of a complete toolkit. However, it may be necessary to convince university administrations that the best statistics education requires small classes. For this to happen, the public view of the discipline of statistics may have to change from a necessary evil to that of a creative and vital subject! If students become excited about the small-class statistics course they are taking, then perhaps the message will get through to administrators that the subject is worth the higher price. So this is another reason to focus on drawing students into the subject matter with material that seems immediately interesting and useful.

EXPLORATORY CONTEXTS:

These days the internet provides a wealth of good examples for teaching material. In fact, there are some conventionally published sources as well: the text mentioned previously, MacGillivray and Hayes(1997), details the teaching experiences in 19 different application scenarios: from fishing to motorcycle accidents to Murphy's Law. The availability of ideas for projects is useful, but the pedagogic effectiveness of an example really depends on how it is presented. Consequently, the instructor's role is still key to the learner's outcomes.

The suggestion in this paper is that exploratory data analysis, suitably guided, will lead a student to understand basic statistical strategies, and the student will learn the basic statistical strategies more thoroughly in a given time frame than if the same strategies are presented in the more conventional way. The reason is that the student will be motivated by the obvious relevance of the strategies since they will be introduced as the data exploration requires them. But will the student be able to apply the strategies to new contexts? This is where the instructor's role is crucial. After several data exploration examples have been worked through, the instructor needs to make sure that the student has the big picture. This is where the logical relationships of the techniques need to be presented.

For students to be able to use their learning of statistical tools and concepts in new contexts, they do need the logical structure clearly in mind. For example, they need to be aware of the different scales of measurement, of the different ways comparisons can be made, and of the difference between parameter estimation and testing parameter credibility. But, as we have argued, to try to teach these in a sequence of techniques has not worked well – better to have them as a framework for techniques motivated by data-based projects, when there is a readiness for aggregating the pieces learned.

To further illustrate the potential of teaching techniques through experience immersion, I will briefly describe three examples. The first one should appeal to students since they get to choose an activity from their own lives. The second one has the advantage of relating to local conditions. The third one relates to a personal characteristic that students hold subconsciously and would often be of interest to a student in comparison with others. All of the examples could be used in a first course, or in an advanced course. Of course, statistics courses with a subject area focus (e.g. life sciences, business, engineering or psychology) would likely include examples more closely related to the subject area, and would build on a more specialized student background. But these examples will suffice to illustrate the point of conveying useful statistical theory through comprehensible applied projects with real-world contexts.

Example 1: Sports Leagues

Students often have at least one sport they are interested in, either as a participant or a spectator. Team sports have the feature that game results are accumulated throughout the season and teams are repeatedly ranked using some points system. Suppose students are given the task of finding the accumulation table of a currently operating league, and commenting on the relative quality of the teams suggested by the table. Students should be advised to choose a sport that interests them, if possible. Questions for discussion might be:

1. If team A has more points than team B, does team A have a better than even chance of winning the next contest with team B?
2. Is there any evidence of a home team advantage?

3. If all the teams have the same chance to win each game, what would the league ranking look like?

Note that many students will have an opinion regardless of their statistics knowledge so far, so a discussion should be easy to stimulate. Where would the discussion lead?
Here are just a few of the possibilities:

a better understanding of "better than and even chance"
a realization that current rankings are, at least in part, subject to "luck" or "random variation"
an opportunity to test a hypothesis by observing data
consideration of conditional probability
an appreciation that randomness can deceive and often does
an opportunity for answering a question via simulation (by coin or computer)
a need to define a measure of variability (in point status)

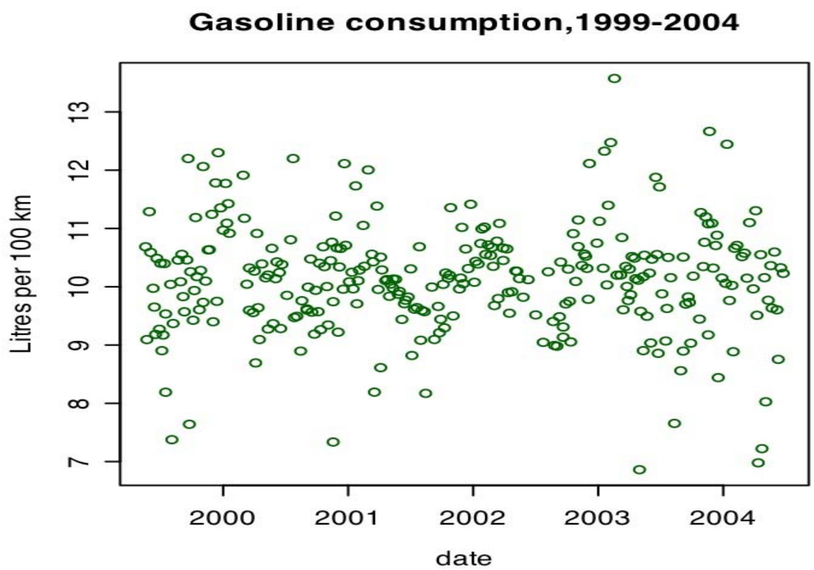
The point of this example is that a context of interest to students can be the platform for introducing many important statistical tools and strategies, and because the answers to the questions are of interest to the students, the tools and strategies that help to get at the answers will also be of interest to the students. The motivation for learning statistics is based on a genuine interest and not only on the need for a good mark in the course. The learning will include the entire process of data-based study and not only an artificially simplified context. Moreover, the freshness of the discussion should make the process stimulating for both student and instructor.

Example 2: Auto Fuel Consumption

An example that I like to use is based on some personal data I accumulated on gas consumption for my car during a five year period. The graph shows the data:

Students are asked if they find any interesting or useful information in this graph, and are asked to analyze the data to see if there are any "trends" or anomalies. The initial response is usually negative. It turns out that fitting ordinary polynomial regression reveals nothing much, just as a visual scan would suggest. But any kind of nonparametric smoothing, even a moving average, shows a nice sinusoidal pattern in sync with the annual seasons. This pattern raises the question of the likely cause, and the many potential explanations include usage, temperature, precipitation, traffic, and tire pressure. Note also the likelihood of a negative serial correlation as a result of the way gas consumption is measured. What does the student learn from this experience?

not all regular patterns in time can be discerned by eye
a correlation can have many causes, and often further data collection is suggested
nonparametric smoothing is a useful exploratory technique
not all interesting data is a sample from a population (time series)
the measurement context must be examined as part of the analysis



Example 3: Crossing of Arms and Clasping of Hands

MacGillivray (2007) reports a project in which a large class of students were asked to report their normal way of crossing their arms and clasping their hands. Students are intially surprised that there is a "natural" way to do these things that is personal for each student. There is the question of whether gender, or handedness, explain the differences, and whether arm crossing is related to hand clasping. An advantage of this example over the other two is that, even if one is an unusual student with no interest in sports or driving, one is still likely to practice arm crossing and hand clasping. The referenced paper gives the full detail of this example, but some of the obvious lessons one learns from the experience are:

the procedure of developing an idea into a data-based study is a non-trivial exercise
careful definitions and protocols need to precede data collection
two features can be related even when exceptions exist
categorical data is summarized by frequencies
descriptive techniques need to be used before inferential techniques
descriptive techniques can often reveal unexpected findings
"obvious" relationships are sometimes not confirmed by data
apparent relationships can be deceiving and require testing for reproducibility

Looking over the twenty "lessons learned" from these three examples, even though they are a partial list, should suggest the richness of the learning experience with respect to statistical practice. Are these lessons as useful to students as the ability to fit a line to x-y data, or test if two groups have the same mean? While students do need to learn about the calculation methods, they need these meta strategies as well to have a useful education in statistics.

These three examples suggest the broad spectrum of statistical tools and strategies that can be conveyed through immersion of students in stimulating applications. One only has to consider how the lessons-learned from the examples would be taught one-at-a-time in a logical sequence to see that the logical approach would lack the charm of the experience immersion. If we want to attract and retain students' interest in statistics, we need to consider charm! And, if we want students to understand the whole process of data analysis, we need to give them experience with the whole process of data analysis.

TARGET AUDIENCE FOR STATISTICS STRATEGIES

The mathematics model of teaching statistics is a product of the 20^{th} century, math-based history of statistics. It seems to work for those few students who relish mathematical abstraction, and the current cadré of statistics instructors are mostly drawn from this group. But today modern statistics is practiced by a wide spectrum of engineers, scientists, and social scientists, and these users need more than a superficial knowledge of statistical strategies. STAT 100 does not adequately prepare these users. These users must be able to identify opportunities for data-based studies, plan data collection, explore data, extract valid information from data, and defend what they have done in words. It is this large group of future practitioners of statistics that needs the most attention at the undergraduate level, not the stat major. The stat major can benefit from math courses, even math courses with little statistical overlap. But the future statistics practitioners need to know how to use statistical software to explore data, and how to allow for study design shortcomings in coming to conclusions.

Practitioners educated in applied disciplines will not be able to handle all data-based problems they meet, and will find it necessary to seek the help of expert statisticians. However, to take advantage of this expertise, they must recognize the opportunity. Do the technique-based courses give students the insight needed to know when an expert can help? A student who has met a situation in which the perfect test is unknown, such as would be likely to happen in an exploratory study, may be willing to admit the need for expert help in later practice.

To become expert statisticians, graduate work would usually be required. This is where the full confluence of mathematics and statistics should be explored. Instead of designing a special undergraduate education for stat majors (future statisticians), it might be efficient to give all the statistical practitioners the same undergraduate education in statistics, but require further mathematical statistics in graduate courses. With this approach, undergraduates learn to appreciate statistics as a vital subject relevant to many careers, and are not deceived into thinking of statistics as a specialized form of mathematics; and graduates will not arrive at the serious study of statistics with a naive view that statistics is a bundle of calculation tools, and will realize that the role of mathematics in statistics is to allow adaptation of available methods to suit particular application contexts.

Mathematics is a powerful technology for clarifying complex ideas, and is essential for expertise in many "applied" disciplines. Theory in engineering, management science, environmental science, and many other fields is assisted by mathematics, but these theories are not only mathematics. The same is true of statistics. It is now clear that statistics is a separate discipline from mathematics, even though this was not always recognized in the past. We need to judge statistics expertise by its relevance to the extraction of useful information from data, and not by the mathematical expression of its tools. Our teaching of statistics should reflect this criterion.

SUMMARY

The content of undergraduate courses in statistics has not changed very much in spite of the reform movement of statistics educators. Better textbooks, and the provision of computer software, have changed the tasks of the student, but objectives as revealed through tests and examinations have not kept up with the recommended reforms. What seems to be missing is the immersion of students in the entire process of data-based research along with frequent interactivity with the instructor. The motivation provided by interesting projects (suggested by the instructor or the students themselves) is an important factor in shaping the image of the discipline, as well as being a powerful stimulus to learning. It may be necessary to depart from large classes to accomplish a useful exposure to statistics: higher education administrators need to see undergraduate statistics as concepts and strategies rather than procedures and formulas. Statistics courses need to include more experience and creativity and less coverage and dogma.

Of course, balance of the old with new is probably the optimal strategy, but change has been so slow that extreme measures need to be contemplated. The basic recommendation in this paper is to provide students with exciting experiences in extracting information from data, and after the student is completely amazed by the many surprising and useful strategies of statistics, proceed to provide the formal structure of the techniques used, including the mathematics as required. We need to recognize that the main audience for statistics at the undergraduate level is future practitioners of statistics, and the principal secondary market is for statistics appreciation courses. The training of future statisticians should not be considered and undergraduate mission. Future statisticians need the practical knowledge of the undergraduate education, as well as graduate work in mathematical statistics. Our focus in undergraduate education should be on experiential immersion in data-based discovery. The conveyance of the logical structure of formal statistical inference will to some extent be achieved simultaneously, but in any case should be relegated to the status of a secondary goal.

REFERENCES

Bentley, D. (1994) Sessions 4A,B,C in *Proceedings of the Fourth International Conference on Teaching Statistics*. Vol 1: pp 17-31. The National Organizing Committee of the Fourth International Conference on Teaching Statistics. Rabat, Morocco.

Carnell, L.J. (2008) The Effect of a Student-Designed Data Collection Project on Attitudes Toward Statistics, *Journal of Statistical Education* v16/n1 www.amstat.org/publications/jse/

Carr, R. (2008) Roadmap Tools for Excel. www.deakin.edu.au/~rodneyc/roadmaptools.htm

Cleveland, W.S. (1993) Visualizing Data. Hobart Press, Summit, NJ.

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) pp 372-396.

Konold, C (2007). Designing a tool for learners. In M. Lovett & P. Shah (Eds.) Thinking with data (pp 267-291). New York, Taylor & Francis.

Lovett,M.C. & Greenhouse, J.B. (2000) *The American Statistician*: 54(3) 196-206.

MacGillivray, H. & Hayes, C. (1997) Practical Development of Statistical Understanding: A Project Based Approach. Queensland University of Technology. Brisbane.

MacGillivray, H. (2007) Teaching Statistics: 29(2): 49-53.

Moreno, L., Gonzalez, C. et al (2007) Applying a constructivist and collaborative methodological approach in engineering education, *Computers & Education*, 49 (2007), 891-915

Phillips, B. (1996) Plenary Lecture Given By David S. Moore (USA) "New pedagogy and new content: The case of statistics". *Proceedings of ICME 8*, Seville, Spain.

Roberts, H. V. (1986) Data Analysis for Managers. *ICOTS 2*. pp 410-414. Victoria, BC.

Rossman, A (1994) "Learning Statistics Through Self-Discovery". Abstract in *Proceedings of the Fourth International Conference on Teaching Statistics*. Vol 1: p30.

The National Organizing Committee of the Fourth International Conference on Teaching Statistics. Rabat, Morocco.

Seneta, E and Wild, C. (2007) Preface. *International Statistical Review* v75(3) 279-280.

Speed, T.(1986) Questions, Answers, and Statistics. *ICOTS 2* pp 18-28. Victoria, BC.

Sterling, D. (2002) Computer Assisted Statistics Teaching (Version 2.1) and for an up-to-date version go to cast.massey.ac.nz/

Taffe, J.(1986) Teaching Statistics: Mathematical or Practical Model. *ICOTS 2* pp 332-336. Victoria, BC.

Tukey, J.W. (1977) Exploratory Data Analysis. Addison-Wesley. New York.

Wessa, P. (2008), Free Statistics Software, Office for Research Development and Education, version 1.1.22-r5, URL http://www.wessa.net/ See also www.freestatistics.org/index.php?action=10 for more information.

Wessa, P. (2008a) Personal Communication.

Zidek, James V. (1986) Statistication: The Quest for a Curriculum. *ICOTS 2* pp 1-17, Victoria, BC.

Invited Paper – Michael Martin

Reproduced here are the PowerPoint slides from the presented paper. The paper is based on the work published in the *Journal of Statistical Education* **(2003), Volume 11(2) online. http://www.amstat.org/publications/jse/v11n2/martin.html**

What lies beneath: inventing new wheels out of old

Professor Michael Martin
Australian National University

Analogical thinking is a powerful cognitive tool that leverages knowledge and understanding of familiar ideas and relationships to form knowledge and understanding in a new setting. For students approaching their first statistics class, fear of the unknown can be a major factor in slowing and even stopping learning. Yet many statistical ideas have their roots in thinking with which students are already familiar. Knowing this fact, and how to exploit it through the use of analogy gives us a decisive advantage in the battle for hearts and minds of students who do not yet know how much they need statistics in their lives. I will describe analogy as a tool for teaching statistics, my experiences with its use, and many examples of analogies I have invented, borrowed, stolen, lost then rediscovered, and otherwise acquired.

Invited Paper – Rob Gould
Reproduced here are the PowerPoint slides from the presented paper.

Technological Literacy and Statistics Education: A call for thought and research

Dr Robert Gould
Director of the Centre for Teaching Statistics
University of California, Los Angeles

Statistics education should include teaching students statistical technological literacy, which I define to be the ability of students to use and criticize technology in the context of doing statistics. Technological literacy is a very important component of the education of data scientists, particularly because Statistics' unique relationship with technology means that changes in technology affect not only how we practice our profession, but the objects we study. After discussing and illustrating aspects of this relationship, this paper reports on the development of a new journal, Technology Innovations in Statistics Education. The journal was founded with the intent of encouraging more research and discussion into the role that technology plays in statistics education.

RANDOM COMPUTER-BASED EXERCISES ABOUT NORMAL DISTRIBUTIONS

Doug Stirling
Massey University, New Zealand

Abstract
Although project work involving analysis and interpretation of real data is important when students are learning statistics, there is an important role for short exercises to help learn specific statistical skills. Computer-based exercises can be much richer than exercises in paper-based textbooks but existing resources do not make full use of the medium. The format can involve multiple-choice, numerical answers, interaction with diagrams (such as sketching a histogram) or a combination of these, possibly in sequence. The exercise can analyse the student response and give helpful hints and feedback about different types of incorrect answer. Random generation of similar questions in an exercise can allow repeated attempts until skills are mastered. Some principles are given for the design of computer-based exercises and a set of nine exercises about normal distributions is described.

PROJECTS AND EXERCISES

Analysis of real data is usually emphasised in the teaching of statistics. Students should leave an introductory statistics course with the ability and confidence to apply appropriate statistical methods to problems in application areas of interest and should be able to correctly interpret the results of these analyses. Projects that involve collection, analysis and interpretation of data are critically important for the learning of these skills and considerable weight is often given to project work in assessment (Kanji, 1979).

However the practice of statistics involves a toolbox of methods that must be mastered before embarking on relatively open-ended projects, so there is an important role for short exercises to help learn specific skills (Garfield, 1994). Although short exercises cannot fully prepare students for project work, they can effectively cover a subset of the necessary skills. Textbooks (both print- and web-based) usually include short exercises about the topics in each section or chapter, often to the exclusion of more open-ended or wide-ranging questions.

It can be argued that such fine-grained exercises should only be used for formative assessment since the main skills that we should assess in introductory statistics courses involve more open-ended questions. However many statistics courses also base a substantial proportion of their summative assessment on mastery of specific skills in short-answer or multiple-choice tests whose questions are similar to these in many exercises.

COMPUTER-BASED EXERCISES

Traditional exercises in textbooks give little feedback about wrong answers. At best, a brief answer is shown at the back of the book, but often this is only for selected questions and sometimes only a single number is provided as the solution. Students are often left without any indication of where they went wrong if their answer is incorrect.

Many paper-based textbooks have associated CD- or internet-based resources that include short-answer exercises and/or tests. For example, many textbooks published by the Thomson group of publishers are linked to CengageNow resources (Cengage Learning, 2008) with multiple-choice tests for their chapters. However these texts and exercises are usually multiple-choice and, although they provide more feedback about wrong answers than textbooks, the multi-choice format limits the type of question and scope for feedback.

Several introductory statistics textbooks have been written and published for direct use in a web browser in recent years, often without payment. Although dynamic and interactive features can allow electronic textbooks to effectively teach statistical concepts, there are often no exercises (e.g. StatSoft Electronic Statistics Textbook, 2008, StatWeb, 2008, Visual Statistics Studio, 2008) or the questions have the same format and limitations as those in paper-based textbooks (e.g. Hyperstat, 2008, StatPrimer, 2008, and Statistics at Square One, 2008).

Perdisco (2008) provides a collection of exercises whose format is somewhat more flexible, including multiple-choice questions, questions requiring numerical answers and a few questions in which something must be dragged in a diagram.

However existing sets of exercises have various limitations. There are often too few versions of any type of question, preventing weaker students from repeating similar questions until a skill has been mastered. Ideally, each computer-based exercise can be randomly generated over such a wide range of question variations that repetition would only make answering the question easier through mastery of the targeted skill.

The format can potentially be more flexible than existing exercises and might require:

- a multiple choice response
- a numerical value (e.g. a guess at a standard deviation or correlation coefficient from a data display)
- interaction with a diagram (e.g. to ‘sketch’ a histogram or draw a least squares line)
- several responses of these types, perhaps in sequence

Exercises might include interactive diagrams and formula templates (with text-edit boxes in which values can be typed to perform calculations) to help answer the questions.

With a more flexible question-answering format, it should also be possible to analyse wrong answers to provide better feedback about errors. And finally, access to the better existing sets of exercises requires payment, so a free set of exercises would be more easily accessed.

Although we emphasise the use of skill-based exercises as formative assessment, it is briefly mentioned here that if the random generation of questions in an exercise includes a wide enough range of question types, the same exercise could in a computer-based test for summative assessment, even after students have had unlimited practice with it.

PRINCIPLES FOR DESIGN OF EXERCISES

The remainder of this paper discusses the design and implementation of a set of exercises that are accessed using a web browser. The requirements of flexible question format, randomisation of questions and intelligent feedback require implementation with a high-level programming language; Java was used due to its widespread availability in web browsers. The exercises make use of library of Java code that was developed for the CAST set of e-books (Stirling, 2007) and will eventually be publicly released as part of CAST.

The exercises will eventually cover most topics in introductory statistics courses, but this paper only uses a set of exercises about normal distributions to illustrate how implementation of each exercise with a separate computer program provides much more flexibility with the question format and potential for intelligent feedback.

The following principles were used in the design of the exercises.

Working panels

Where possible, exercises should be self-contained and should not require the use of pencil-and-paper or a calculator. Panels containing interactive diagrams and formula templates can be provided to help answer the questions. Checking whether an exercise has been correctly completed may be based on these working panels, but the answer is often entered in separate area of the exercise.

Feedback about wrong answers

If possible, the exercise should identify the mistake in reasoning behind a wrong answer and explain this to the student.

Hints

When a wrong answer is given, the exercise should give hints and allow further attempts. These hints may be supplied as textual messages or displayed in the working panels. An option may be provided to show hints, even before the answer is checked.

'Tell me' button

Each exercise should include a button to show the correct answer. Clicking this button completes any working panels and gives explains the solution textually. (This feedback is an important function of the working panels.)

Randomisation of questions

Exercises should contain a button to randomly generate another version of the question by randomising as many aspects as possible, including the context. Independent generation of successive questions does not work well. For example, consider an exercise that asks questions of the form $P(X<a)$, $P(X>b)$, $P(c<X<d)$ and $P(X<e$ or $X>f)$. Independent randomisation of the question type can present questions of the same type two or three times in a row, and some students would need to try an unreasonable number of questions before seeing all types. A workable solution to select one of n options was to constrain the randomisation by ensuring that there were no repeats in the most recent $(n-1)$ choices and that all options occurred in the most recent $(n+1)$ choices. Separate randomisation of different aspects of a question should ensure that the same question is never repeated exactly.

Sequences of questions

Since the exercises are intended as a learning tool, harder concepts should be associated with a sequence of exercises of increasing difficulty so that students can build their knowledge in small steps.

The following sections illustrate these principles with a set of nine computer-based exercises that have been written about normal distributions.

PARAMETERS AND SHAPE OF NORMAL DISTRIBUTIONS

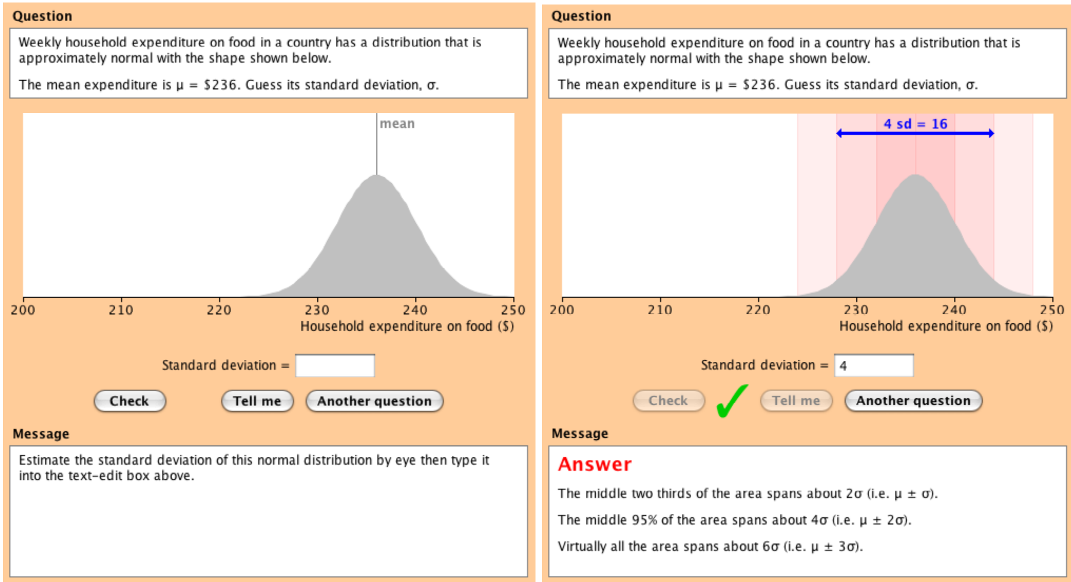
The first two exercises assess understanding of the parameters and shape of a normal distribution. Students should understand how the mean, μ, and standard deviation, σ, are linked to the shape of the distribution's density function and, in particular, that the density becomes close to zero at $\mu \pm 3\sigma$.

In the first exercise, a normal distribution is graphically displayed and the student is asked to guess the value of the standard deviation. A reasonably large error is permitted.

Wrong answer: If the answer typed under the diagram is not close enough, bands are displayed behind the normal curve showing $\mu \pm \sigma$, $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$ and the message states the proportions of the total area that should be in each of these bands.

Tell me: The diagram on the right below shows the result of clicking the 'Tell me' button.

Another question: This button randomly changes the context and wording for the question, the scale on the axis, and the mean and standard deviation of the displayed distribution.

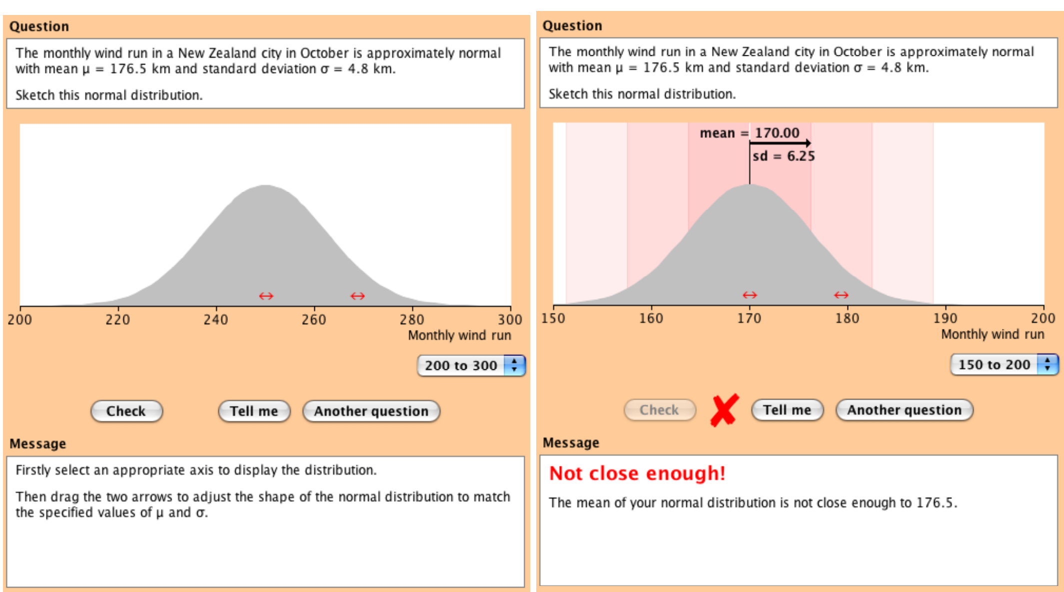


In the second exercise, the numerical values of μ and σ are given and the student must 'sketch' the distribution's density function by (a) selecting an appropriate axis from a pop-up menu, then (b) dragging arrows to change the centre and spread of the displayed normal density function.

Wrong answer: If the sketched distribution is not close enough, the mean and standard deviation of the sketched distribution are displayed and bands show μ ± σ, μ ± 2σ and μ ± 3σ for this distribution, as shown on the right of the following diagram. These 'hints' are dynamically updated if the arrows are dragged to make another attempt at the same question.

Tell me: The correct normal density function is shown with a message saying that the density virtually disappears at μ ± 3σ.

Another question: The context and the target mean and standard deviation are randomised.



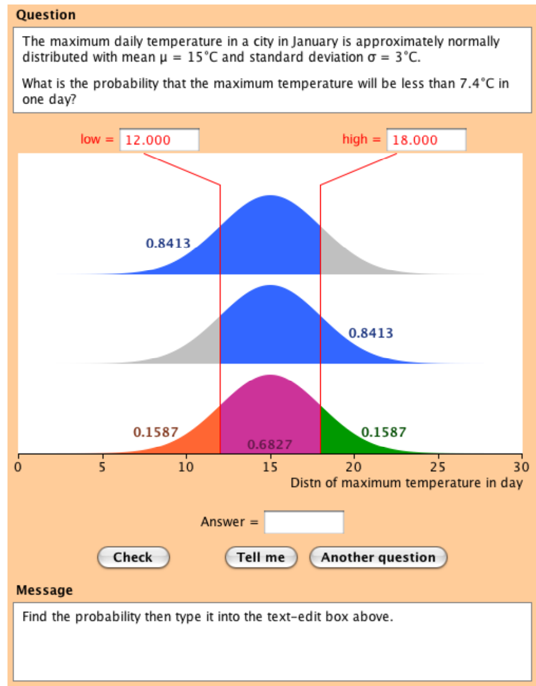
NORMAL PROBABILITIES

Most statistics courses expect students to be able to evaluate probabilities relating to normal random variables. The relationship of probability to area under the normal density function underlies the method, so an initial exercise assesses this. It includes a working panel that displays three copies of the normal distribution, two vertical red lines and the values of five areas relating to these lines (as shown in the diagram below). The two red vertical lines can be dragged or values can be typed in the text-edit boxes to position them exactly.

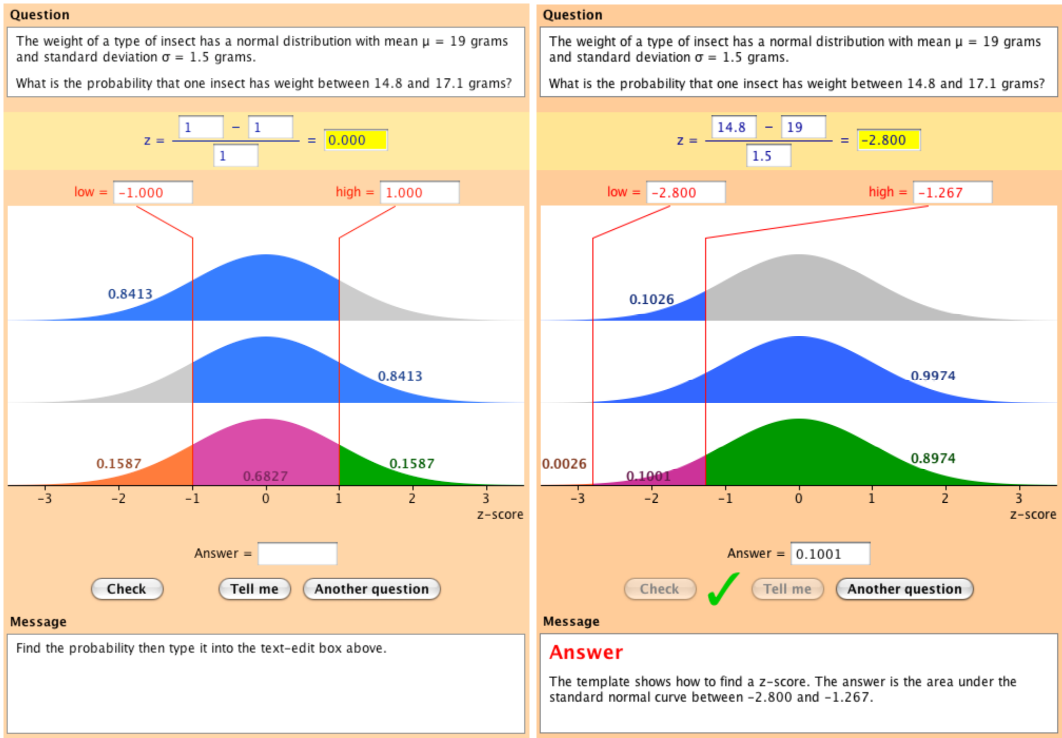
Wrong answer: If the answer typed under the diagram is incorrect or not close enough, the question is reworded in the 'message' box in terms of an area under the normal density.

Tell me: The red lines in the diagram are moved to the end-point(s) of the interval and the relevant area is shown in the answer box.

Another question: The context, mean and standard deviation are changed, and the question is randomised between the forms P(X<a), P(X>b), P(c<X<d) and P(X<e or X>f).



It can be argued that the use of z-scores to find normal probabilities is unnecessary when statistical software and even spreadsheets such as Excel provide functions to directly evaluate cumulative probabilities for any normal distribution. However the concepts of standardisation and z-scores are important in other contexts and their use to find normal probabilities is a good introduction. The next exercise therefore asks the same questions as the previous exercise, but the normal density function is replaced by a standard normal density function and a formula template for calculating z-scores. The following diagram shows a typical question and the result of clicking 'Tell me'.



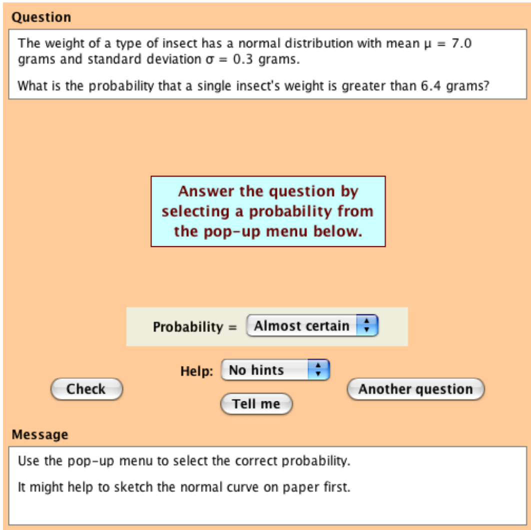
A fifth exercise (not illustrated here) asks the same questions but replaces the standard normal density function with a table of cumulative standard normal probabilities.

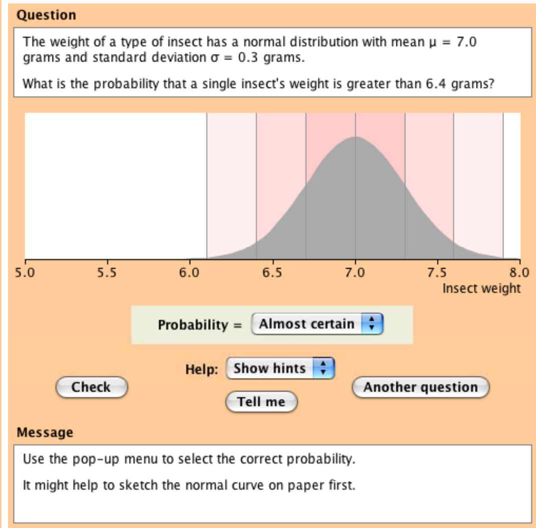
The next exercise requires more understanding and is somewhat less mechanical. It asks for approximate probabilities relating to values σ, 2σ and 3σ from the mean and students should be able to answer questions of this form without detailed calculation. The exercise initially provides no graphical display of the distribution, as shown on the left of the following diagram. To answer the question, a probability must be chosen from a menu containing the values *Almost certain*, 0.975, 0.95, 0.85, 0.7, 0.5, 0.3, 0.15, 0.05, 0.025 and *Almost impossible*.

Hint: There is an option to show hints in this exercise. This displays the normal density function and lines at $\mu\pm\sigma$, $\mu\pm2\sigma$ and $\mu\pm3\sigma$ as shown on the right below.

Tell me: The normal density function is again displayed with the required area highlighted and the answer selected from the menu.

Another question: The context, mean and standard deviation are changed, and the question is randomised into ones of the form $P(X<\mu\pm k\sigma)$, $P(X>\mu\pm k\sigma)$, $P(\mu-k\sigma <X< \mu+k\sigma)$ and $P(X< \mu-k\sigma \text{ or } X> \mu+k\sigma)$ where $k = 0$, 1, 2 or 3.

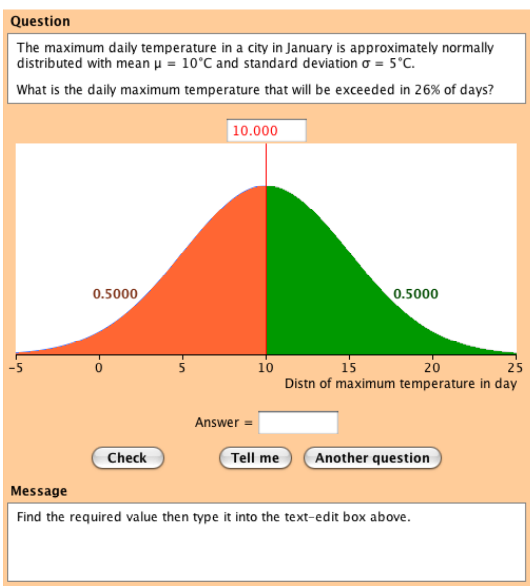


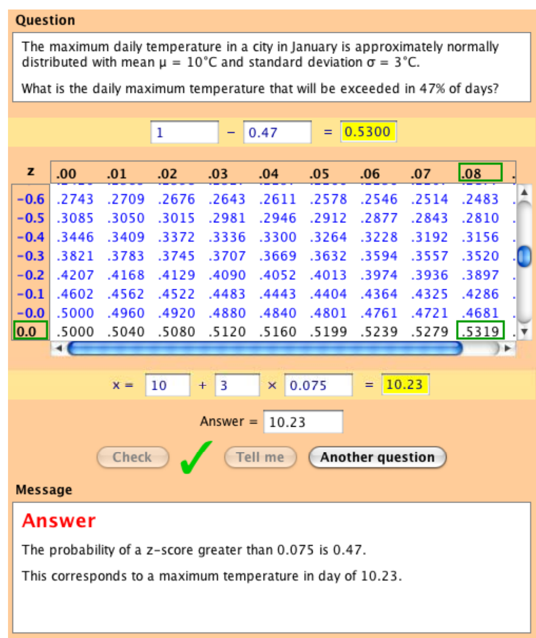


NORMAL QUANTILES

Two exercises relate to the inverse problem of finding quantiles of the normal distribution. The first of these is shown on the left below. The quantile can be determined approximately by dragging the vertical red line on the normal density function or more accurately, with a little trial-and-error, by typing values in the text-edit box above it.

In the exercise on the right, the appropriate quantile is found for the standard normal distribution using the table and translated into the correct units with a formula template. The diagram on the right below shows a typical question after clicking 'Tell me'.





A final exercise that is not shown here asks a mixture of question types including ones requesting probabilities, expected numbers out of n values, and quantiles. This exercise includes several formula templates and a standard normal table as working panels and the student must decide which are relevant to the question.

CONCLUSION

Traditional exercises in textbooks usually provide minimal feedback to weak students. Computer-based exercises have greater potential as a learning resource by giving feedback about wrong answers and explanation of the correct answer. They can also be randomly generated to allow students to make repeated attempts at similar questions until the concepts have been mastered. Well-designed exercises that include interactive features are also more ‘fun’ than paper-based exercises.

It takes a lot of work to design and program effective computer-based exercises. The nine exercises that were used as illustrations in this paper are a very small proportion of the number necessary to cover all topics in an introductory statistics course, but a complete set of exercises of similar style would provide a much better learning resource than the exercises in most textbooks and current web sites.

A full set of exercises will eventually be published as a free e-book within CAST (Stirling, 2007).

REFERENCES

Cengage Learning (2008) CengageNow, http://cvg.ilrn.com/ilrn/

Garfield, J. B. (1994) Beyond testing and grading: Using assessment to improve student learning. *Journal of Statistics Education*, 2, 1.

Hyperstat (2008) Lane, D., http://davidmlane.com/hyperstat/questions/index.html

Kanji, G. K. (1979) The role of projects in statistical education, *The Statistician*, 28, 19-27.

Perdisco (2008) http://www.perdisco.com.au/workbook

StatSoft, Inc. (2008) Electronic Statistics Textbook. Tulsa, OK: http://www.statsoft.com/textbook/stathome.html

StatWeb (2008) Hulme, P, Sherratt, T & Moss, J., http://www.dur.ac.uk/stat.web/

StatPrimer (2008) Gerstman, B. B., http://www.sjsu.edu/faculty/gerstman/StatPrimer

Statistics at Square One (2008) Swinscow, T. D. V,, http://www.bmj.com/statsbk

Stirling, W. D. (2007) Computer-Assisted Statistics Textbooks (CAST) release 3.4, http://cast.massey.ac.nz

Visual Statistics Studio (2008) Cruise Scientific, http://www.visualstatistics.net

JOINING THE DOTS: DIRECTED ONLINE TUTORIALS FOR STATISTICS

WILSON, Richard and BULMER, Michael
Department of Mathematics, The University of Queensland

Abstract
Two issues often arise in the teaching of statistical models and methods to large groups of students in non-statistics programmes. Firstly, such students struggle with the relevance and the applicability of the material to their own discipline. Secondly, these students struggle with the notion of randomness, how to model it and how to account for it in their analyses. One approach to dealing with the second issue is to expose students to as much data as possible, for both single and multiple contexts, by providing individual data sets to each student for tutorial and assignment work. This has been the approach taken in developing DOTS, Directed Online Tutorials for Statistics, for students in engineering and the health sciences at the University of Queensland. The current version of DOTS will be presented and discussed, as well as future directions.

INTRODUCTION

Teaching and learning involves many challenges for both the educator and the student. This is particularly true when the subject matter covers the knowledge, understanding and use of statistical models and methods, and even more so when the students are studying a discipline apart from statistics, such as engineering or health science. Such students did not realise when they enrolled in their programme that they would be required to undertake a course which involved collecting, analysing and interpreting data, and are often resistant to their need to do so, especially if they are later year students (say, in the third year of their programme). They frequently struggle with both statistical concepts and the motivation to learn material they feel is superfluous and not in their area of expertise. Engineering students, after several mathematics courses, struggle with the indeterminate nature of the analyses after seeing only deterministic systems, while health science students struggle with even the notion of quantitative rather than descriptive approaches to learning. Both groups struggle with the requirement that they must solve problems, rather than just listen to the solutions to problems, in order to learn.

A key approach to motivating students to learn statistics is to show that the material to be covered is relevant to their particular discipline through the use of examples and exercises from their discipline. This requires an excellent working knowledge of the use (misuse and abuse) of statistical methods in a variety of discipline areas. In teaching large courses, there may also be a number of different discipline areas represented within the course. At the University of Queensland, all engineering groups, apart from electrical engineering, are taught in a single short course (19 one-hour lectures), which includes mechanical, civil, chemical, mining, metallurgical and environmental engineering. Similarly, all students from the School of Health and Rehabilitation Sciences, studying physiotherapy, occupational therapy and speech pathology, are taught in a single course (39 one-hour lectures). Providing examples and exercises relevant to each discipline within such courses is difficult to do and time consuming to prepare.

One of the features of many statistics courses over many years has been the collection of data by students. This has enabled students to explore data they own and has provided them with insight into the random nature of data and how it should be modelled, analysed and interpreted. Such an approach has been one of the most powerful tools at the statistical educator's disposal. The use of team projects in which students design a statistical study and collect their own data, then analyse their data and prepare a report on their analysis for a major project is an excellent example of this approach. Another option is for each student to analyse their "own" data for a variety of contexts during their tutorial sessions, whether they have collected it themselves or not. However, in time-hungry programmes, the time required to prepare, co-ordinate and assess such activity, especially for large classes, is almost nonexistent.

Before describing and presenting a short demonstration of the current version of DOTS, it is worth commenting on some of the features of the types of courses in which it may be used. Firstly, the objectives of these courses are not set just by the course coordinator, but are established by the client school to which the students belong. Consequently, there is often an emphasis on methodology rather than on understanding concepts, even though the methods may not be used clearly in later courses. The topics covered, both in number and depth, are pre-determined to a certain degree. Frequently, there is an emphasis on the simpler material rather than providing a taste of the more complex modelling and analysis that may be required in real applications. Although it may be difficult to do much on the more complex material, it is desirable to indicate that it exists (and is complex) as this may be the only statistics course the students undertake. The background of the students obviously has a major impact on the approach taken. Health science students have difficulties with modelling, so many courses for such students take an analysis based approach rather than a model based approach, leaving students bewildered as to why they must do some aspects of the analysis. Providing a balanced course in which students understand the modelling of randomness through probability and are enabled to "do statistics" is not easy. An aim of DOTS is to enhance their understanding of the conceptual basis of the methods.

TUTORIALS ONLINE

Most statistics courses have students spending some or all of their tutorial time in a computer laboratory, analysing data and then presenting a report on the analysis. Over the last three to four years, we have been developing online tutorials for students studying either engineering or health science. One of the key objectives of this development has been to provide each student with their own data sets for each tutorial so that they can obtain better intuition into the nature of randomness. As each student has different data, they are encouraged to observe other students' data and results. The data is obtained in a variety of ways depending on the course and on the methodology and concepts being covered by the tutorial. For the health science students, many of the samples are obtained from data collected from the students in the class and previous years' classes in an online survey carried out at the start of each semester. This survey was initially developed for a large first year biological sciences course. It includes questions related to physiology and special interest questions. This database provides data of special interest to these students (as it is about them) and is not too far from their discipline interests. For the engineers, the data are simulated from models fitted to real data as a general rule.

The materials for the tutorials are structured to enable the students to prepare adequately for the analysis of their data and the preparation of their report during the tutorial time. Prior to the tutorial, the students are able to access the current tutorial page. On this page there are links to the learning objectives for the tutorial, instructions (context of the data, statistical package commands and other information depending on the course) and trial data (the same for all students and different to that which they will receive for their report). At present, SPSS is used in the health science course and MINITAB is used in the engineering course. The students are encouraged to prepare for the tutorial by analysing the trial data and answering the tutorial questions for this analysis (without assistance). Each student can then identify aspects which they find difficult prior to the tutorial and then hopefully have these resolved during the tutorial by the tutor.

During the tutorial time, the students have access to their own data, which they copy and paste from a table on a web page into the statistical package and must analyse during the tutorial. They must then compile a report in response to given questions. As each has different data, they are encouraged to look at the properties of their neighbours' data (as far as time allows). Initially, the reports do not form "proper reports", but as the semester progresses, they (should) gradually become better formatted. The available time to carry out the analysis and compile the report is restricted to the tutorial time. The report is prepared through web pages which contain text boxes (with a basic toolbar for minimal formatting given the time constraint) for each part. Additional text boxes are provided for non-graphical output and tables from the statistical package, though these do not form part of the report but act like an appendix for the tutor when marking the report (see below). There is also the opportunity to store graphs from the analysis in the report. The availability of the text boxes and graph storage facilities commence at the start of the tutorial and conclude at the end of the tutorial. Both the content of the text boxes and the files for the graphs are stored in a MySQL database. All the web pages run from php scripts and are password protected.

BEING DIRECTED

One critical component of the tutorials is that they are directed; that is, there is a tutor available during the tutorial. This is essential as students in these types of courses often struggle with their understanding of statistical concepts and methods, and require assistance at critical parts of their learning to clarify unclear aspects or, on occasions, help them unlearn aspects learnt earlier but which are not true. The tutors are available to guide them through the tutorial both as a class and individually. One aspect which can work well is that the tutor can lead class discussions at given times during the tutorial to draw together the differing results students obtain. For example, for the health science students, the mean height of students in the database is known so the tutor can lead a discussion on the nature of confidence intervals by discussing with the class how many obtained a 90% confidence interval which contained the true mean. Similar discussions can be carried out for the results of hypothesis testing on the basis of different significance levels, enhancing the students understanding of the use of p-values to gauge the evidence against a null hypothesis.

Given the current shortage of statisticians, all of the tutors involved are employed on a part time basis. This presents special challenges in running tutorials such as these as the tutor needs to be both knowledgeable and confident enough to carry out class discussions. To assist the tutors in this regard, weekly tutors meetings are held (each 30 minutes), in which the forthcoming tutorial is discussed. These are reasonably well attended though not compulsory. These meetings also assist in addressing the marking criteria for each tutorial and resolving any issues which may arise in the operation of the tutorials or disputes between students and tutors regarding the marking of the reports.

ONLINE MARKING AND FEEDBACK

The tutorial reports for DOTS are marked online. This makes some aspects of marking easier and some aspects more difficult. For instance, paper and pen marking of reports naturally allows the marker to highlight errors and well done work in a reasonably easy fashion. To enable the marker of the online tutorial reports to give feedback on the work done, text boxes (similar to those which the students use to complete their reports) are provided. Typing a response takes longer for most markers than using a pen on paper. To enable the markers to provide informative feedback more easily, a slightly more complex toolbar than that used for the students text boxes is provided. This allows, for example, the marker to copy and paste parts of the student's report into the feedback box, highlight the part which is in error, and briefly comment on the error. As the data for each student are different, the answers obtained by each student are also different. Again, this makes the marking more difficult. However, having the reports marked online resolves this issue as answers for the individual reports can be generated. This enables markers to check quickly if the student's answers are correct. (Of course, such checking could also be done automatically, though this aspect is not implemented yet). As much of a statistical report is based on the output from the statistical package, having the relevant output saved by the student and accessible to the marker in a window separate to the report, allows the marker to check, when necessary, the analysis carried out by the student. Graphs required for the report are placed near the section of the report which refers to it. Providing the student's data also allows the marker to carry out the analysis for the report in rare cases when this may be needed.

It is obviously not possible to provide detailed solutions for each tutorial report for the tutor to use for marking. So, to provide a guide for marking, a marking criteria window indicates where marks should be given or deducted, and includes special aspects to consider for the particular tutorial. In addition, a "model" report based on the trial data (to which the tutors have access and are encouraged to try out in preparation for the tutorial) is provided so that the tutors can see the types of discussion the students should be writing. Additional notes provide an indication of where differences between the results for the trial data and the results for an individual's data may occur. The tutors, after adjusting to a very different system to the usual tutorials and marking, have been reasonably positive about the online tutorials and willing to suggest ways of improving it.

After all students have completed the tutorial and a little time has elapsed for the tutors to at least start marking the reports, a link to a feedback page becomes available. On this page, the student sees their report as the tutor has seen it, with the feedback from the tutor shaded so the student can readily identify it. The answers provided for the tutor are also available.

FUTURE DIRECTIONS

Currently, there is still much which can be done to improve both the organisation of the preparation of the material and the actual materials themselves. On the logistical side, there is a need to streamline some aspects of the preparation of materials to make it easier to implement new tutorials. Also, as the students must complete the analysis and report during the tutorial, attendance is compulsory. Arranging the allocation of students initially to tutorial groups is also time consuming as there must be fewer students in a computer laboratory than computers. If students have legitimate reasons for not being able to attend at their usual time, then an alternative time (and place) need to be found if they are to complete the work. Usually this will be in another tutorial class where possible, though this may not be an option as many classes may be full.

While they are encouraged to do the tutorial using the trial data prior to the tutorial, few students seem to be doing so. This is partly due to time constraints, but it is also due to the difficulty in accessing the statistical package used. One solution to this is to use a free alternative or to encourage the students to obtain a student edition of the package.

Although it is not the intention to remove the presence of a tutor from the tutorials, it is desirable to provide better online help for especially the use of the statistical packages, but also for some basic statistical concepts. An investigation into doing this using a basic help window for each of these is underway.

As indicated above, there are multiple disciplines covered by each course and there is a need to provide multiple contexts for each tutorial to motivate students. For instance, it is often difficult to convince a mining engineer that a mechanical engineering problem is of any relevance. At present, journal articles are being collected from which relevant contexts can be collated and transformed into tutorial settings (with data – this aspect is more difficult to arrange) to provide a multi-disciplinary version of the tutorials. Once these are implemented, students will be able to indicate their discipline area and then be streamed through tutorials covering similar material to students in other disciplines but focussing on a relevant context to their own. A further enhancement would be possible if better links with other courses took place.

One possible way of encouraging these links would be to enable the preparation of reports of statistical analyses of data for other courses to be prepared using the same online structure. This aspect is important for another reason. Many of the students do not have statistical role models within their own discipline as statistical methodology used is often outdated or poorly applied. If such links are forged with courses in their own discipline, then this encourages those in their discipline to utilize the statistical models and analyses currently ignored. DOTS provides a clear framework for enabling those links.

TOWARDS AN UNDERSTANDING OF PROCESSES AND TASKS THAT FOSTER THE DEVELOPMENT AND ASSESSMENT OF STATISTICAL THINKING.

Sharon GUNN and Roslyn STEEL
University of Melbourne, Australia

Abstract
Early attempts to define statistical thinking revolved around discussions of the need for data, understanding the nature of variability and how statisticians go about solving statistical problems. More recently, Wild and Pfannkuch (1999) have proposed a framework in which they identify five types of thinking they perceive as being fundamental to thinking statistically. We understand these types of thinking as a mapping onto the statistical problem solving cycle (real world problem, statistical problem, statistical solution, real world solution) and as providing us with a beginning definition for statistical thinking. Wild and Pfannkuch (2004), in their paper on understanding statistical thinking discussed the historical development of statistical thinking, emerging from this discussion was a broader view of statistical thinking, namely, statistical thinking is a way of making sense of the world, a particular world view.
We believe understanding statistical thinking as a world view may provide additional insights into how we, as educators, can recognise, develop and assess statistical thinking within our students. We explore the links between these constructs and what we observed when we introduced a new teaching and learning strategy into a statistics design and analysis subject.

INTRODUCTION

In this paper we reflect upon our experiences with a second year design and analysis statistics subject. Our focus is on the introduction of online multiple-choice *quiz questions* (quizzes) as an enhancement to the existing teaching model. Our primary objective for introducing this enhancement was to encourage the development of deep learning strategies in our students. In our evaluation of the project we found that students had made changes in their approaches to the subject, the most pronounced changes being in their discussion of their work with other students particularly the nature of these discussions. Tutors reported that student engagement in tutorials was more focused on the discussion and exploration of statistical concepts as well as the interplay between real world contexts and the ‘big’ ideas of statistics. They also commented that they were finding it easier to emphasise the connections between ideas delivered in lectures and material reinforcing these ideas that was covered in tutorials and computer laboratory classes.

These outcomes prompted a reflection on this innovation. Rather than considering tasks as tools embedded in complex learning and teaching systems that are defined in terms of processes, our attention was drawn to considering the nature of the processes that characterise our teaching model and the role they play *in the enculturation* our students into a statistical way of thinking.

We begin by providing background to this innovation, and then we describe the innovation making comments on its effectiveness and follow up by contrasting and comparing a current variation to this model in 2008. We conclude with a statement of our current position on developing statistical ways of thinking in our students.

BACKGROUND

Experimental Design and Statistical Methods is a *service course* subject taught by the Department of Mathematics and Statistics at The University of Melbourne. The subject is compulsory for Agriculture, Food Science, Horticulture, Environmental Science, Forestry and Animal Science Degree students – it represents the only statistics subject that is offered in their courses.

Although the cohort of students is across six degree courses, the students tend to know each other very well through their shared studies, particularly the many compulsory field trips.

Consequently they are very accustomed to discussing and sharing ideas about their work. On a less positive note, there are times when field trips have cut across the teaching time for the subject and students miss vital classes, this lack of continuity and inability to maintain a continual study cycle is somewhat problematic, especially as they find statistics challenging.

The subject is taught using a combination of three lectures per week, a small group tutorial and a computer laboratory session often involving two or three tutorial groups working together. The material covered in the subject is integrated across the various teaching environments. Concepts introduced in the previous week's lectures are reinforced in tutorials by means of tutorial activity sheets and problem sets. The problems from the tutorial class are often re-visited in the computer laboratory session by generating and using Minitab output.

Observations of the teaching staff at end-of-semester evaluations of the course prior to 2004 included:

- Although excellent tutorial questions are provided students do not necessarily attempt questions prior to tutorials;
- there is considerable variability in the academic ability of students;
- tutorial discussions are often very 'fractured';
- student participation in tutorials is less than optimal;

These observations were the motivation to enhance our teaching/learning model in 2004.

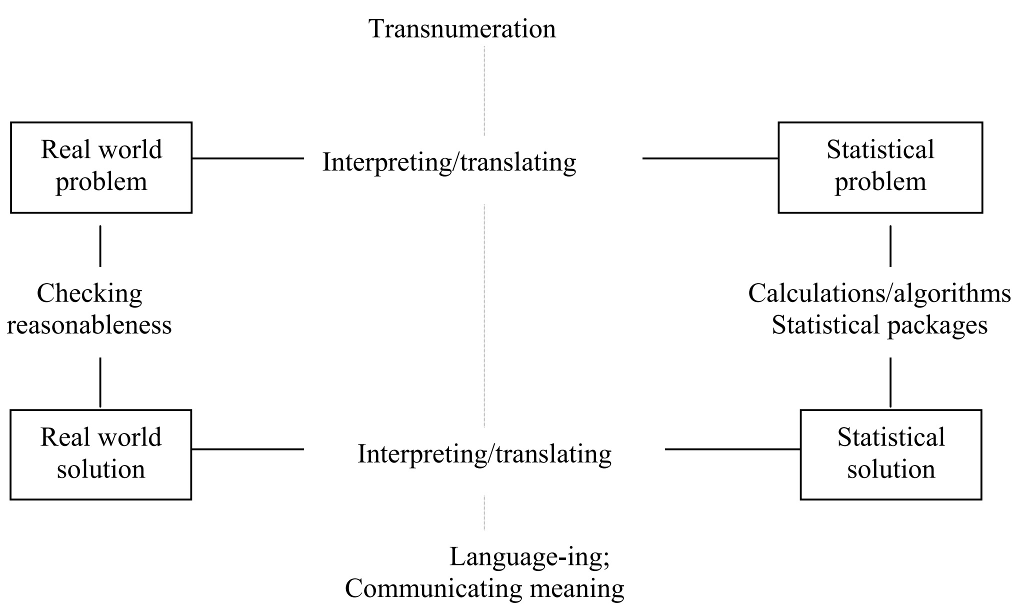
THE TEACHING AND LEARNING MODEL

Problems and Tutorial Activity Sheets

Application problems were often framed within a Tutorial Activity Sheet for each tutorial. They were designed specifically to reinforce the statistical problem solving cycle. This statistical problem solving cycle involves the type of thinking that Wild and Pfannkuch (1999, 2004) refer to in their four dimensional model as transnumeration.

These activity sheets also included relevant Minitab output such as summary statistics/graphical displays/statistical analysis, etc for the problems being discussed as students generally would not have attempted the problems before coming to the tutorial. Students initially were asked to identify the questions of interest to the researcher, the response and explanatory variables and to conduct a preliminary analysis *of the data* using the information provided. They then modelled the data and answered the rest of the question as asked in full.

THE STATISTICAL PROBLEM SOLVING CYCLE



Incorporating the quizzes into the existing model

The enhancement we introduced involved the introduction of weekly online multiple-choice quiz questions, usually two or three related questions The questions were short, did not require extensive calculations, explored one main statistical concept and had been authored with due care to provide meaningful and informative distracters. In addition they were designed to complement and consolidate the work of the tutorial and laboratory sessions of the week concerned. Students had the opportunity of obtaining 5% of their assessment in the subject through their participation in the quizzes.

The quizzes remained open for student access for a week. At the end of the week a summary of responses were e-mailed to tutors so that they could use this information to diagnose student difficulties and hence assist their planning for the following week's tutorial classes. Answers were not provided online; students had to attend the tutorial class and participate in the discussion in order to know whether they 'got the right answer'. Tutors were instructed to facilitate discussion but not provide answers directly, they were to encourage students to justify their answers and explain their solution process to the group. In practice animated discussions of the quizzes often became the focal point of the tutorial class.

There seems to be something about puzzles/quizzes that really engages human beings!

EXAMPLE 1: CONFIDENCE INTERVAL (CI) AND LEAST SIGNIFICANT DIFFERENCE (LSD**)**

The Context

A celery grower wanted to compare two new fertilizers (F1 and F2) with a natural product (fowl manure), in their effect on growth of celeries. Fifteen plots were set aside to do this. The fertilizers were randomly allocated to plots, and at harvest the average weight (in kg) of celeries in each plot was recorded. The weights were as follows:

Fertiliser	F1	F2	fowl
	1.56	1.44	1.53
	1.51	1.50	1.48
	1.52	1.46	1.57
	1.49	1.41	1.47
	1.60	1.48	1.58

Students commenced this problem by completing the focusing questions on the Tutorial Activity Sheet, thereby entering the *Statistical Problem Solving Cycle*. (Graphs and relevant summary statistics were provided on the sheet.)

The questions

(a) Do the following without using Minitab:

i) Formulate a model for the data.

ii) Estimate the parameters in the model except sigma .

iii) Construct an ANOVA table.

iv) Compare the F statistic with the appropriate tabulated value, and use this to compare your model with an overall mean model, hence test the significance of the fertilizer effect.

(b) Enter the data into Minitab, and run an ANOVA analysis to check that your calculated ANOVA table is correct.

i) Using the ANOVA table find the standard error of the difference (s.e.d.) between the means for fertilizers F1 and F2.

ii) Using this s.e.d. and the t distribution table, find a 95% confidence interval for the true mean difference between fertilizers F1 and F2.

iii) Check your 95% CI using Minitab. Use your confidence interval to test the hypothesis that the mean yields for fertilizers F1 and F2 are equal.

Related quiz questions

Q1. You have taken a random sample of 25 cows on a farm and measured the percentage of fat (%fat) in the milk from each cow. A 95% confidence interval for the mean %fat for all cows on the farm (i.e. the population mean) is (3.6, 4.2). Which of the following statements gives a valid interpretation of this interval?

A. 95% of the sample of cows has milk with %fat between 3.6 and 4.2.
B. 95% of the population of cows has milk with %fat between 3.6 and 4.2.
C. If the procedure were repeated many times, 95% of the resulting confidence intervals would contain the population mean.
D. If the procedure were repeated many times, 95% of the %fat readings would be between 3.6 and 4.2.
E. If the procedure were repeated many times, 95% of the sample means would be between 3.6 and 4.2.

This quiz question uses *language-ing* as a tool for teasing out understandings of a CI. Students spent a great deal of time talking about what each statement actually meant and the *conceptual* differences between each statement.

Q2. An experiment was conducted to examine the yield of 4 different varieties of barley. An ANOVA was performed on the data, and each pair of means compared using the least significant difference (LSD) at the 0.05 level. The variety means are as follows, with different letters signifying which means are significantly different from each other.

Variety	1	2	3	4
Mean	4.6*a*	4.9*ab*	5.3*b*	5.9*c*

We can conclude that the LSD is:

A. less than 0.4
B. greater than 0.6
C. between 0.3 and 0.4
D. between 0.3 and 0.6
E. between 0.4 and 0.6

In this quiz question students were being asked to approach the concept of LSD's from a very different angle, solving the puzzle required them to work *backwards* from the magnitude of the differences between the means to determine the size of the LSD. This approach was contrary to the *forward* process necessary in the tutorial question. Students were quite uncomfortable with this approach, it seemed to 'push them out of their comfort zone' refusing to allow them to rest with a simple procedural understanding, and thereby encouraging them to develop a more complex conceptual understanding of LSD's.

A further example of approaching problems *from very different directions* can be seen in the questions used to explore the concept of interaction. Again the gains for the students seemed to be an enhanced understanding of a statistical concept.

EXAMPLE 2: INTERACTION BETWEEN TWO CATEGORICAL VARIABLES

The Context

A farmer was interested in the weight gain (kg) of pigs and the administration of Vitamin B12 and Antibiotics as food supplements. All pigs were of similar ages but from different litters. The four combinations of the two factors were each used on three pigs over a four week period.

The weight gains were as follows:

	Antibiotics	
B12	No	Yes
No	1.30	1.05
	1.19	1.00
	1.08	1.05
Yes	1.26	1.52
	1.21	1.56
	1.19	1.55

Again, the problem was framed within the Tutorial Activity Sheet with relevant Minitab output of ANOVA analysis and asked the usual initial questions about the problem.

The questions

(a) Produce a cross table of means and an interaction plot. Use the display of the data to make some tentative conclusions about the interaction.
(b) Perform an ANOVA in Minitab and discuss your conclusions in (a).
(c) Calculate the LSD for assessing the effect of Antibiotics in the presence of Vitamin B12 and find a 95% confidence interval for this effect.
(d) Calculate the LSD for assessing the effect of Vitamin B12 in the presence of Antibiotics.
(e) Give a suitable display to show which of the four means are significantly different from each other.
(f) Calculate the predicted value and the residual for the first observation (1.30 kg).
(g) Suppose that the treatments had been described as a single factor at 4 levels, corresponding to the four possibilities. Construct the ANOVA table for this using the ANOVA table from (b).
(h) Create a single factor in Minitab and run an ANOVA to check your ANOVA table in (g).

Note

These questions are enquiry driven from the data that was given and conceptual understandings were encouraged from the context. Part (g) opens the students to explore the notion that there is more than one way to model the data (Pfannkuch and Wild, 2004).

Associated quiz question

Each of the 2 *X* 2 tables below shows the means from an experiment involving two factors, A and B, each with two levels 1 and 2. Which one of the cross tables indicates the greatest interaction between the two factors?

A.

		B 1	B 2
A	1	14	10
	2	12	2

B.

		B 1	B 2
A	1	10	14
	2	16	20

C.

		B 1	B 2
A	1	14	14
	2	14	10

D.

		B 1	B 2
A	1	10	14
	2	20	18

E.

		B 1	B 2
A	1	10	14
	2	12	24

Comments on student approaches to these questions

- When approaching the quiz question students said 'they had the numbers but they needed to attach a meaning to the numbers before they could begin to solve the puzzle'. They invented various contexts and then proceeded to analyse each table.
- Students never actually came up with the model estimates for the interaction terms in the quiz question but they did devise a measure of 'degrees of parallelism'. Understanding of interaction as 'degrees of parallelism' was a challenge for many.
- Students acknowledged that interaction, like variation, exists in all *real world* contexts, they were asking such questions as 'was the interaction observed in the sample due to sampling variation or systematic variation?'
- Although students already had a good contextual understanding of interaction they experienced some difficulty in finding the *right* words to describe it and mapping their contextual understandings onto the graphical displays of the interaction plots. (*Reading the story* from the summary measures and graphical displays was encouraged across all our activities through discussions surrounding the tutorial activity sheets.)

This was a long and interesting discussion week for us all!

EVALUATION OF THE 2004 MODEL

To evaluate the introduction of the quizzes we administered online questionnaires both at the start and end of the course - all responses were anonymous. We also maintained a journal documenting ongoing informal discussions with tutors.

Tutors did not think the 'carrot' of five percent assessment was the main motivation for the student participation in the online quizzes because the conversations in the tutorials and laboratory sessions were so rich and students engaged so readily and enthusiastically with the quizzes. Tutors reported increased attendance at, and greatly enhanced discussion of ideas in the tutorial sessions. They also commented that the diagnostic reports from the quizzes were helpful to them when planning tutorials, and that the quiz questions provided an excellent opportunity for the consolidation of *ideas as discussion* of the statistical ideas associated with the quiz questions *often* continued until a consensus was reached. Tutors also comment that being able to *listen* to students justify their answers and explain their understandings made it easier to identify student misconceptions, monitor use of statistical language and assess whether or not students had actually understood the statistical ideas involved.

Student responses of the project were also very positive. Student participation rates were reasonable with sixty five percent of students attempting six or more of the ten weekly quizzes, the average participation rate ranged from fifty five to seventy five percent.

Comments made by students on the post questionnaire indicated that students also considered the quizzes valuable in terms of helping them to develop understandings. No measures were available to 'quantify' enhanced learning however the qualitative analysis of the questionnaire responses indicated that the primary goal of encouraging students to maintain meaningful communication with the subject had been achieved with students often emphasizing how important it was to 'stay in touch with the material'. Students also commented on the language of statistics making comments such as:

> *"To recognise that even though it uses the same words as English, stats is another language and to compile a glossary of terms, symbols and stuff right from the beginning"*

As with all innovations in teaching one needs to be cautious about assigning cause to the innovation, however our exploration of our experiences with this enhanced model appears to bring the hope that new understandings about the teaching and learning of statistics may emerge.

REFLECTIONS ON PROCESSES THAT DEFINE OUR TEACHING MODEL

With the initial introduction *in 2004* of the online multiple choice quizzes we noticed that new *conversation spaces (*where students could discuss and further explore their understandings of statistical concepts) appeared. The process we used enabled us to engage the students in an ongoing statistical conversation that lasted for the entire semester. It also provided tutors with the opportunity to *hear* students' statistical thinking and facilitate understandings in a timely manner. These conversation spaces emerged from a teaching and learning model that was philosophically and pedagogically integrated across the teaching and learning environments of lectures, tutorials, and laboratory sessions.

In the preceding examples we draw attention to the *types of tasks* that were used in the course and how they meshed together to encourage deeper understandings. We also note the statistical ways of thinking that were being encouraged. In identifying these statistical ways of thinking we have drawn heavily on Pfannkuch and Wild's (2004) exploration of understandings of statistical thinking, their earlier four dimensional framework (Wild and Pfannkuch, 1999), and Chance's (2002) discussion on the components of *statistical thinking*.

We are cautious with our use of the term *statistical thinking* primarily because our understanding of the themes that are emerging from the literature relates more to what is referred to in cultural studies as a 'worldview'. For us, these themes reflect particular values and views on what constitutes legitimate knowledge, types of knowledge and how we come to know, preferred ways of communicating and language-ing, and in particular ways of thinking, perceiving and making sense of the world – that is they reflect a particular *culture of statistics*.

CONCLUSION

The interplay between quizzes and tutorial exercises encouraged students to continually map between contexts and conceptual understandings – an important aspect of statistical ways of thinking (Pfannkuch and Wilde, 2004). Our process enabled a meshing of the various teaching materials, and we believe that the meshing plays a major role in the type of learning and thinking that is developed.

Supporting this view were our experiences with the quizzes in later years and a further variation in 2008 when the quizzes were incorporated into weekly homework exercises. These exercises consisted of two quiz questions and an analysis question that required students to use the problem solving cycle. Students commented that they found the quizzes very demanding, conversations took place at the university cafeteria and on the mobile phone but no longer in the classroom! The nature of the questions allowed for diagnostic testing at this point but it was almost too late for meaningful discussions in tutorials.

From cultural studies we know that conversation and language-ing are essential for enculturation to take place. We suggest that by providing an opportunity for students to develop their understanding of statistics through language-ing a *space* is provided for enculturation to take place. The nature of the particular conversations that emerged in our 2004 model encourages enculturation into *statistical* ways of thinking.

We would like to emphasize that, in our experience; it was not so much the tool that was critical but the process in which the tool was embedded.

REFERENCES

Chance, B.L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistical Education,* 10:3

Pfannkuch, M and Wild, C. J. (2004). Towards an understanding of statistical thinking. In *The Challenge of Developing Statistical Literacy, Reasoning and Thinking.* D. Ben-Zvi and J. Garfield (editors). Kluwer Academic Press: Netherlands.

Wild, C. J., and Pfannkuch, M. (1999). Statistical thinking in empirical enquiry, *International Statistical Review,* 67:3, 233-265.

ACKNOWLEDGEMENT

We would like to acknowledge Dr Graham Hepworth, the lecturer for this subject, for making it possible for us to run this innovation, for his authoring of most of the materials and his continued support.

Contributed Paper – Alice Richardson and Felecia Zhang

LANGUAGE SUPPORT FOR STATISTICS LEARNERS

Alice Richardson, Faculty of Information Sciences and Engineering, University of Canberra
Felicia Zhang, Faculty of Communication and International Studies, University of Canberra

Abstract
In this paper we report on the results of an experiment conducted in the unit Introduction to Statistics *at the University of Canberra. A variety of strategies, referred to as games, sets and matches, were employed, many of which emanate from the teaching of foreign languages. These included personal strategies for the lecturer such as recording lectures, individual strategies for the students such as the use of* Hot Potatoes *software, and group strategies for the students such as vocabulary cards and in-class activities and discussions. The aim of the experiment was firstly to improve students' use of statistical language, and secondly to see if we were indeed still teaching statistics, and improving overall student performance in the unit..*

LANGUAGE ISSUES IN INTRODUCTORY STATISTICS

Who hasn't heard students complain that learning Statistics is like learning a foreign language – there are so many words to learn and learn to use properly. Overcoming the language barrier in the learning of science subjects has recently been addressed in molecular biology and genetics (Zhang and Lidbury, 2006). Many foreign words need to be assimilated, along with specialised meanings for ordinary English words.

Specialised meanings for English words are also common in Statistics: these could be referred to as jargon. Firstly, ordinary English words such as 'normal' have a restricted meaning. A student who had not acquired the statistical meaning wrote "The whiskers of the boxplot are normally spaced" to describe a boxplot with whiskers of equal length.

A second language problem that arises in Statistics is the 'scattergun' approach to answering questions. In other words students often seem to throw two or three statistical terms into a sentence (without understanding them) and hope for the best. For example, a student was asked to select either the mean or median as a more appropriate measure of location, based on a boxplot. The student wrote, "I would like to describe the location with the median, because there are not too many outliners and though median appears to sit lower on the scale even in the interquartile range." Notice that in this answer a range of statistical terms was used, such as the 'median', 'scale', and 'interquartile'. But as an answer to describe location with the median, from 'median' on, the sentence is basically nonsensical. The non-statistical term 'outliner' was also used instead of 'outlier'.

This and other interesting uses of Statistical language by students could well arise from the teaching model employed. Many introductory statistical courses are taught using a traditional transmissive model. This basically means that the lectures are usually instructor centred with the lecturer talking almost 100% of the time to a class of students. Students do not usually work in cooperative groups during the lecture. In a week of lectures (3 hours with 1 hour of tutorial/lab) covering regression, for example, 20-30 new terms would have been covered, some text-based and some graph-based. However, in the traditional model of delivery, there is no time for students to reflect on the learning nor is there time for the lecturer to stop to check students' understanding.

As a result, while many students do well in introductory statistics courses, it is not clear that they retain the information for very long or that they are able to make use of it in their studies. As an example of this, an experiment was conducted in semester 1, 2008, in *Forensic Statistics*, a 2nd year unit taught at the University of Canberra with prerequisite consisting of any introductory statistics unit. Students were asked at the beginning of week 1 to supply definitions for the following terms: population, random, sample, observation, value, mean, median, standard deviation. A data set of 5 observations (10, 10, 25, 30, 5) was provided for students to use for illustration if required. Twenty students responded. For example, eleven students gave completely correct definitions of "standard deviation" (s.d.). Two of the eleven reported a correct value for sample s.d. and one reported the population s.d. It should be noted that students were not required to calculate the s.d. On the other hand there is some evidence, particularly given the student who calculated the population s.d., that students can provide correct definitions but are unable to support them with correct calculations.

CHANGING HOW STATISTICS IS TAUGHT

Over 20 years ago, Arthur and Gamson (1987) stated that good practice in undergraduate education included developing reciprocity and cooperation among students, using active learning techniques, giving prompt feedback and respecting diverse talents and ways of learning. In our experiment to investigate the effect of language teaching strategies on Statistics teaching to a small cohort (less than 50 students) in the unit *Introduction to Statistics*, we have implemented strategies for each of these aspects.

In order to apply active learning techniques it is necessary to identify teachable moments. Teachable moments (Rea, 2003) arise when students' questions lead to discussion of a topic without much, if any, lecturer preparation. By preparing the lecture alongside an intelligent novice, likely teachable moments can be identified in advance and worked into group activities. In this experiment, the Statistics lecturer worked alongside an educationalist (an "intelligent novice" at statistics), by meeting an hour before each lecture to convert transmissive-style classes to learner-centred ones. During this hour, the content of the lecture was taught to the educationalist and she in turn asked questions which were used as "teachable moments" in the lecture.

Knowing when to give prompt feedback requires listening to students intently. However, to encourage students to speak up about what they do not understand requires the developing of reciprocity and cooperation among students. So in order to develop a more cooperative learning atmosphere in class, students are divided into 3-4 person groups in each lecture. During the lecture student groups are involved in activities at appropriate points during the lecture. The appropriate points are either pre-determined before the lecture or arise from questions from the students. This frees the lecturer from being the sole source of information within a lecture, and enables the lecturer to check students' understanding. In terms of 'how' to give feedback, once students have attempted to solve the problem themselves in groups, the groups report back to class and, if required, get further information from the lecturer. Most of the time, discussions led by students can solve most of the problems thus creating a sense of achievement in students.

PERSONAL STRATEGIES

Using active learning techniques such as seeking feedback and conducting group work in lectures require changes in the personal teaching strategies of lecturers. For instance, it may be necessary for lecturers to consciously slow down their speed of delivery and remember to elicit feedback either through asking a question or doing a group activity.

In language teaching, new language is constantly modelled (shown in correct usage) and checked via different sensory modalities e.g. listening, speaking, writing. This experiment also involves importing these techniques into the teaching of Statistics. For instance, instead of being shown how to work out a problem only once during the lecture, students are now shown how to work out a problem once and then given the opportunity to work on examples themselves to demonstrate that they understand the reasoning modelled by the lecturer.

The lecturer needs to listen intently to students' needs. Bulmer et al. (2007) have worked in retrospect to discover what it is that students find hard as opposed what lecturers think students find hard. It is also possible to respond to students' needs dynamically. In this case the lecturer needs to be ready to change a lecture in order to cater for student needs. For instance, a problem involving the Normal distribution (adapted from Moore (2007)) might start out like this. "Consider babies born in the 'normal' range of 37 – 43 weeks gestational age. Extensive data supports the assumption that for such babies born in Western countries, birth weight is Normally distributed with mean 3432 g and s.d. 482 g. What is the probability that the birth weight of a randomly selected baby of this type is less than 2000 g?"

A typical solution is as follows: "$P(W < 2000) = P(z < (2000 – 3432)/482) = P(z < -2.97) = 0.0015$." But solutions of this type ignore the fact the first thing required of the student is to decide which number in the problem is μ, which number is σ and which tables or software command is going to generate the required probability (denoted by P). Sometimes lecturers forget that how to start thinking about a problem is usually more difficult than applying a formula to solve the problem.

GROUP STRATEGIES

Group activities provide students with an opportunity to get involved and try out ideas and techniques. Two examples are discussed in this section.

Firstly, consider a lecture introducing the topic of simple linear regression. The lecture begins with discussion of a motivation question. "When a scatterplot shows a linear relationship, we would like to use one variable to predict values of the other. Consider a study of 21 children that finds that children who start to speak when very young are more likely to have high Gesell test scores than children who start to speak at an average age, while children who start to speak the oldest have the lowest Gesell test scores. Which variable is used to predict? Which variable is being predicted?"

The lecturer then discusses the logic behind the least squares regression line, estimation of the regression line using SPSS, and various facts about least squares regression. At the end of the lecture, students work in groups to attach the terms "X axis", "Y axis", "intercept" and "slope", "a" and "b" to a scatterplot with superimposed regression line. For instance, during this lecture there was an interesting debate about the word 'intercept'. The debate was around whether it signifies the point at which the regression line intercepts with the Y axis or whether it means the distance from zero to the point of the line intercepts with the Y-axis. Such a question from the students showed that they were tuned in to language used in Statistics through the word 'intercept'. An activity like this raises students' awareness of the visual aspect of regression, increases their confidence in the definition of terms, and provides a connection to SPSS terminology (where, for instance, "constant" is used in the output to refer to the intercept).

The second hour of the lecture is spent on collecting data on height and armspan from students in the class, and using SPSS to analyse the data dynamically.

Secondly, as part of this experiment, every two weeks a set of vocabulary flash cards grouped around one or two lecture topics was prepared. For example, under the topic of regression the cards included the phrases "direction", "$\hat{y}$", "slope", "strength", "intercept", "scatter plot", "association", X axis" and so on. Groups of up to 4 students arranged the cards on the desk in a manner meaningful to them, in the manner of the concept maps of Novak and Gowin (1984) but with the added flexibility of being able to move cards around without using up paper. In a large tutorial, after 15 minutes, groups can compare the arrangements and discuss similarities and differences. This activity enables students to create relationships between the terms themselves thus leaving them with a feeling of success, and reducing anxiety about studying statistics. It also encourages students to talk about statistics to each other, invoking the mode of speech to assist in Statistics learning. Furthermore, dialogue between students frees the tutor from being the sole source of information in a tutorial. The tutor now can walk around the class offering guidance if needed. Furthermore, debate amongst students can also highlight misconceptions such as a confusion between $\hat{y}$ and y.

ONLINE STRATEGIES

In order to respect diverse talents and ways of learning, in particular for students who were working online, three strategies were implemented.

Firstly, lectures were recorded through a microphone connected to a laptop sitting on the table in the lecture room.

Secondly, lecture overheads, complete with notes made during a lecture, were scanned into pdf format and uploaded into WebCT after each lecture. Not all teachable moments can be captured before a lecture, but it is important to respect the students' need to close the loop on questions raised in class. A written record of loops requiring closure appears in these slides.

Thirdly, online quizzes using *Hot Potatoes 6* (http://hotpot.uvic.ca/) software for online learning were administered. This software was created by the Research and Development team at the University of Victoria Humanities Computing and Media Centre in Canada and is free for educational institutions to install. We chose this software because of it is freely available and studies of its application to the sciences have already been undertaken at UC (Zhang and Lidbury 2006; Richardson, Lidbury, and Zhang 2007).

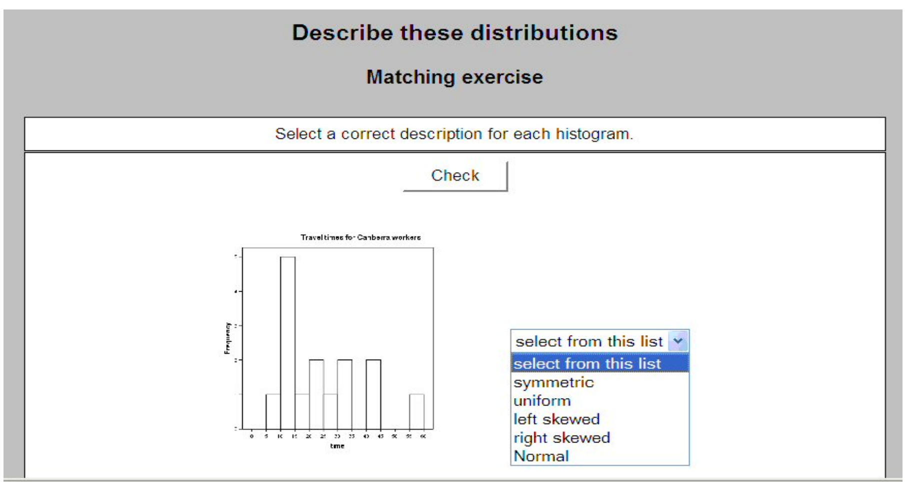
Example exercises from *Hot Potatoes 6* are shown below. In the examples, we have taken data from Moore (2007, p. 37) and adjusted it for local conditions. Students are provided with the sample of 15 travel times to work, in minutes, for a sample of 15 Canberra workers.

30 20 10 40 25 20 10 15 60 40 5 30 12 10 10

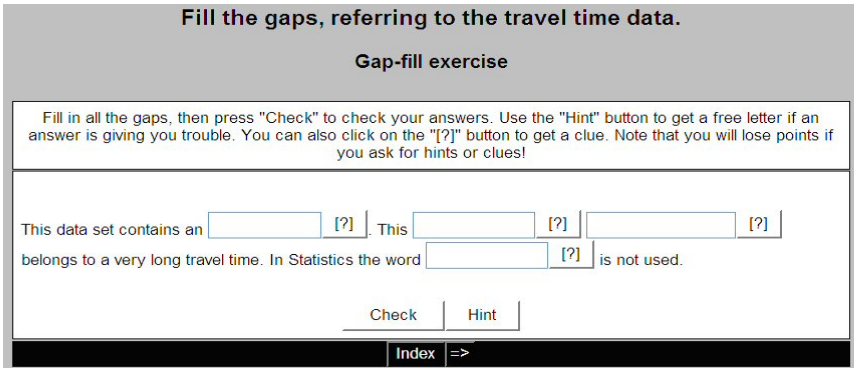
The real-world question posed initially is “Does the data support the claim in the *Canberra Times* last week that Canberra workers spent an average of 10 minutes travelling to work?”

Exercise types in Hot Potatoes include matching, cloze, multiple choice and jumbled sentence. Two examples are shown below. They need not be regarded as stand-alone exercises as the different types of exercise test different aspects of understanding and act to reinforce each other.

Matching: A *Hot Potatoes* exercise matching concepts in hypothesis testing to their definitions is shown at http://www.mathsnet.net/asa2/modules/s24hyptest.html. Our example below is designed to assist students with one of the first steps in answering the question above, namely matching terms for describing histograms to the histogram shapes.



Cloze: Highly specialised statistics software utilises the cloze or gap-fill approach, hence our use of it provides an additional educational benefit in preparing students for future study. In particular, the *Bioconductor* software package has a report-writing package *affyQCReport* (Parman and Halling, 2007) which generates a report on a data analysis by filling in the gaps in a standard report structure. *Hot Potatoes* make that sort of technology available to students of introductory statistics as well as providing a tool for statistics students to learn the language of statistics. Here we show a cloze exercise commenting on an unusual aspect of the distribution of the travel times.



The [?] buttons define provide [hints] for the missing words as follows.

- This is a noun which describes a value which is a long way from the bulk of the values.
- This adjective goes before the noun and refers to a value which is a long way from the bulk of the observations.
- This noun refers to a number in the data set (observation / value).
- This noun refers to a tool for drawing around the edge of a picture.

The meta-language used here such as ‘nouns’, ‘adjectives’, ‘verbs’ and so on are a way of establishing a common language that students can use in future discussions on the nature of words used. For instance, without such language, it is not possible to explain to someone the difference between ‘outlier’, ‘outlying’ or ‘outliner’. An answer that simply states that using ‘outliner’ is wrong is not sufficient to equip students with tools for self-monitoring of their own progress in future studies.

RESULTS

A variety of data pertaining to the success of this experiment is in the process of being collected, including an attitude survey, a language test and counts of website hits along with test and final exam scores.

The Survey of Attitudes Towards Statistics (Shau et al, 1995) has been administered to students at the beginning and end of semester 2, 2007. Students are asked to select from a 5 point scale *(1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree)* the number that best represents your point of view. The median scores from the start and end of semester are shown in the table below. Nine students responded at the beginning of semester and six students responded at the end of semester. The voluntary nature of the survey and the small sample sizes prevented any matching of data tacking place.

	statement	Median (start)	Median (end)
1 √	*I think I will enjoy/have enjoyed taking a statistics unit.*	3	4
2	Statistical skills will make me more employable.	4	4
3	*Because it is easy to lie with statistics, I don’t trust them at all.*	2.5	3
4	Understanding probability and statistics is becoming increasingly important in our society, and may become as essential as being able to add and subtract.	4	4
5	Statistics is not particularly useful to the typical professional.	2	2

6 √	*You need to be good at mathematics to understand basic statistical concepts.*	3	2.5
7 √	*To be an intelligent consumer, it is necessary to know something about statistics.*	3	4
8	Statements about probability (such as what are the odds of winning a lottery) seem very clear to me.	4	4
9 √	*I can understand almost all of the statistical terms that I encounter in newspapers or on television.*	3	4
10	*I could easily explain how an opinion poll works.*	4	3
11	Given the chance, I would like to learn more about probability and statistics.	4	4
12 √	*I often use statistical information in forming my opinions or making decisions.*	4	4.5
13	*I feel insecure when I have to do statistics problems.*	3	3.5
14 √	*Statistics should be a required part of my professional training.*	3	4
15	When buying a new car, asking a few friends about problems they have had with their car is better than consulting an owner satisfaction survey in *Choice*.	4	4

Nine of the fifteen statements (in italics) showed movement in the median. Of these, six (marked with a tick, √) were in a "positive" direction i.e. further towards the response that would be expected from students who have studied introductory statistics. These results suggest that the unit as a whole has positively influenced students' attitudes towards statistics. The movement towards the left (i.e. in the negative direction) in statement 6 is interesting as it suggests that students are starting to understand that doing Statistics is not just about applying formulae to numbers. Perhaps they realise that getting the process or thinking right is more important.

The language test described in *Forensic Statistics* earlier was also administered in semester 2, 2007 to *Introduction to Statistics* students at the beginning and end of semester. One student at the beginning of semester defined Population as "the total amount of something" and Random as "an element chosen which is not in any particular sequence", but could not come up with definitions for any of the other terms. By the end of semester, the student defined Population as "all university lecturers who drove here last Friday" and Random as "the 5 samples above were chosen at random" (referring to the 5 University lecturers in the example given, but with an inappropriate use of the term sample). Another student defined Standard deviation at the beginning of the semester as "average difference of a value to the mean" and at the end as "average distance from the mean = measure of spread in symmetric data" (adding a little extra about the shape of distributions). This small selection of results shows that outcomes will be uneven, but that language acquisition can successfully be demonstrated by the end of a semester.

Our study also confirms the work of Bulmer, O'Brien and Price (2007). Their survey of Statistics students found that the approaches to learning that were highly valued by students were a focus on the basics, taking small steps and providing plenty of opportunities for practice.

FUTURE DIRECTIONS

A drawback of Hot Potatoes is that while it can display images such as graphs in a straightforward way, it does not handle matrices or statistical formulae or even subscripts very easily. We would like to investigate how to overcome these difficulties: some work has already been done in this direction in sciences such as Chemistry (see http://www.wpbschoolhouse.btinternet.co.uk/).

The interaction between lecturer and student, and between students themselves, is critical to the implementation of the strategies discussed in this paper. In the future we would like to consider how to implement these techniques in larger classes (for example, Business Statistics at the University of Canberra attracts upwards of 250 students each semester). Some of the possible group work configurations in large lecture type classes are described in detail in Ebert-May, Brewer and Allred (1997). Another possibility would be to use clickers. Anecdotal evidence suggests that some institutions have been wary of introducing clickers because of the cost involved in purchase, and possible problems in the management of issuing and collecting the clickers themselves. Other institutions have circumvented this problem by issuing the clickers in the same way as a library book.

One outstanding observation from lectures that incorporated group work in this experiment is the amount of energy that is generated by students through discussions. We would like to be able to capture this energy and transfer it to the students who do not attend the class. However, how this can be done is still not clear. For example, is it possible to upload the teachable moments (i.e. questions from students and staff during the lectures) to the unit website and use them as stimulus questions for discussions? If so, how should the online discussion be managed in order to make sure that students do benefit from the discussions? An ability to transfer energy online will be extremely useful in other Statistics units, as the University of Canberra offers *Forensic Statistics* at 2nd year and post-graduate level in a fully online version only.

REFERENCES

Allard, D.W and Barman, C.R. (1994). The learning cycle as an alternative method for college science teaching. *Bioscience* 44, 99 – 101.

Arthur, A.W and Gamson, Z.F. (1987). Seven principles for good practice in undergraduate education. *American Association for Higher Education Bulletin* [reprinted online]. Available: http://honolulu.hawaii.edu/intranet/committees/FacDevCom/guidebk/teachtip/7princip.htm [2008, April 17]

Bulmer, M., O'Brien, M. and Price, S. (2007). Troublesome concepts n statistics: a student perspective on what they are and how to learn them. In UniServe Science *Symposium Proceedings, Science Teaching and Learning Research including Threshold Concepts.* Sydney: University of Sydney 9 - 15.

Ebert-May , D., Brewer, C. and Allred, S. (1997). Innovation in large lectures: Teaching for active learning. *Bioscience* 47, 601-607.

Moore, D.S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review* 65, 123 – 165.

Moore, D.S. (2007). *The Basic Practice of Statistics*, 4th edition. New York: Freeman.

Novak, J.D. and Gowin, D. (1984). *Learning How to Learn.* Cambridge: Cambridge University Press.

Parman, C. and Halling, C. (2007). *affyQCReport: A Package to Generate QC Reports for Affymetrix Array Data.* [online] Available: http://bioconductor.org/packages/2.0/bioc/vignettes/affyQCReport/inst/doc/affyQCReport.pdf . [2007, June 18].

Rea, D. (2003). Sustaining teachable moments on the complex edge of chaos. American Educational Research Association Conference [online]. Available: http://www.ccaerasig.com/papers/03/SustainingTeachable.htm [2008, April 17].

Richardson, A.M., Lidbury, B.A. and Zhang, F. (2007). One potato, two potato, three potato, four: the use of *Hot Potatoes* software in science language comprehension. In UniServe Science *Symposium Proceedings, Science Teaching and Learning Research including Threshold Concepts.* Sydney: University of Sydney 83 - 88.

Schau, C., Stevens, J., Dauphinee, T.L. and Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Towards Statistics. *Educational and Psychological Measurement* 55, 868 – 875.

Zhang, F. and Lidbury, B.A. (2006). Its' all foreign to me: learning through the language of genetics and molecular biology. In UniServe Science *Symposium Proceedings, Assessment in Science Teaching and Learning*. Sydney: University of Sydney 153 - 159.

WHY DO STATISTICS EDUCATORS USE EXAMPLES TO TEACH STATISTICS?

GORDON Sue and NICHOLAS Jackie
The University of Sydney

Abstract
Examples have long been an integral component of statistics teachers' instructional repertoires but tend to be in the background of pedagogical knowledge. We explore the diverse ways that university statistics educators use examples, drawing on data from recent research (Gordon, Reid & Petocz, 2007). Three overlapping categories are proposed: examples are developed and presented by educators in basic instruction, examples are generated by students, under teacher direction, to aid learning and examples connect statistics with students' future professional work. Expressions in the second category were sparse suggesting an opportunity for statistics educators to develop teaching. We review models of exemplification in mathematics education and relate these to the empirical findings to begin the development of a framework for characterising examples in statistics education. We conclude that examples help promote statistical literacy.

INTRODUCTION

The use of examples to teach statistics at tertiary level is well documented in the published literature. These publications refer to various roles for examples. Niglas and Osula (2005) mention the importance of giving students examples of good statistical practice and examples to explain statistical concepts. Empirically they found that students valued practical examples to help them understand explanations, a finding that is echoed by Wulff and Wulff (2004). Online and journal resources for statistics education include data sets, applets and detailed examples that can be used in a lecture setting, such as resources published by causeweb.org (discussed by Bloomer Green, McDaniel & Holmes Rowell, 2005) and Teaching Bits in the Journal of Statistics Education. Macnaughton (2004) presents a conceptual approach to teaching introductory statistics emphasising relationships between variables. Primary concepts are presented to students in terms of numerous practical examples. Standard statistical principles and methods are then developed, again with emphasis on practical examples. Sowey (2001) and other educators demonstrate their ideas and provide exemplars as a resource for teaching statistics.

This brief 'sampling' of the statistics education literature suggests that the use of examples is central to the practice of teaching university statistics but is often in the background of pedagogical knowledge. Teachers may not necessarily articulate their reasons for using examples. Further, as Atkinson and Renkl (2007) point out, while example-based learning environments can be effective in supporting learning in some students, they may promote passive processing and hence poor student performance on subsequent, more complex tasks.

In this paper we present empirical data on why and how international university educators use examples based on recent research. We then review exemplification in mathematics and outline a framework relating these mathematical models to the empirical results. We discuss how examples help develop statistical literacy.

EXAMPLES IN EDUCATORS' REPORTS ABOUT TEACHING AND LEARNING STATISTICS AS A SERVICE COURSE

We report data from a research project investigating international educators' ideas about teaching statistics in 'service' courses—for students studying statistics as part of their various major study areas. Detailed descriptions of the methodology and project findings are available in publications on this project, including: Gordon, Reid & Petocz (2007); Gordon, Petocz Reid (2007), Petocz, Gordon & Reid (2006) and Reid, Petocz and Gordon (2008). Hence the methodology is summarised here only briefly.

Participation was invited through IASE (International Association for Statistics Education) and Faculty bulletin boards of Australian universities. In all, 36 IASE members from

12 different countries and 8 other Australian statistics educators took part in the project. The investigation consisted of a three-phase series of e-mail interviews based on an initial set of six open-ended questions. These questions included: *What do you consider to be the most important aspects of statistics for your teaching?* and: *What are the attributes of a good statistics teacher at university?* Two rounds of follow-up questions probed participants' responses in depth enabling a personal dialogue in which educators could reflect on and expand their initial responses to questions and communicate their ideas.

Although participants were not specifically asked about their use of examples to teach statistics, two thirds of the educators spontaneously referred to examples in their e-interviews. From a content analysis of the interview transcripts we derived three categories for the use of examples from the respondents' perspectives: educators' examples used in basic instruction; student-generated examples as an aid to learning and examples presented to demonstrate statistics as a methodological tool in future professional work. There is no suggestion that these categories are mutually exclusive. Indeed, since the overall aim of teaching is to promote learning, all pedagogic tools are in essence learning tools. However, the categories are separated analytically to highlight the different foci that emerged from the data.

A. EXAMPLES AS INSTRUCTIONAL DEVICES USED BY EDUCATOR

In this category the educator develops examples to engage students (and themselves), explain concepts or illustrate procedures, build skills and guide students' thinking. We summarise the varied uses of examples in this category in Table 1. All pseudonyms were chosen by participants themselves and brief excerpts are reported under these self-chosen pseudonyms.

Table 1: Ways that examples are used in instruction

Educators' use of examples	**Illustrative quote**
To engage students (and oneself)	Ron Fisher: *I am constantly updating my examples, and looking for new applications that will interest my students. Not only do I do this for the students' sake, it also makes the class much more interesting for me, since I am interested in the world around me.*
Illustrate ideas of lectures	Andrew: *The large methods courses have a one hour tutorial each week where examples are worked on that illustrate the lectures of the previous week.*
To ground concepts	Joyce: *[How do real world examples help students to learn statistics?] To use an educational psychology phrase, it gives them "an anchor".*
As a template for students to follow	Natalie: *By giving students worked examples they can use these as a "template" for their own work until they are comfortable creating their own non-technical explanations and conclusions.*
Develop skills	Margaret: *To develop good case examples for students to work on so that they develop their skills in a step by step fashion. To learn how to guide students and keep their interest as they go from simplistic examples to more complicated ones.*
Extra practice	Kay: *(Struggling) students get extra worksheets with examples.*
Differentiate statistics from mathematics	Primavera: *Students believe that Statistics is a branch of Mathematics. They change their mind with the use of real examples.*
Way into theory	Baz: *Even the students who can handle theory can learn from illustrative examples, and students who can't have no other choice.*
Develop critical thinking	Jane Johnson: *I sometimes use examples of incorrect analyses – as a warning to those who do not think critically.*
Build conceptual complexity	Despina: *I try to structure the problems I use as examples and as tutorial work so that we begin with a basic problem and slowly add complexity.* Kay: *We go over a number of examples, spread throughout the course –*

	distributed versus massed practice, "spiralling" to repeat earlier concepts.
Demonstrate the statistical process	John: *We will use an example reported in the media to illustrate how we can identify the statistical investigative process and understand statistical aspects of the study as reported.*
Indicate variation	Daria: *(We show students that variation is present in everything we do) by means of many "real" examples in the course of the lecture, and tutorials in the computer.*
Personalise teaching	Natalie: *Even though (many of us) use the same master resources, each of us adapt these materials slightly – add in our own examples, tweak to our own preferences.*

Many participants expressed the idea that the capability to develop and apply relevant and real examples—from the home discipline—is the hallmark of a good teacher. Conversely, poor examples, irrelevant to students' interests, can lose students. As Tilito summed up: *My students are from different disciplines, and all those examples of card or balls in probability, are obsolete, because they don't see where to apply them.*

B. LEARNER GENERATED EXAMPLES

Expressions fitting into this category were sparse.

Horace reported that if students could: *draw a picture, give the definition, state an example, and show they know when to use something, then that's a pretty good operational definition of understanding?! (If they can write the formula, that's a nice bonus!)*

Cara encouraged students: *to find examples of misuse (of statistics) on their own and present them in class.*

The few reports in this category indicated that student generation of examples helped students construct their own knowledge. As Janet Cole explained: *(The process) helps students put together/construct their own frameworks for learning. If a student can understand the process of constructing a confidence interval for a proportion (including checking conditions, mechanics, and presentation of results), then it should ideally be easier for that student to transfer this process to the construction of a confidence interval for any other situation. The student is not learning something entirely new, but is rather doing what I call a "variation on a theme."*

C. EXAMPLES PRESENTED TO DEMONSTRATE APPLICABILITY OF STATISTICS AS A METHODOLOGICAL TOOL IN FUTURE PROFESSIONAL WORK

In this category examples indicated the work of statisticians, or related to students' disciplines or future professional work.

Andrew: *It is important to convey to students the view that what they learn in a preparatory course on statistics is going to be essential for their future work in their major subject, something that many first year students do not appreciate. // Therefore, the statistics teacher must be able to access appropriate examples from the areas of application. //It is with these types of applications that students suddenly realise the greater picture of statistics. Such examples abound but they must be chosen very carefully. They must not be artificial class exercises in my view.*

Glee added that from his experience: *students respond very well to the statistics if the examples given in class relate to their specific disciplines. The motivation is that when they leave University they will be relevant and employable in their respective job markets.*

In turn, Johanna found that starting the semester by asking students why they think statistics has been made a compulsory unit in their degree program, led to discussion about the role of statistics in their particular field—how, when and why it would be used. *I use many examples that relate to the degree program that the students are enrolled in but also ask them to consider examples from different areas (highlights the diversity of statistical applications).*

John felt that in evaluating reports in the newspaper there should ideally be a balance of examples, some good reporting and some not such good ones. *Even although examples (of poor* *quality reports) demonstrate precisely what we want to make the students to be aware of, I don't feel comfortable about it—there is a certain element of glee and self-righteousness on our part when we come across such blunders in the newspaper*.

Statsboy: *So, in teaching statistics to medical students you MUST use medical examples – show me how the authors have analysed their data and what it means. Don't write a regression model with alphas and betas without giving me a relevant example of how this works in real life.*

Henry VIII: *What I try to show medical students is that, even if they don't ever intend to do any research, they still need some basic knowledge of stats in order to be able to fully understand the concepts of "statistical patterns" and "typical values", and the probabilistic nature of the decisions they have to make every moment during their practice. I try to do this by highlighting, through examples, the probabilistic nature of the patterns and decisions, and by trying to steer them away from the sort of deterministic thinking they are exposed to during most of the other courses they attend at college. // My most successful lecture on the Probability subject, for instance, is the one in which I discuss the interpretation of sensitivity, specificity and predictive value of diagnostic tests—these are very practical application of the rather abstract Bayes' theorem, and the students love it.*

The spontaneous reports about examples by participants, reported in section 2, indicate that the majority of them used examples, chosen by themselves, for a range of instructional goals. A few participants reported instances of examples generated by students in learning activities, while many educators used examples to indicate the utility of statistics in the student's own discipline.

The first two categories have much in common with the ways examples are used in mathematics and we review some models in the literature.

EXEMPLIFICATION IN MATHEMATICS

Bills et al (2006), in their overview of the role of examples in teaching and learning mathematics, give a broad definition of what constitutes an example—anything used as raw material for generalising, including intuiting relationships and inductive reasoning, illustrating concepts and principles, indicating a larger class, motivating, exposing variation and change and practising technique.

Examples play a central role in learning mathematics (Bills et al, 2006). Watson and Mason (2002a) note that people learn mathematics through engagement with examples rather than through formal definitions and techniques. They argue that it is only through this engagement that definitions assume any meaning. They advocate that students are exposed to a range of examples, as identifying the commonalities and differences in these examples is crucial for concept development (Watson & Mason, 2002b).

Bills et al (2006) argue that examples play a key role as communication tools between teacher and learner—fundamental to explanation and mathematical discourse—and their choice depends on factors including the teaching goals and the teacher's awareness of their learners' preconceptions and dispositions. However, the example chosen does not always fulfil its intended purpose. Mason and Pimm (1984) explored the inherent difficulties in teachers presenting their students with a generic example of a technique or theory. They argue that unless a teacher explicitly draws attention to the characteristics of the example that exemplifies the technique or theory, their students may focus on the particular example, resulting in students trying to "learn the example".

Recently, there has been a change in emphasis away from teacher-oriented to student-oriented activities in the mathematics classroom, leading to an increased emphasis on students constructing their own examples (Bills et al, 2006). Students may be asked to generate their own examples for assessment purposes or motivate interest in a new topic (Watson & Mason, 2002b). Watson and Mason (2002b) argue that student-generated examples can play an invaluable role in concept development. Dahlberg and Housman (1997) studied the use of student-generated examples in concept development in a mathematics undergraduate course. They showed that the generation of and reflection on examples provided powerful stimuli for learning. They found that students who employed an example generation learning strategy were more effective in attaining an initial understanding of a new concept than those who did not.

The ideas discussed above have implications for teaching statistics as well as highlight possible pitfalls about examples in learning statistics.

The central role of examples in mathematics has lead to a number of researchers characterising their use. Watson and Mason (2002b) proposed a framework of five types of intended experience for students when engaged with the task of constructing examples. Michener (1978) considered examples in a conceptual framework for understanding mathematics categorising (noteworthy) examples into four, not necessarily disjoint, epistemological classes: start-up examples that help motivate fundamental concepts and set up useful intuitions in a new subject; reference examples that are basic, widely applicable and used repeatedly in a branch of mathematics; model examples that are indicative of the general case and counterexamples which demonstrate what something is not. These classes resonate with some participants' reports of using examples to teach statistics, presented in section 2. However, a major difference is the emphasis on statistics as a tool for shedding light on real-life problems, rather than illustrating abstract mathematical concepts or theories.

Ideas about exemplification in mathematics education could develop example use in statistics pedagogy. We begin a framework, informed by ideas of Watson and Mason (2002b), Michener (1978) and the empirical categories that emerged from this study. The outline is sketched in Table 2. Illustrative examples fitting three boxes (a, b, c) follow the table.

Table 2: Framework for characterisation of examples in statistics education

Classification of example → Role of example in practice ↓	Start-up example for motivation	Developmental example to underpin/build concept or procedure	Example to illustrate generic structure	Counterexample
Educator's use in basic instruction		a		
Learner generated			b	
Methodological tool to demonstrate application in discipline	c			

Our data suggest some illustrations of the framework. John used examples from the media to underpin the "statistical investigative process" (box a); Janet Cole explained that student construction of a confidence interval for a proportion illustrates the generic structure of confidence intervals (box b); examples applying Bayes' theorem motivate the study of probability theory for medical students, according to Henry VIII (box c). The framework provides opportunities for statistics educators to reflect on, characterise and assess their own pedagogical use of examples.

CONCLUSION: EXAMPLES TO PROMOTE STATISTICAL LITERACY

Our empirical data show that statistics educators use examples in diverse ways for teaching including, importantly, promoting understanding of statistics applications in many disciplines. The findings suggest, too, that examples play an central role in enhancing statistical literacy, portrayed by Gal (2002) as the ability to interpret, critically evaluate, and communicate about statistical information and messages. Examples help develop understanding of concepts underpinning statistical knowledge, assist in developing awareness of context and promote critical thinking skills. Our participants' reports support many of these aspects, such as Kay's "spiralling" examples to develop concepts; Jane Johnson's "warning" examples of incorrect analyses and

Statsboy's applications of regression models in a medical context. By engaging and motivating students, examples enhance the affective components of learning statistics, which, Gal (2002) stresses, support the knowledge bases of statistical literacy. Respondents' expressions in Category C are at heart about making statistics personally meaningful to students.

Research is needed to assess the effects of different types of examples on student learning and how best to utilise examples for enhancing statistical literacy. Mathematics frameworks and models of exemplification could stimulate ideas on teaching statistics and expand statistics educators' repertoires in practice by offering a basis for a more reflective and theorised understanding of examples. Our framework begins this task.

REFERENCES

Atkinson, R. & Renkl, A. (2007). Interactive example-based learning environments: Using interactive elements to encourage effective processing of worked examples. *Educational Psychology Review*, 19, 375–386.

Bills, L., Dreyfus, T., Mason, J., Tsamir, P., Watson, A. & Zaslavsky, O. (2006). Exemplification in mathematics education. In J. Novotna (Ed.), *Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education*, 125-154, Prague, Czech Republic: PME.

Bloomer Green, L., McDaniel, S. & Holmes Rowell, G. (2005). Online Resources for Non-Statisticians Teaching Statistics, *MERLOT Journal of Online Learning and Teaching*, 1(2). Retrieved February 18, 2008, from: http://jolt.merlot.org/documents/Vol1_No2_rowell.pdf.

Dahlberg, R. & Housman, D. (1997). Facilitating learning events through example generation, *Educational Studies in Mathematics*, 33, 283-299.

Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review*, 70(1), 1-25.

Gordon, S., Petocz, P. & Reid, A. (2007). Tools, artefacts, resources and pedagogy—stories of international statistics educators. In *Australian Association for Research in Education 2006 Conference Papers*, compiled by P. L Jeffery, AARE, Adelaide. Available online at: http://www.aare.edu.au/06pap/gor06358.pdf.

Gordon, S., Reid, A. and Petocz, P. (2007). Teachers' conceptions of teaching service statistics courses. *International Journal for the Scholarship of Teaching and Learning*, 1(1). Available online at: http://www.georgiasouthern.edu/ijsotl/v1n1/gordon_et_al/IJ_Gordon_et_all.pdf.

Macnaughton, D. (2004). The Introductory Statistics Course: The Entity-Property-Relationship Approach. Available online at: www.MatStat.com/teach.

Mason, J. & Pimm, D. (1984). Generic examples: Seeing the general in the particular. *Educational Studies in Mathematics*, 15(3), 277-289.

Michener, E. (1978). Understanding understanding mathematics, *Cognitive Science*, 2(4), 361-383.

Niglas, K. & Osula, K. (2005). University-level data analysis courses with the emphasis on understanding and communication of statistics – a ten years action research project. IASE /ISI Satellite conference.

Available online at: http://www.stat.auckland.ac.nz/~iase/publications/14/niglas.pdf.

Petocz, P., Gordon, S. & Reid, A. (2006). Recognising and developing good statistics teachers. In Rossman, A. and Chance, B. (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, ICOTS7*, Salvador, Brazil: ISI, Voorburg, The Netherlands (On CD). Available online at: http://www.stat.auckland.ac.nz/~iase/publications/17/5B2_PETO.pdf .

Reid, A., Petocz, P. & Gordon, S. (2008). Research interviews in cyberspace. *Qualitative Research Journal*, 8(1), 47-62. Available online at: http://search.informit.com.au/documentSummary;dn=121182566086364;res=E-Library.

Sowey, E. (2001). Striking demonstrations in teaching statistics. *Journal of Statistics Education* [Online], 9(1), www.amstat.org/publications/jse/v9n1/sowey.html.

Watson, A. & Mason, J. (2002a). Extending example spaces as a learning/teaching strategy in mathematics. In A. Cockburn & E. Nardi (Eds.) *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education*, 377-385, Norwich, UK: PME.

Watson, A. & Mason, J. (2002b). Student-generated examples in the learning of mathematics. *Canadian Journal of Science, Mathematics and Technology Education*, 2(2), 237-249.

Wulff, S. & Wulff, D. (2004). "Of course I'm communicating; I lecture every day": Enhancing teaching and learning in introductory statistics. *Communication Education*, 53(1), 92–103.

Technology for insight into student beliefs about statistics in large classes

Michael Bulmer and Emma Low
University of Queensland

Abstract
Prior beliefs and attitudes of students can have a significant impact on their learning experience but it is usually difficult to engage with student beliefs in depth when dealing with large classes. As part of an ALTC Associate Fellowship project, we have developed technologies and strategies for facilitating connections between staff and student beliefs through the embedding of student reflective writing in statistics courses. Students were free to write whatever they liked in their journals but weekly themes were also provided to give them a starting point if needed. The aim of this paper is to give an overview and analysis of entries around the themes that were particularly related to beliefs about statistics, as well as to demonstrate the use of text mining tools in this context.

INTRODUCTION

Journal writing is widely used as a means of promoting deeper learning by students through reflection (Beveridge, 1997). In earlier work (Bulmer & Rodd, 2005) we extended this by using online reflective writing in a medium sized class (around 100 students) to also gain insight into emotional responses to the curriculum design and the associated learning environment. This has the potential to support a pedagogic resonance (Trigwell and Shale, 2004) through rich dynamic feedback on learning but in practice, the volume of writings that needs to be read limits this approach. With large classes it becomes impossible for an individual lecturer to fully engage with the text. One aim of our current work is to trial the use of text mining software (such a *Leximancer*) to assist in connecting lecturers with their students.

At the same time we have also been interested in using student creative works to learn about student beliefs and attitudes towards mathematics and statistics in the context of an introductory statistics course at university. These creative works have included drawings (Bulmer & Rolka, 2005; Rolka & Bulmer, 2005) and poetry writing (Bulmer & Rolka, 2008) and have been used to identify coarse categories of differing worldviews. While being a somewhat crude and indirect lens on beliefs, the constrained nature of these creative works makes them an efficient tool for the analysis. In contrast, the open-ended reflective writings give an opportunity for greater insight but again the volume makes them inaccessible. By focussing writing in a particular week on a theme related to beliefs, and making use of text mining software, we will explore the use of reflective writings in this context.

WRITINGS

The project has involved developing online technologies for students to make their writings and for lecturers to then access the writings by student, by date, or through search functions. Each student was given a random pseudonym by the system so that the lecturer did not know the identity of the student authors but could still follow entries over time from a particular pseudonym. The initial pilot of this technology in 2007 involved a class of 526 students in an introductory statistics course. Students were asked to keep regular journal entries (at least 100 words in most weeks of the course) as part of their assessment during semester. Together they made 6940 entries for a total of just under a million words of text.

Students were free to write whatever they liked in their journals but weekly themes were also provided to give them a starting point if needed. Table 1 shows the themes chosen for the first four weeks. These had a particular focus on background knowledge, attitudes and beliefs, with themes in later weeks being more tied to content topics or assessment tasks in the course.

Table 1. Journal themes for Weeks 1 – 4

Week	**Theme**
1	*I'm doing this course because...* Start your Learning Journal with thoughts about why you are doing this course, what 'statistics' means to you, what you are looking forward to or dreading in the course ahead, ...
2	*After one week I feel...* With the first week behind you and all your courses underway, use your Learning Journal to summarize your feelings.
3	*My relationship with maths...* This week we come to our first serious maths in the course, using logarithms to transform nonlinear equations. Take this opportunity to summarize your relationship with mathematics over the years and at the moment.
4	*Randomness...* With a focus on probability in the course this week, use your Learning Journal to talk about randomness and probability in your life.

WEEK 1

In the first week of semester students were encouraged to write in their journal "with thoughts about why you are doing this course, what 'statistics' means to you, what you are looking forward to or dreading in the course ahead, ..." (see Table 1). A total of 301 entries (around 42000 words) were written during the week. Some students may have also written on this theme in the second week, since they would have still been writing their first entries, but here we will restrict attention to students who wrote their entries before the second theme was posted (on Sunday morning).

Figure 1 shows the concept map produced by *Leximancer* from the full collection of 301 entries. The points in this picture are the concept entities that have been extracted automatically from the data. The arrangement of entities on the map comes from a method related to multidimensional scaling so that two entities that are close together on the map appeared in similar conceptual contexts. The circles are concept themes, groups of entities, which can be added to the map. (The number of concept themes can be changed interactively, often giving very different looking maps, though still with the same general arrangement of concept entities).

The largest concept theme shown here relates to the lectures, including 'lecturer', 'book', 'time' and (pleasingly) 'interesting'. These come from students describing their first week in the course, rather than addressing the specific theme that was given. Similarly another large concept theme, 'data', arises from the topics discussed in the second lecture of the course, with students using their journal to simply summarize what they had "learned" in the lecture. (For many students it took several weeks of guidance before they engaged more deeply with the reflective aim of the journal.) Neither of these concept themes are directly linked with the journal theme we proposed. Of course this is fine – students didn't have to write to the theme – but for the purposes of this paper we would like to analyse the entries specifically related to the Week 1 theme. We skimmed the 301 entries and identified 181 (around 28000 words) that seemed to do this, at least in part.

Figure 2 shows the concept map for this subset of entries. Concept themes around lectures and variability are still present but there is now a much larger theme around 'statistics'. This theme appears away from the 'lecture' theme, suggesting that the two are appearing separately in the student writings and so entries contributing to the 'statistics' concept theme may be addressing the topic of why the students are doing the course and what statistics means to them. It also includes concept entities such as 'school', 'maths', and 'science', which might be expected in writings on this topic.

Looking at the text underlying the 'statistics' concept, most entries focusing on why the student is doing the course but some did also link it with their previous experiences, as in the following example:

> I've never really studied statistics as a science/subject unto itself before. Like most people I learnt a little about statistics in maths and science during highschool. However the importance of stats and data analysis always seemed to be dwarfed by the actual collection of data or experimentation.

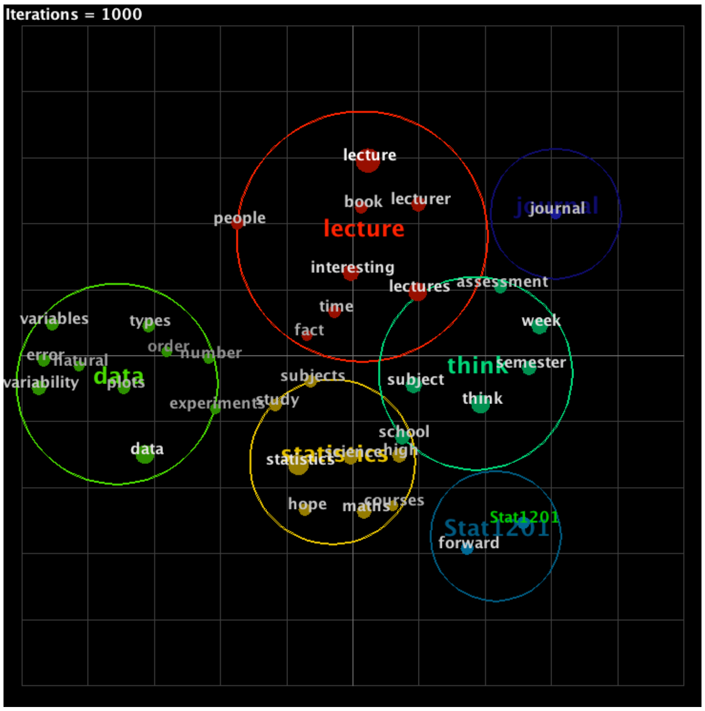


Figure 1. Concept map output for full Week 1 entries

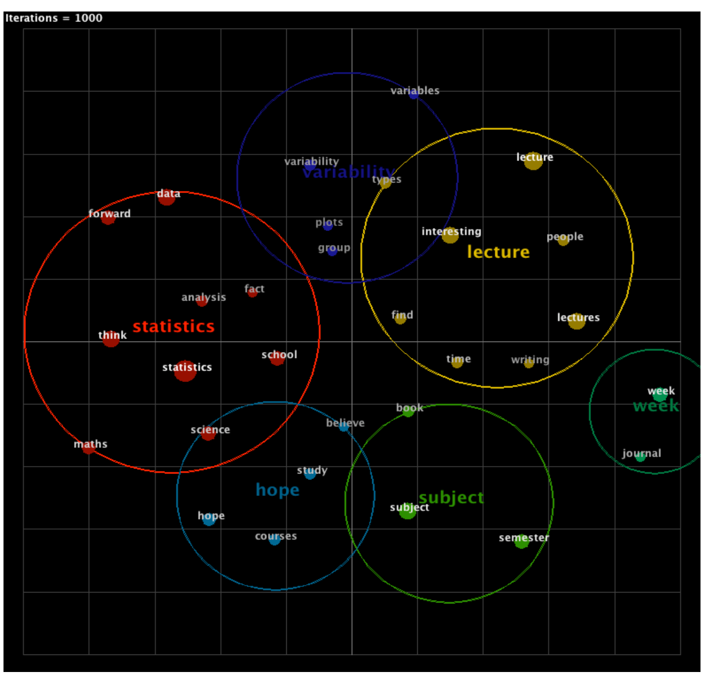


Figure 2. Concept map output for subset of Week 1 entries

A few students did show a broader appreciation of the role of statistics in science. The course was mainly taken by biology students, such as this one:

> I've always realised that statistics is important in biology to filter out significant data from “background” variations in data. I've come across scientific papers before but I’ve never really understood the many statistical terminology used in those papersa and the deductions drawn from the many statistical tests used.

WEEK 4

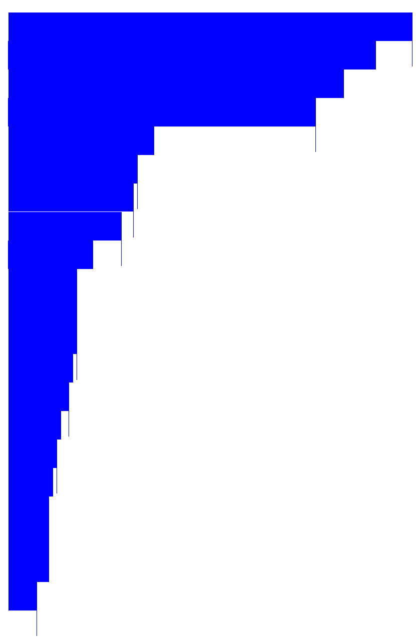
Simple probability and random variables were introduced in the fourth week of the course. It was felt that this was an area where students may have mixed backgrounds and attitudes and so a theme was given, encouraging them to reflect on “randomness and probability” in their lives.

In this week there were a total of 501 entries (around 71000 words). Most of these were actually not about the theme, focussing mainly on the assessment item that had been handed out that week and on the public holiday that fell in the middle of the week. A simple search for ‘random’ in the text extracted 96 entries (around 15000 words). For the purposes of this study we then skimmed each entry to determine whether it was specifically related to the theme, removing entries that were just a summary of the lectures, resulting in 75 entries (around 12000 words) for analysis.

Table 2 shows the concepts automatically identified by *Leximancer* in these 75 entries, together with their frequency in the entries. There is scope within *Leximancer* to add or delete concepts, or to merge concepts that are really synonyms. Here we avoided this, partly because we wanted to test the default behaviour (which would be most useful for novice or automated usage) but more specifically because many students used the roles of “randomness” (noun) and “random” (adjective) quite differently. A concept like “statistics” **however,** was mostly used as the name of the course being studied, rather than a discipline based on probability. This is reflected in the concept map shown in Figure 3, where “statistics” and “study” are grouped in the same theme.

Table 2: Concept entities for “randomness” theme

Concept	**Count**
probability	90
randomness	82
random	75
life	69
think	33
people	29
time	28
day	26
interesting	19
chance	16
statistics	16
uni	16
lecture	15
study	14
back	12
happen	11
friend	10
topic	9
high	9
find	9
variables	7



As with the Week 1 entries, the automated concept map gives a valuable starting point for exploring the text in more detail. A useful approach is to look at the entries that involve pairs of the common concepts. The ‘life-random’ pair included many entries such as:

> So this week's theme is randomness and probability. This got me thinking about randomness in general. There is a high probability that some random things will happen to you in your life - opportunities arise seemingly out of nowhere, people become your friends from random chance encounters, you see funny things that make for a good anecdote when you get home ("It was so random!").

Similar ideas came through ‘people-randomness’, with the idea of controlling randomness:

> There are many different examples and occurrences of randomness in my life. So I don't quite remember at the moment, but random all the same.Looking at the people around me, it's weird to think that we're all somehow linked together. Everyone in Brisbane is linked somehow...I feel amazed when I think of it.For example, a friend I knew from tutoring, went to school with a guy who is now friends with my ex-boyfriend's sister, through dancing. It's quite crazy! I'm sorry if I'm completely off the topic, but yes, this is the best I can do at 9 in the morning. Randomness is everywhere. I try to control most that occur around me, but it's far too hard.

Our earlier work (Bulmer & Rodd, 2005) focussed on emotional and attitudinal dimensions in student writings. Several entries discussing ‘life-randomness’ also illustrated these broader themes:

> Probability - What is probable? That throughout the course of the day I’ll probably get annoyed, bored, angry and/or lonely. Usually all of the above. I feel like I should do something about this, but I am apathetic by nature and just can’t be bothered. I’ll wait patiently instead. If anything, I’m good at that.

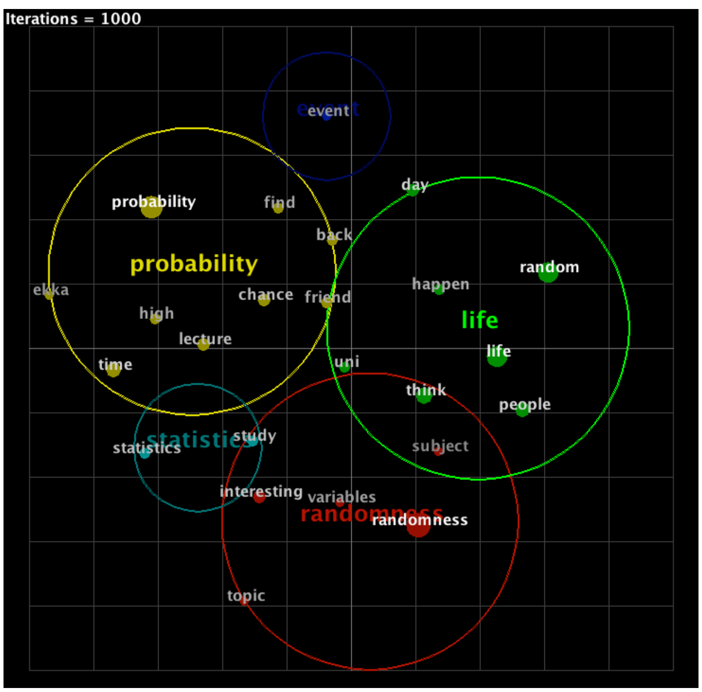


Figure 2. Concept map output for “randomness” theme

DISCUSSION

In the previous sections we have used samples of student writings to illustrate the concepts found in the data. Of course the choice of these samples is rather subjective and it may be quite misleading to imply we are representing 103000 words of student reflections with just a handful of such samples. An interesting benefit of using text-mining software is that it may help “make the analyst aware of the global context and significance of concepts and to help avoid fixation on particular anecdotal evidence, which may be atypical or erroneous.” (Smith and

Humphreys, 2006). The interplay between the automated concept generation and the subjective analysis of the underlying text seems particularly valuable in this context.

Clearly the wording of the themes is crucial if reflective writings might also be used for insight into more specific questions, such as those related to worldviews and attitudes. In Week 1 the vast majority of students did not get beyond answering "why are you doing this course" when in fact we were much more interested in what 'statistics' meant to them, particularly as a basis for establishing a pedagogic resonance. Our earlier creative projects (Bulmer & Rolka, 2005) involve the single task of expressing the students' understandings of statistics and consequently gave results that were easier to analyze.

However, the concepts that came through the writings on 'randomness' were successful in highlighting the rich informal, even colloquial, understanding of randomness and chance that students bring to their learning of statistics. These vignettes provide valuable starting points for discussion in the classroom and in turn through further reflective writing.

ACKNOWLEDGMENTS

This work was supported by the Australian Learning and Teaching Council (ALTC) through the first author's Associate Fellowship project.

REFERENCES

Beveridge, I. (1997) Teaching your students to think reflectively: the case for reflective journals. *Teaching in Higher Education*, 2 (1), 33-43.

Bulmer, M. and Rodd, M. (2005) Technology for nurture in large undergraduate statistics classes. *International Journal of Mathematics Education in Science and Technology*, 36 (7), 779-787.

Bulmer, M., and Rolka, K. (2005) The "A4-Project" – Statistical world views expressed through pictures. In *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (H. L. Chick & J. L. Vincent, eds.), Vol. 2, 193-200.

Bulmer, M. and Rolka, K. (2008) Student beliefs and attitudes from haiku writing in statistics. Submitted for publication.

Rolka, K. and Bulmer, M. (2005) Picturing Student Beliefs in Statistics. *Zentralblatt für Didaktik der Mathematik*, 37 (5), 56-61.

Smith, A.E. and Humphreys, M.S. (2006) Evaluation of Unsupervised Semantic Mapping of Natural Language with Leximancer Concept Mapping. *Behavior Research Methods*, 38 (2), 262-279.

Trigwell, K. and Shale, S. (2004) Student learning and the scholarship of university teaching. *Studies in Higher Education*, 29 (4), 523-536.

A DISCUSSION OF SOFTWARE CHOICES IN TEACHING OF DATA MINING TECHNIQUES

Siva GANESH
Massey University, Palmerston North, New Zealand

Abstract
Many tertiary institutions now include 'Data Mining' as a topic in their Statistics curriculum, both at undergraduate and postgraduate levels. The choice of software for learning the topic of Data Mining is an interesting issue to think about. There is a wide range of such software available, from commercially popular ones such as SAS/Enterprise Miner, Statistica/Data Miner and S-plus/Insightful Miner to free ones such as R and Weka. The main aim of this paper is to discuss the pros and cons of such software, including their capabilities to handle and manipulate large volumes data, all from a teaching/learning point of view at both introductory and more advanced levels.

INTRODUCTION

From a statistical perspective, *Data Mining* can be viewed as the exploration and analysis of large volumes of data. In other words, Data Mining, like *Statistics*, is the science (or art) of discovering useful summaries such as patterns, structures and interesting relationships in complex data. The combination of fast computers and cheap storage makes it easier to extract (useful) information out of everything from supermarket buying patterns through Banking and Insurance assessments to email filtering and micro-array analyses. As described by Hastie *et.al.* (2001), the common feature of Data Mining and Statistics is "learning from data".

Teaching of Statistics involves developing and adapting robust procedures for understanding underlying concepts, and for the management and analysis of statistical data. The field of Statistics is constantly challenged by problems arising from business, science and industry. Traditionally, the statistics curriculum at *tertiary* or *university* level deals with data often collected to answer specific questions. However, in the modern *information age*, vast amounts of data are collected, stored and retrieved, often automatically, with the advent of powerful computers. Since *computation* plays a major role in the process of *making sense of data*, computer scientists have a significant claim over the ownership of the field of Data Mining. Modern databases often contain several thousand or millions of records (megabytes, gigabytes or even terabytes in size), and data of such magnitude clearly put into context the futility of standard statistical techniques. This means, if all (or even a sample) of such large data is to be processed during an analysis, adaptive or algorithmic techniques need to be utilised. These techniques have been the concern of computer scientists working in the areas of machine learning and pattern recognition. Note also that, the knowledge of fields such as Information Systems (Database management & Data warehousing) would also be a necessity to handle large data.

Nevertheless, modelling techniques in Data Mining, in general, have a statistical base; and statisticians are now showing a significant interest in the area, including offering tertiary courses in data mining with a statistical perspective, sometimes in partnership with computer scientists. Example prescriptions of such courses are shown below.

PRESCRIPTIONS OF SOME UNIVERSITY COURSES IN (STATISTICAL) DATA MINING

At present, Massey University offers a 3rd year course entitled 'Statistical Machine Learning' aimed students majoring in statistics, mathematics or computer science, and another course entitled 'Statistical Data Mining' aimed at postgraduate students in statistics. The corresponding prescriptions are, *Statistical Machine Learning:* "Introduction to artificial intelligence methods and statistical learning; supervised learning; neural networks; linear methods of regression and classification; Bayesian and kernel classifiers; tree based methods; unsupervised learning; k-means; self-organizing maps; principal components and statistical clustering;

optimization and genetic algorithms. (Offered jointly by Statisticians and Computer scientists.)"; and *Statistical Data Mining:* "Principles of data mining with statistical underpinning of techniques for supervised and unsupervised learning: classification and regression trees; multi-layer neural networks; nearest neighbours; support vector machines; bayesian classifiers and belief networks; association rules; dimension reduction; segmentation; self-organising kohonen maps; ensemble models with hybrid including bagging, boosting and random forests; text mining; rare event prediction; use and assessment of modern software. Examples from recent research literature and case studies will be used to illustrate techniques."

The description for a course at Australian National University (2008) reads, "The main focus of the course will be supervised learning, primarily for classification. The emphasis will be on practical applications of the methodologies that are described, with the R system used for the computations. Attention will be given to: Generalizability and predictive accuracy, in the practical contexts in which methods are applied. Low-dimensional visual representation of results, as an aid to diagnosis and insight. Interpretability of model parameters, including potential for misinterpretation. There will be very limited attention to regression methods with a continuous outcome variable. Relevant statistical theory will mostly be assumed and described rather than derived mathematically. There will be somewhat more attention to the mathematical derivation and description of algorithms. Topic to be covered include: Basic statistical ideas - populations, distributions, samples and random samples; Classification models and methods - including: linear discriminant analysis; trees; random forests; neural nets; boosting and bagging approaches; support vector machines; Linear regression approaches to classification, compared with linear discriminant analysis; The training/test approach to assessing accuracy, and cross-validation; Strategies in the (common) situation where source and target population differ, typically in time but in other respects also; Unsupervised models - kmeans, association rules, hierarchical clustering, model based clusters; Low-dimensional views of classification results – distance methods and ordination; Strategies for working with large data sets; Practical approaches to classification with real life data sets, using different methods to gain different insights into presentation; Privacy and security; Use of the R system for handling the calculations."

These are two classic examples of Data Mining courses (with a statistical flavour) offered at University level. Note from the coverage of topics that, the techniques explored in the courses fall into two main categories: *Supervised learning (or Predictive modelling)* mainly referring to building models from data with the specific goal of predicting future outcomes, in particular via classification and regression models, and *Un-supervised learning (or Descriptive modelling)* referring mainly to segmentation or clustering, association rules and general EDA.

SOFTWARE USAGE

The last decade or so have seen hundreds of computer software manufacturers jumping onto the Data Mining bandwagon. Major statistical software packages such as SAS/Enterprise Miner (www.sas.com), S-plus/Insightful Miner (www.insightful.com), SPSS/Clementine (www.spss.com) Salford Systems (www.salford-systems.com) and STATISTICA/Data Miner (www.statistica.com) are being marketed as Data Mining solutions more so than 'Statistical'. With the sophistication and competitive edge of software packages, some would (still) attempt to have you believe that, armed with large data and software tool, answers will come pouring out! Despite this *myth*, the field of Data Mining is having a major impact on business, industry and science, and it affords research opportunities for analytical and methodological developments.

Many university courses in (Statistical) Data Mining use one or more of the above mentioned and other commercially successful software packages, usually assisted by freeware software such as the R system (www.r-project.org) and Weka (www.cs.waikato.ac.nz/ml/weka/).

From a teaching/learning point of view, a Data Mining software tool must include, ability to manipulate data: data selection, sampling, cleansing, transforming etc.; scalability: ease of handling large volumes of data; powerful analytical and modelling capabilities including ensemble techniques; visualization, evaluation and assessment capabilities: concept of validation and testing data (via bootstrapping, 10-fold CV etc.); performance metrics such as misclassification error, ROC, lift, loss functions, etc.; ability to provide accurate and usable information efficiently; and easy-to-use interface (preferably a good GUI).

Brief descriptions of the two software suits used at Massey University are given below.

SAS/ENTERPRISE MINER
(Ref: http://www.sas.com/technologies/analytics/datamining/miner/)

SAS Enterprise Miner streamlines the data mining process to create highly accurate predictive and descriptive models based on analysis of vast amounts of data from across the enterprise. Forward-thinking organizations today are using SAS data mining software to detect fraud, anticipate resource demands, increase acquisitions and curb customer attrition.

The key features of Enterprise Miner are: 1. A broad set of tools supporting the complete data mining process providing flexible software that addresses complex problems regardless of the user's data mining preferences or skill level; 2. The data mining power is delivered via an easy-to-use, drag-and-drop interface designed to appeal to experienced statisticians as well as less-seasoned business analysts, with the advanced analytic algorithms organized under the core tasks that are performed in any successful data mining endeavour; 3. Provides superior analytical depth with an unmatched suite of predictive and descriptive modelling algorithms, including decision trees, bagging and boosting, neural networks, memory-based reasoning, hierarchical clustering, linear and logistic regression, associations, sequence and Web path analysis, and state-of-the-art methods such as gradient boosting, partial least square regression, support vector machines and integrated text mining for analysis of both structured and unstructured data; 4. Sophisticated set of data preparation, summarization and exploration tools ease the burden of data preparation, usually the most time-consuming aspect of data mining endeavours (tools include, data sampling, data partitioning, outlier filtering, merge and append, batch and interactive plots, segment profile plots, easy-to-use graphics explorer wizard, interactively linked plots and tables, data transformations, interactive variable binning and data replacement); 5. Model comparisons via lift curves, statistical diagnostics, ROI metrics and interactive examination of posterior probability distributions; 6. Automated scoring process (applying fitted model to new data) via interactive scoring and generation of score code in SAS. C, Java and PMML; and 7. Open, extensible design for easily adding tools and personalized SAS codes including interactive editor features for training and score code.

SAS/Enterprise Miner works on Windows, AIX, Solaris, Linux and many other platforms except for MacOS!

THE R SYSTEM (Ref: http://www.r-project.org/)

R is free software for statistical modelling, graphics and a general programming environment, thus can be regarded as a combination of a statistics package and a programming language. R is completely free and we can make any modifications we want and share it with other users! It works on Windows, MacOS, Linux, and many Unix variants. R is not supported by any commercial enterprise, but it has a very active development community, and there are institutions that offer training courses etc.. The R manuals are complete (if not always helpful) and there are many books on it. R has an enormous number of standard and cutting-edge statistical (and data mining) functions built in, a wide variety of (free) add-in packages that add even more functions, and we can extend it further. Almost every standard statistical analysis can be carried out in R. R is mostly command-line driven (although various graphical interfaces have been developed), thus makes it harder for many (students) to use but allows flexibility and documentation and repetition of analyses. (Ref: http://wiki.r-project.org/rwiki/)

Data Mining capabilities of R are substantial. A good start would to explore the GUI oriented tool named *Rattle* (www.togaware.com). According to the developers, Rattle (the R Analytical Tool To Learn Easily) is a data mining toolkit used to analyse very large collections of data. Rattle presents statistical and visual summaries of data, transforms data into forms that can be readily modelled, builds both unsupervised and supervised models from the data, presents the performance of models graphically, and scores new datasets. Through a simple and logical graphical user interface based on *Gnome* (www.gnome.org), Rattle can be used by itself to deliver data mining projects. Rattle also provides an entry into sophisticated data mining using the open source and free statistical language R. The aim is to provide an intuitive interface that takes you through the basic steps of data mining, as well as illustrating the R code that is used to achieve this. Whilst the tool itself may be sufficient for all of a user's needs, it also provides a stepping stone to more sophisticated processing and modelling in R itself, for sophisticated and unconstrained data mining.

A sample of R packages available for data mining are, *nnet* for single-hidden-layer neural network, *rpart* for recursive partitioning - tree-structured models for regression, classification and survival analysis, *mvpart* for an adaptation of *rpart* for multivariate responses, *knnTree* for a tree algorithm fitting nearest neighbors in each node, *maptree* for graphical tools for the visualization of trees, *lasso2* and *lars* for regularized and shrinkage methods, *randomForest* for the implementation of the random forest algorithm for regression and classification, *ipred* for bagging for regression, classification and survival analysis as well as bundling (a combination of multiple models via ensemble learning), *gbm* and *boost* for various forms of gradient boosting, *svm* together with packages such as *tune* and *svmpath* from library *e1071* for support vector machines, *kernlab* for implementing a flexible framework for kernel learning, *arules* for implementations of Apriori algorithm for association rules (and for mining frequent itemsets), functions *hclust* from package *stats* and *agnes* from package *cluster* for agglomerative hierarchical clustering, function *diana* for divisive hierarchical clustering, *dendrogram* for improved visualization for cluster dendrograms, function *kmeans* from package *stats* (with several algorithms) for computing partitions with respect to Euclidean distance, function *pam* (with wrapper *clara* for large data) from package *cluster* for partitioning around medoids, *mclust* for fitting mixtures of Gaussians using the EM algorithm, *som* for self-organizing maps, *daisy* (and *randomForest*) for computing proximity measures between objects, *cmdscale* for creating MDS plots and *RWeka* for an interface to a rich toolbox of partitioning algorithms available in Weka,

One of the major problems reported in the literature relating to R is its inability to handle very large data sets in an efficient manner. However, there have been remedies suggested and an example can be found in Maindonald (2007).

CONCLUSION

Course evaluations at Massey University suggest that students benefit from starting with a menu-driven GUI oriented software (such as SAS/Enterprise Miner and R/Rattle) for learning basics of Data Mining ideas. Students are then happy to continue with exploring advanced and modern techniques via coding or programming within both software tools.

The author has some experience in using software such as S-plus/Insightful Miner (and S-codes), STATISTICA/Data Miner, Salford Systems tools such as CART, MARS, TreeNet and RandomForests and Weka tools (Ian H. Witten, Ian H and Frank, Eibe (2005)), and the presentation of this paper at the conference will consist of further information on these packages with example screenshots. SAS/Enterprise Miner and R (including Rattle) will also be demonstrated, briefly. An extensive list of software for Data Mining can be found at www.kdnuggets.com.

REFERENCES AND ACKNOWLEDGEMENTS

Australian National University (2008)
http://studyat.anu.edu.au/courses/MATH6210;details.html

Hastie.T., Tibshirani.R. and Friedman.J.H., (2001) "Elements of Statistical Learning: Data Mining, Inference and Prediction", Springer Verlag.

Ian H. Witten, Ian H and Frank, Eibe (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

Maindonald (2007)
http://www.maths.anu.edu.au/~johnm/courses/dm/statminers/labs/largishDataset.pdf

STATISTICS LEARNING: MAKING IT EASY FOR NON-STATISTICIANS

Ayse Aysin BILGIN, Kehui LUO and Subramanyam VEMULPAD
Macquarie University, Sydney, Australia

Abstract
Nowadays, statistics is taught to university students in many non-statistics disciplines, such as chiropractic and health. In this study we examined and compared chiropractic students' experiences of using two statistical software programs, one web based and another Excel based, to identify the strength and weakness of each program that might influence statistical learning.
Data were collected through a survey of students, and analysed using the chi-squared and the McNemar tests. The web based program appeared to have a good potential due to its particular features and functions, flexibility and easy accessibility. Although the web based program has better graphical displays, broader range of analytic capabilities and remote access, students found the Excel based program relatively easier to use than the web based one. This study suggests that more online instructions and explanations are needed in the web based program for non-statistics students, to aid their statistical learning.

INTRODUCTION

Chiropractic is the third largest primary healthcare profession (the other two being medicine and dentistry) in the world. Its origins dating back to 1895, it is the youngest of the three. Chiropractic education could be considered to be in its infancy, with very few training programs being situated in the academic environment of a University. Being a primary healthcare profession, chiropractic education is regulated by accreditation bodies. In Australia there are three University-based training programs (Macquarie, RMIT and Murdoch Universities) and the training standards are regulated by the Council for Chiropractic Education Australia (CCEA, 2007).

In the recent decades public acceptance of chiropractic is increasing, owing to improving quality and quantity of chiropractic research (Feise, 2001). The importance of imparting research literacy including critical appraisal and appreciation of statistical analyses for promoting Evidence Based chiropractic has been reiterated by educators (Maloney, 1990; Meeker, 2000; Cramer, 2000; Long *et al*, 2004). The study of statistics provides ideas and methods that could be utilised to better understand the environment at academic level and for day-to-day activities (Garfield & Ben-Zvi, 2007). Utts (2003) also emphasized the importance of statistics education for everyone. He stated that educated citizens should understand basic statistical concepts so that they can detect any misuse of statistics by policy makers, physicians and others.

At Macquarie University, the chiropractic curriculum includes 'Research Methods for Health Sciences' as a third year subject in which students are taught, *inter alia*, statistical analyses in the context of appraising research publications and critical appraisal of published work. This subject was introduced in 2005. In the past, EcStat (McNeil *et al*, 2006), an add-in program to Excel, was the only statistical software package that was provided to students who were studying that course. EcStat has been in use by the first year statistics students at Macquarie University since 1999. Although user friendly and easy to install, EcStat is limited in its statistical procedures, and does not provide any information about the procedures (no built-in help function), nor any relevant statistical knowledge. To improve students' statistics literacy and learning experience, students in the 'Research Methods for Health Sciences' unit were required to use not one but two separate statistical software packages in the second semester 2007: an Excel based statistical add-in and a newly developed web-based statistical program, WebSTAT (Bilgin *et al*, 2007). WebSTAT includes all the basic and some advanced statistical procedures, including a number of graphical procedures, which are available in most other statistical packages including EcStat. It was conceived to enhance learning and teaching of statistics, and also improve flexibility and accessibility. Furthermore, as its name suggests, WebSTAT is a web-based program, and it is accessed through the web (http://statquiz.efs.mq.edu.au/webSTAT/). A snapshot of a WebSTAT home page is given in Appendix (Figure A.1).

WebSTAT was initially developed to meet the needs of Macquarie university students with various backgrounds, such as students in first year statistics courses, research students in statistics, and researchers in non-statistics areas and also students majoring in non-statistics areas, such as chiropractic science. The developers of the program had one main objective, apart from creating a user-friendly and widely accessible statistical software, which was to increase the level of understanding of statistics including statistical concepts, applications and results' interpretations, and the degree of appreciation of the statistical tools for statistical analysis by the non-specialist users.

In this study, we aim to compare users' experience in using WebSTAT and EcStat for statistical analysis. The data were collected from third year chiropractic students who used both WebSTAT and EcStat programs during their study in second semester, 2007. The results from this study, including the understanding of students' experience of using WebSTAT, would provide valuable information for us to further develop and modify WebSTAT, and to make it a learner-centred statistical software which assists with the understanding of statistics and encourages statistical thinking by users, particularly non-statisticians. Also, in the light of obtaining valuable suggestions from the students in comparison with the features and functions of EcStat, this could guide the further improvement of WebSTAT.

BACKGROUND

The students in this university research methods course are introduced to a variety of statistical tools that have been commonly used in the chiropractic profession to test study hypotheses and answer research questions, and widely referred to in journal articles in the area (for example, Foster & Bagust, 2004; Haas *et al*, 2004). Practical sessions are instituted for students to learn "how to do", *i.e.* practicing with techniques learnt in lectures. In the practical sessions, students are given opportunities to gain hands-on experience with the techniques and their applications. Through answering a number of well-designed questions, they are encouraged to think "why" and "when" the techniques can be used, what statistical results would be produced from a computer software package and how to interpret the statistical results.

Assessment is an important part of curriculum development, and from students' point of view, assessment always defines the actual curriculum (Ramsden, 2003: p.182). In other words, students are most likely to give higher importance to what was assessed than what was taught in a course. The practicals in this 'Research Methods for Health Sciences' course at Macquarie university form a part of formal assessments. They are designed to help students to learn "how to do" certain analyses by using computer software packages (Figure A.2 and Figure A.3 in Appendix). They are also used as preparation for summative assessment tasks in the course, such as assignments. In one such practical session, students were asked to read a paper by Haas *et al* (2004), and then analyse the data set used in this paper themselves to gain hands-on experiences of "how to". After that, they were required to answer the following questions.

(a) The authors appear to have treated the pain level itself as the outcome variable, rather than the pain reduction. How is this justified? Explain.
(b) Did the authors treat the dose level as a nominal variable or as a continuous variable? Explain.
(c) The authors do not appear to have used analysis of variance. What method did they use?

These questions require higher order thinking skills (Biggs 2003) since the students need to understand different statistical methods and then critically judge a research article which has been published in their studying area, chiropractic science.

Rubin (2007) stated that no statistical software package would ever teach statistics without a curriculum and a teacher. However, meaningful error messages or warnings from software packages could be very valuable in guiding the use of the statistical tools for learners of any software. If they are informative enough to flag possible problems and require thinking and decision making by users (learners), the outcome of the learning process for the users would be more effective and informative.

In light of the suggestions from earlier users, we have started creating "descriptions" and "warning messages" for individual tools, to improve the documentation of WebSTAT. Figure 1 shows an example of the help page for the "Scatter Plot Matrix" (more details can be accessed from WebSTAT home page).

Figure 1: Example of the WebSTAT Scatter Plot Matrix Help in WebSTAT.

Warning messages
"Maximum is 20 variables for scatter plot matrix": No more than 20 variables are allowed.
"There is No data in selected variables!!"
"Can not process Categorical Variables for Scatter Plot. Maximum category is 10 groups" : The Other variable has more than 10 categories.
"Y axis data isn't numeric and more than 10 category"
"X axis data isn't numeric and more than 10 category"
"All data are missing"
"Samples must be equal size": WebStat can generate scatter plots from data in different tables, but the sample sizes must be equal.
"Non-numerical data problem": WebStat can generate scatter plots from non-numerical data but each such data value must start with a number. For example, '1: Male', '2: Female' or '3: Unknown'. For example, labels like 'one: Male', 'two: Female', and 'null' cannot be plotted.

The assignments as a part of summative assessments require students to answer questions of different order of thinking. Low level thinking (straight forward) questions such as "Explain the difference between taking a random sample from a target population and doing a randomized control study." are mingled with a higher order thinking questions such as "Rewrite the abstract of the paper that you have read in week 3" and/or "Imagine that the Health Authorities are so impressed with your report that they are willing to provide a large research grant to enable you to undertake further studies addressing the research question. Briefly indicate what further studies you would undertake". Interpreting written reports, rather than just performing computations is identified as one of the important ways for assessing statistical thinking (Watson, 1997). We consider 're-writing an abstract' and 'formulating a proposal for further study for a published work' also require statistical thinking and can not be done by simply learning to perform different analyses. In addition to that, Wood & Petocz (2008) highlighted the importance of statistics thinking in statistics education in a recently published paper, and stated that "A student should leave the course confident that a new problem can be approached with a way of thinking, not just a basketful of techniques." This is one of the motivations behind the present study.

DATA SET AND METHODS

A short survey questionnaire was developed with 17 questions about users' experience with WebSTAT and EcStat. Ninety students were enrolled in the 'Research Methods' course and given the opportunity to use both WebSTAT and EcStat in completing their coursework in second semester 2007. Unfortunately not all 90 students enrolled in the course were available to be included into this sample due to some logistic limitations, and instead a practical class of 30 students participated in the survey, which can be considered as a reasonable random sample of the 90 students, since each practical class in this course was formed in a rather random manner.

Cross-tabulations and bar charts were used to explore our data, comparing students' experience in using WebSTAT and EcStat. The McNemar test (Simonoff, 2003: p.279) was used to further evaluate the differences in the student experience in using the two programs, given the paired nature of the data.

RESULTS

Of the 30 students in our sample, 63% reported their preference to be EcStat with only 17% preferring WebSTAT and 20% having no preference. Compared to EcStat, WebSTAT's accessibility over the internet was more appealing to 15 out of the 27 students (*i.e.*, 56%).

Majority of the students reported that the computer output generated from both WebSTAT and EcStat were sufficient and useful.

In terms of the best features, a greater proportion of students, 23%, were impressed by the flexibility offered by WebSTAT, compared to EcStat (17%), but a much higher proportion of students, 41%, thought that EcStat was rather easy to use compared to WebSTAT (27%), as shown in Table 1. The fact that none of the observed differences in the proportions are statistically significant may mainly be due to the small sample size.

Table 1: Comparison of the best features between WebSTAT and EcStat.

Best feature	WebSTAT (30 responded)		EcStat (29 responded)		P-value
	n	%	n	%	
Flexibility	7	23	5	17	≈ 1
User Friendly	10	33	11	38	0.804
Easy to learn and use	8	27	12	41	0.332

The students who preferred EcStat over WebSTAT (19) mainly complained about WebSTAT being rather difficult to use and expressed interest in wanting more explanations and instructions in using WebSTAT. Of those who preferred WebSTAT over EcStat (6), most complained about the latter's complexity and suggested that EcStat can be too fiddly and is less accessible. Furthermore our students' preference for one of the two packages compared was not associated with the extent of their paid-work commitment, nor with their locality, *i.e.*, how long it would take for them to travel to the university campus.

A slightly higher proportion of students believed that WebSTAT provided more informative output (18.5%) than EcStat (14.8%), while the majority (66.7%) of students believed they were similar in this aspect. 44% of students thought that WebSTAT and EcStat gave similar quality graphical displays; while 16% chose WebSTAT and 40% chose EcStat as a better program to produce quality graphical displays, ie, a difference of 24% between the two programs. However the difference is not statistically significant (p-value > 0.05). This might be due to the small sample size.

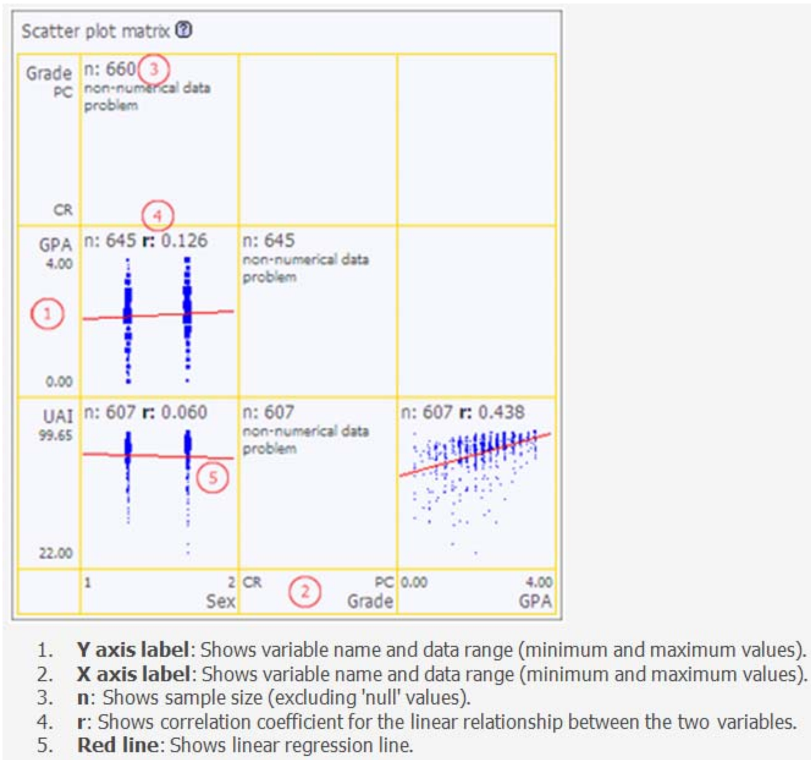
DISCUSSION AND CONCLUSION

WebSTAT in its current status does not appear to be superior to EcStat, at least from the chiropractic students' perspective. However, it produces comprehensive outputs compared to Ecstat. It also provides a wider range of statistical procedures compared to EcStat, including more complicated statistical model procedures which could not be evaluated by the students participated in this study.

As identified by Wild and Pfannkuch (1999) statistical thinking has four dimensions: an investigative cycle, an interrogative cycle, types of thinking and dispositions. The warning messages and limited statistical knowledge provided within the current version of WebSTAT, and more documentation and statistical knowledge to be added in the future version of WebSTAT, would certainly assist users with their statistical thinking in at least the first two dimensions suggested by Wild and Pfannkuch.

To make WebSTAT a learner-centred statistical program, the documentation of WebSTAT would need to be improved further, with more instructions to each procedure and more relevant statistical knowledge to guide interpretation of results. Figure 2 shows part of the "Scatter Plot Matrix" help page which is designed to reduce the worry of "how to do?" and enable students to concentrate on "why are we doing this?" and "what does that mean?" questions. However, WebSTAT needs more detailed warning messages to alert users to think about the techniques that they are applying and to provide more statistical knowledge within the package to help them with interpretation of statistical results, and thus may assist them to make intelligent decisions in relation to their research questions. More explanations (interpretation) for the outputs are also needed to make it more useful for non-statisticians.

Figure 2: A scatter plot matrix from WebSTAT.



It is reasonable to believe that the perceptions of non-statistics students towards statistical softwares would be very different from statistics students and/or lecturers in statistics. In general, chiropractic students do not like statistics and find it hard to learn, which might be the reason they had some negative feelings about WebSTAT and the use of it. Whether they like the software or not may not make much difference on their liking statistics. Hubbard (1997), based on the research of Gal and Ginsburg (1994), and with no quantitative data, states that for many students statistics is, "merely a hurdle to be overcome on the way to obtaining a degree". If so, could we make statistical education pain free by developing statistical software which not just provides required methods but also guides statistical thinking?

Wells claimed that "statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write" (quoted in Castles, 1992). The American Statistical Association's (2005) GAISE report emphasised the development of statistical thinking and conceptual understanding rather than mere knowledge of procedures. Also, Australia is facing a skill shortage in mathematics and statistics graduates who could perform analyses when required by other professionals (Australian Academy of Science, 2006). Therefore it is important for every professional to acquire statistical literacy to help them to analyse their own data.

To conclude, WebSTAT has got a good coverage of statistical procedures and functions for their targeted users, and provides neat and informative outputs and quality graphical displays. It also has got a good potential due to its features and functions, and in particular its flexibility that it is accessible via the web wherever and whenever. However, it needs to be improved in a number of dimensions. First of all, having more comprehensive information on statistical concepts built within WebSTAT is desirable. Such provision of statistical knowledge will guide and assist non-statistics users in doing their statistical analyses with ease, and also help them to understand and thus be able to interpret the results from the analyses. The built-in statistical knowledge within WebSTAT also provides opportunities of learning or re-learning statistics for users while they are practicing the use of statistical analysis procedures. This would encourage users to concentrate more on the statistical concepts behind rather than the procedures. Because of this, users including our students are more likely to think statistically. They will also be encouraged to relate their statistical knowledge learnt in their lectures, to their experience gained in using WebSTAT and possibly to statistical applications in their daily life and/or work.

As WebSTAT will be mainly used by non-specialist users, online help including error-warning messages on many basics is essential. A more detailed online documentation on all procedures, features and functions of WebSTAT is also in need for all users, particularly first-time users.

ACKNOWLEDGEMENTS

We would like to acknowledge Emeritus Professor Don McNeil for his continued dedication to students' learning. His infectious enthusiasm and thought provoking discussions continue to inspire us.

REFERENCES

American Statistical Association (2005). Guidelines for assessment and instruction in statistics education. Retrieved on 16 January 2008 from
http://www.amstat.org/education/gaise/GAISECollege.htm

Australian Academy of Science (2006). Mathematics and statistics: Critical skills for Australia's future. Retrieved on 16 June 2008 from
http://www.review.ms.unimelb.edu.au/Report.html

Biggs, J. (2003). Teaching for Quality Learning. Buckingham: Open University Press.

Bilgin A.A.B., Luo K., Middledorp, J. (2007). Web-based Student Learning of Statistics. in Enhancing Higher Education, Theory and Scholarship, Proceedings of the 30th HERDSA Annual Conference [CD-ROM], Adelaide, Australia, 8-11 July. (E1)

CCEA (Council for Chiropractic Education Australia) Accessed on 10 December 2007 from http://www.ccea.com.au

Cramer G. D. (2000). Development of research infrastructure within the chiropractic profession. Presented to Fifth Chiropractic Research Agenda Conference Chicago, 23 July 23.

Feise R. J. (2001). Evidence-based chiropractic: The responsibility of our profession. Journal of the American Chiropractic Association 38(1): 50-52.

Foster I.A., & Bagust J. (2004). Cutaneous two-point discrimination thresholds and palpatory sensibility in chiropractic students and field chiropractors. Journal of Manipulative and Physiological Therapeutics 27(7): 466-471. http://www.sciencedirect.com/science/journal/01614754

Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. Journal of Statistics Education 2 (2), 1-15.

Garfield, J. B. & Ben-Zvi, D. (2007). Developing students' statistical reasoning: Connecting research and teaching practice. Emeryville, CA: Key College Publishing (in press).

Haas M., Groupp E. and Kraemer D.F. (2004). Dose-response for chiropractic care of chronic low back pain. The Spine Journal 4: 574-583.

Hubbard, R. (1997). Assessment and the process of learning statistics. Journal of Statistics Education 5 (1), 1-8.

Long C.R., Nick T.G., Kao C. (2004). Original Research Published in the Chiropractic Literature: Evaluation of the Research Report. Journal of Manipulative and Physiological Therapeutics 27: 223-228.

Maloney P. (1990). Integrating statistics courses into chiropractic curriculums: a developing need. Abstract from Symposium of the Pacific Consortium for Chiropractic Research, 2 July, 1989, Monterey, California; cited in Journal of Manipulative and Physiological Therapeutics. 13(1): 56-57.

McNeil, D., Chilvers, M., Middledorp J., Petersons M, Shaw, P. (2006). Modern Statistics: A Graphical Introduction. Southwood Press Pty. Ltd. Marrickville, Australia.

Meeker W. C. (2000). Concepts germane to an evidence-based application of chiropractic theory. Top Clin Chiropr. 7: 67-71.

Ramsden, P. (2003). Learning to Teach in Higher Education (2nd Ed.). London: Routledge Falmer
Rubin A. (2007). Much has changed; little has changed: Revisiting the role of technology in statistics education 1992-2007. Technology Innovations in Statistics Education 1(1): Article 6. Retrieved on 10 March 2008 from http://repositories.cdlib.org/uclastat/cts/tise/vol1/iss1/art6
Simonoff J. S. (2003). Analyzing Categorical Data, Springer: New York.
Utts, J. (2003). What educated citizens should know about statistics and probability. The American Statistician 57(2), 74-79.
Watson J. M. (1997) Assessing Statistical Thinking Using the Media. In Gal I. & Garfield J.B. (editors) The Assessment Challenge in Statistical Education (pp.107-121). IOS Press. http://www.stat.aucland.ac.nz/~iase/publication/
WebSTAT Home Page, Macquarie University, Department of Statistics, Retrieved on 10 March 2008 from http://www.statquiz.efs.mq.edu.au/webSTAT/
Wild C. J. & Pfannkuch M. (1999). Statistical Thinking in Empirical Enquiry. International Statistical Review 67(3): 223-265.
Wood G. & Petocz P. (2008) One in a Million: An Alternative Transformation of the First-Year Statistics Course. Asian Social Science, 4(3): 44-48. Retrieved on 10 March 2008 from www.ccsenet.org/journal.html

Appendix

Figure A.1: WebSTAT Homepage.

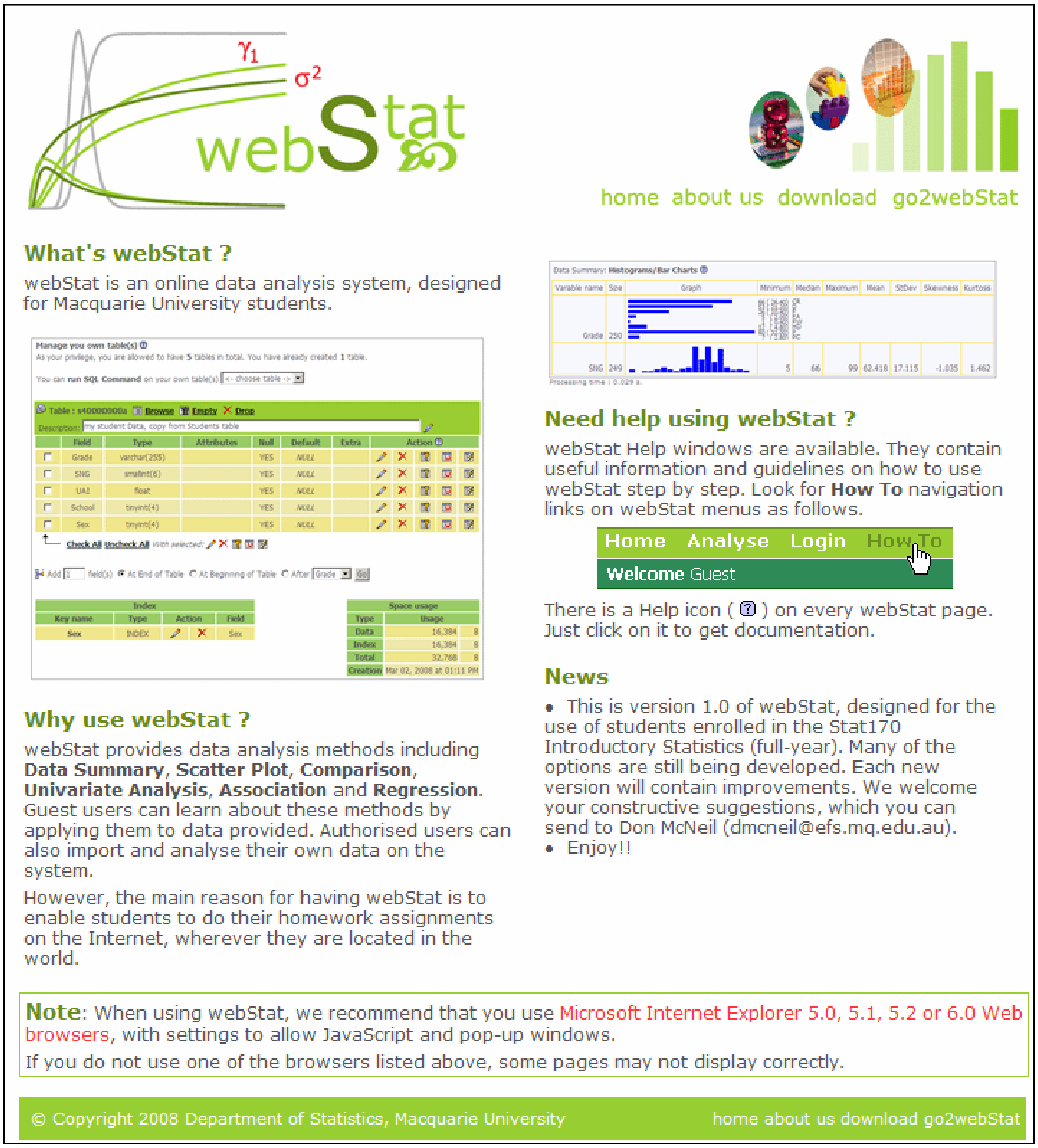


Figure A.2: WebSTAT Output.

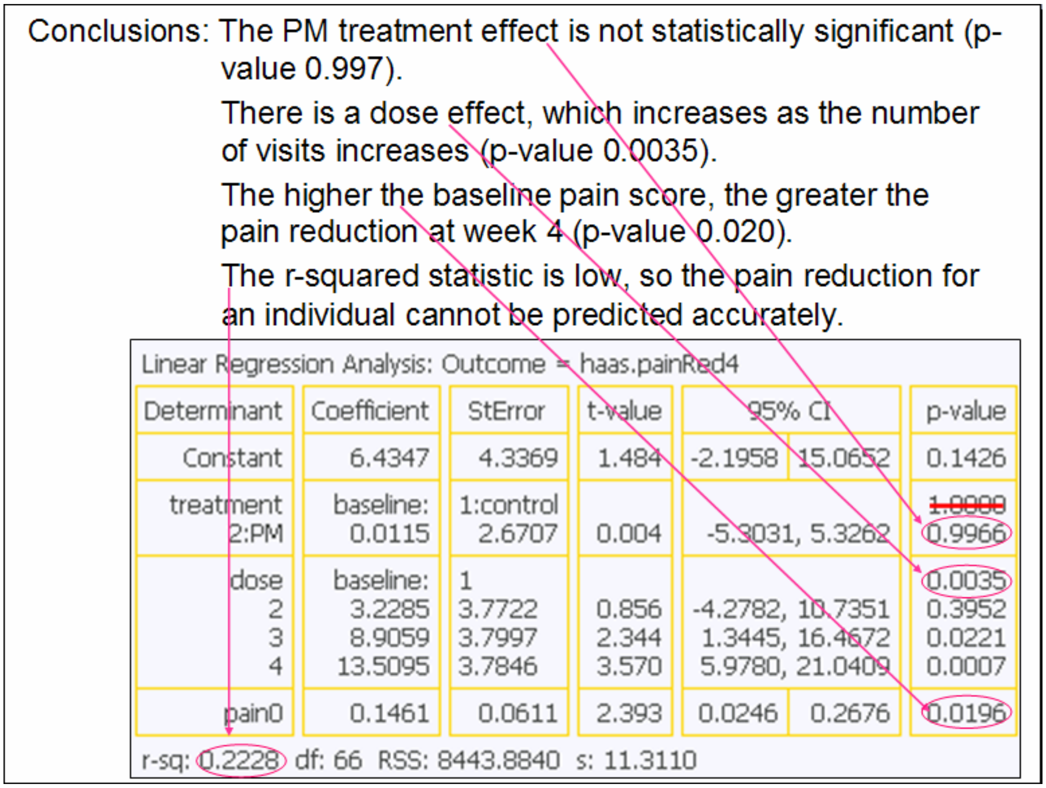
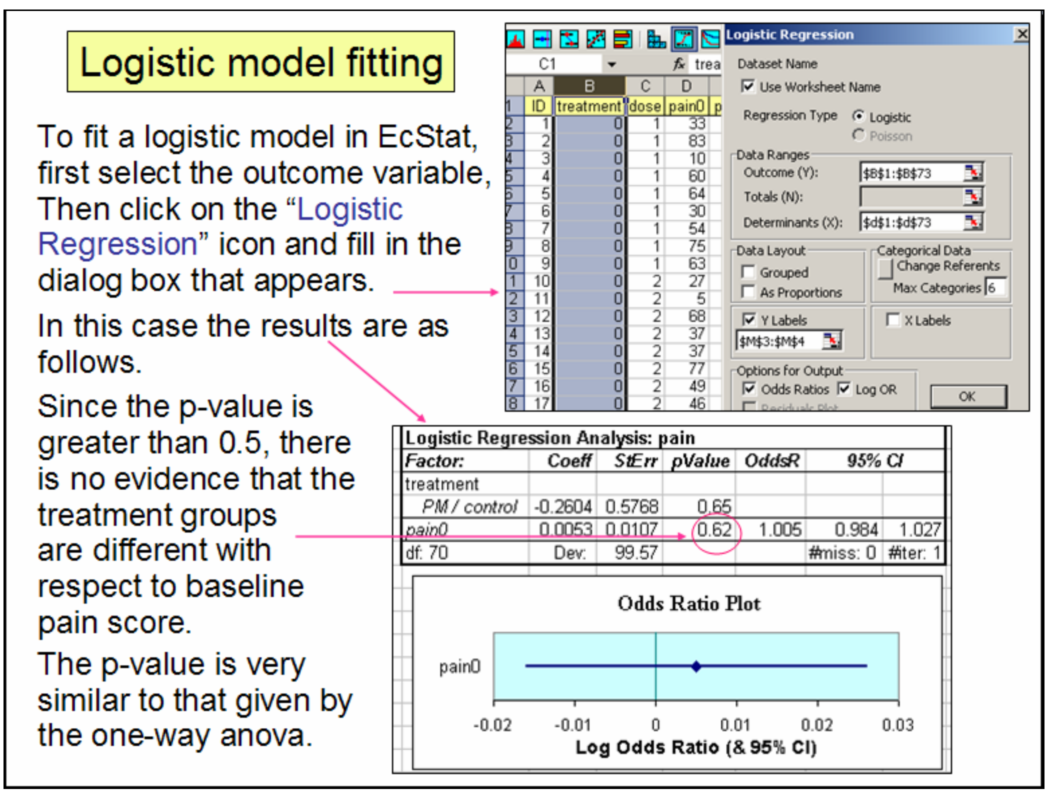


Figure A.3: EcStat Output.



Contributed Paper – Caro-Ann Badcock

DRUG DEVELOPMENT – WHAT STATISTICAL SKILLS ARE REQUIRED BY WHOM AND WHEN?

BADCOCK Caro-Anne
Shimsco Consulting, Australia

Abstract
The development of safe, efficacious and cost effective medicines and medical devices requires dedicated teams of people with diverse backgrounds including all manner of the sciences, regulatory affairs, marketing, health economics and so on. Statistics and statisticians also have an integral role in the drug development industry. In fact, it is probably one of the few industries where there is a regulatory requirement that statisticians MUST be involved in each project from the design of an experiment through to the report. The life cycle of a medicinal product, focusing on the statistical needs of various team members and others exposed to the product will be described to understand the range of statistical skills required by the various disciplines within the pharmaceutical, medical device and biotechnology sectors.

PHARMACEUTICAL INDUSTRY

Globally there are now more than 200 major pharmaceutical companies, jointly said to be more profitable than almost any other industry. The large companies often participate in a broad range of drug discovery and development, manufacturing and quality control, marketing, sales and distribution with many qualified staff in each area. Smaller organisations, such as the local biotechnology companies are more likely to focus on specific aspects such as discovering therapeutic candidates or developing formulations independently or collaboratively with larger pharmaceutical companies. As such they are more likely to lack staff with some of the critical knowledge required to take their product to market.

Each year only about 25 truly novel drugs are approved for marketing – this is after about 1000 potential drugs being tested before one reaches the point of being tested in a clinical trial. Marketing approval comes only after heavy investment in pre-clinical development and clinical trials, as well as a commitment to ongoing safety monitoring. Drugs or devices which fail part-way through this process often incur large costs while generating no revenue in return. If the cost of these failed products is taken into account the cost of developing a successful new product has been estimated as up to 1.7 billion USD in 2003 (including marketing and other business expenses) [Bain].

A product life-cycle typically consists of discovering a new chemical entity or designing a new device, undertaking pre-clinical and clinical studies, manufacturing and quality control, gaining approval to market and then promoting and selling the product with sales and marketing teams. Statisticians are not commonly found through all these stages in a product life-cycle but strong statistical skills are certainly helpful and in some cases required.

DRUG DEVELOPMENT PROGRAM

It typically takes up to 15 years to develop a new drug from the time of finding a new active substance to gaining approval to market a new product! Apart from meeting the regulatory requirements companies are always looking for ways to shorten the time to approval. Providing scientists with a strong understanding of the practical benefits of statistics may encourage them to seek more formal statistical advice. For example, can the process of selecting potential new chemical entities using high throughput screening be optimised?

Once a potential new chemical entity is discovered there are the pre-clinical studies that include the toxicology and animal studies. In addition, the process by which the chemical is made will be optimised so that it can be manufactured in large enough amounts for general use and its suitability to be made into capsules, tablets, aerosol, intramuscular injectable, subcutaneous injectable or intravenous formulations assessed. Together these processes are known as CMC:

Chemistry, Manufacturing and Control. Again statisticians and scientists and manufacturers with strong statistical skills would be beneficial to any of these processes.

On completion of the pre-clinical studies an Investigational New Drug (IND) application is submitted to the relevant regulatory authorities to gain approval to start testing the new drug in humans. Following IND approval three phases of progressively larger human clinical trials may be conducted and form the New Drug Application (NDA) to gain permission to market this product. Generally phase I trials study toxicity on healthy subjects (50-100 subjects), phase II trials study the effects on limited numbers of patients usually to understand the pharmacokinetics and determine safe dose levels (100-200 patients) and phase III trials are studies comparing the new drug to placebo, to prove efficacy and comparing new drug to current standard treatment, to prove it is at least as efficacious and hopefully safer (500-5000 patients). Once the drug is marketed the research doesn't stop there. In Australia a drug needs to be proved to be cost-effective before being reimbursed by the government and hence listed on the pharmaceutical benefits scheme (PBS) and affordable to patients. Once a drug is on the market further trials, commonly called phase IV can be undertaken for a variety of reasons. For example, treating special populations, assessing how doctors are prescribing the product, comparison with other products and so on. Data on the safety of the product is also being continually collected. A similar program is followed for medical devices and other therapies provided by the biotechnology sector. The statistical skills required by those working on clinical trial programs are the focus of this discussion paper.

CLINICAL TRIAL PROGRAM

For any clinical trial there are many functions represented. There are the medical experts, clinical research associates, monitors, data management staff, statisticians, regulatory affairs and drug safety associates and medical writers who may or may not be part of the pharmaceutical company (sponsor). These people develop the protocol, design the data collection forms, manage the data management process, monitor any adverse drug reactions, analyse the results, write the clinical study report and combine the information with the other available information on that product for submission to the regulatory authorities. It is mandatory that a pharmaceutical, device or biotechnology company can prove that the specialists involved in any product development are suitably qualified for the role that they played. This means that apart from undertaking the statistical analyses it is expected that a statistician is involved in and approves the protocol, the data collection form, the database structure and the clinical study report.

It is becoming more common for health economists to be part of the clinical trial program but generally they are more often brought in once registration has been approved and reimbursement is sought. These specialists are usually keen to work with statisticians as they are required to undertake quite a lot of modelling-type activities.

In addition there are the doctors, some of whom may participate in the development of the protocol, all of whom are involved in recruiting patients for the study and interested in the results, the study nurses and, of course the patients. From this varied list of participants it is not surprising that there is a wide range of exposure to statistical thinking and methods, and, even though some participants can fulfil their role without any statistical literacy some understanding, such as the need for 'good' and 'complete' data could assist the research.

STATISTICAL ADVICE AVAILABLE VIA GUIDANCE DOCUMENTS

The drug development industry is, not surprisingly one of the most regulated industries in the world and is driven by the three regions of Europe, United States and Japan. About 10 years ago the International Conference on Harmonisation (ICH) spear-headed the amalgamation or development, and continued development of global guidelines that the various regulatory authorities agree to (www.ich.org). These guidelines are in four categories, quality, safety, efficacy and multidisciplinary. All of them contain advice around the handling of statistical issues. For example, ICH E9 Statistical Principles for clinical trials, ICH E10 Choice of control group and related issues in clinical trials and ICH E12 Principles for clinical evaluation of new antihypertensive drugs. One might think that with all these guidance documents available the need for statisticians and statistical thinking has been removed. This is not the case, and in some ways the guidance documents help the statistician work with other functions as the issues have been raised for them.

TYPES OF GRADUATES

A large variety of graduates find positions within the pharmaceutical industry. These include those with medical qualifications such as doctors, nurses and pharmacists, other scientists including chemists, biologists, statisticians, computer programmers and so on. In addition to the variety of backgrounds is the variety of levels from certificates and diplomas to PhDs and other post-graduate qualifications. This variety results in a large range of exposure to statistical education both during their training and once they are in the workforce.

STATISTICAL EDUCATION FOR STATISTICIANS

There is a shortage of statisticians within the pharmaceutical industry world wide and indeed, in the health industry. The Biostatistics Collaboration of Australia (www.bca.org.au) was established to go part of the way to addressing this need.

Apart from a sound knowledge of statistical theory, including the designs and analyses common in clinical trials it is also helpful if graduates have a good understanding of issues that may affect the validity of the analysis methods. For example, what impacts do missing data, bias, multiplicity, randomisation schemes or lack there-of, main and interaction effects, sub-group analyses and extrapolation have on the generalisability of results? In addition graduates need to have good problem solving, questioning and listening skills, they need to be comfortable working to tight deadlines, working as an integral part of a cross-functional team and able to explain statistical concepts to non-statisticians, have good writing skills and good programming skills.

There are two main career paths a statistical graduate can follow within the pharmaceutical industry. One is as a statistician and the other is as a statistical programmer. Both have a general minimum education requirement of a Master's level qualification that has a very strong component of statistics, although programmers may be sourced with less statistics and stronger programming skills. Both paths take you through the entry level position, a more experienced level, a senior level, a Principal type level and then into management.

With the statistician route you will start under supervision but become responsible for providing the statistical input to the trial designs, sample size calculations and protocols, writing statistical analysis plans, working with the programmer(s) to develop the tables and listings that present the results of the analyses and the medical writer to develop the clinical study report. If you are in a large pharmaceutical company you may become a product life-cycle team member where you become and expert on a particular product and stay involved throughout the development, marketing and final decisions to withdraw a product should such a decision be required.

If the programming route is chosen, again you will start under supervision but will become responsible for providing advice on displaying the data, programming the outputs, supervising teams of programmers, developing macros to increase the efficiency of the programming team, work with the data management group on designing databases, extracting data and generating complex data queries.

Once in the industry there is a need for continued education in statistical methodology. At present this need in Australia appears to be filled by short courses generally provided by local or visiting academics through clinical trial units and medical schools attached to universities. This activity should continue to be encouraged. Internationally these types of courses may be offered specifically through the statistical association(s) specifically for pharmaceutical statisticians.

STATISTICAL EDUCATION FOR NON-STATISTICIANS

The other players in the drug development team require different levels of statistical literacy depending on the role they are playing. Many of those who have at least a degree level qualification will have obtained some basic statistical education and will quite happily admit to not understanding statistics or that it was too long ago for them to remember. Those with only certificate or diploma qualifications may not have had any exposure to statistical education. Some, like the health economists, obtain their additional statistical education via further educational qualifications.

Within the clinical trial arena the medics or clinical members involved in designing studies need a general knowledge of designs, sample size estimates and how they change with continuous or binary data, issues such as bias, missing data and multiplicity. They also need to know how to respond to any queries they receive from any ethics committees reviewing the protocol. They will obviously gain this over time, particularly for their therapeutic area but some basic knowledge would be a helpful starting point. The clinical research associates or monitors collecting and verifying the data from the sites and the data management team need an understanding of the implications of poor data both for data collection and for focusing on data queries. The medical writer requires quite sophisticated statistical knowledge as they need to understand the design, any discussion of data issues and the description of the results that the statistician provides. Sales and marketing people could also benefit by an understanding of how to interpret results and where they must draw the line as they are using the results in published papers to convince doctors to prescribe.

Apart from their specific requirements as mentioned above, all of these people read the medical literature and a common request is that they would like training to help them assess the quality of published papers, including the statistical methods! There appears to be no realisation that even for a statistician this skill is built over time.

Can the cross-discipline training in statistics be improved for those undertaking degree qualifications, maybe by the use of more practical examples? Who should provide or how can statistical skills be obtained or improved once you are in the workplace? The first question is one that needs to be investigated by statistical departments within universities that offer qualifications in subjects that may lead to work within the pharmaceutical or broader health arenas. The second question is a bit more complex in some ways. The author's view is that the offerings are somewhat haphazard in that short courses, for example on introductory biostatistics may be advertised by a university one year and then not for a few years. Overseas there are a number of companies that provide a variety of training, including statistics specifically for the pharmaceutical industry.

In order to address the possible gap in the industry in Australia for professionals in the product development area the author and a colleague recently integrated various training courses they had provided in-house and externally to pharmaceutical companies into what is currently called a foundations course (2 days), a fundamentals course (1 day) and an advanced course (2 days). All courses specifically focus on the needs of the industry in Australia. The two day courses are deliberately small and provide a mix of theory, practice and discussion, particularly for attendees to share their experiences with clinical trials. The discussion sessions have proved extremely popular as they have allowed the attendees to synthesise their experiences with what they have learnt or re-learnt in the course. The aim is to provide these courses on a regular basis so that those in the industry can plan their training requirements in advance.

As the majority of pharmaceutical companies in Australia do not have an in-house statistician it is the experience of the author that at this stage the only training that sales forces get in understanding statistics and the results published in the literature is provided in-house by someone in the medical department. Those few companies who have in-house statisticians will ideally receive some, usually very brief training from one. These are then the people who are trying to convince the health professionals, who have their own varied levels of understanding of statistical concepts to prescribe their product! And so the training needs expand.

SUMMARY

The pharmaceutical industry is a large employer of staff requiring varying levels of statistical literacy. Drug or new product development takes many years and is extremely costly. Although the industry is highly regulated with many guidelines approved by the regulatory authorities these do not obviate the need for skilled thinkers and, in particular, thinkers aware of the implications on the development and subsequent sales of a drug of issues that come under the umbrella of statistical thinking such as bias, missing data and generalisability. It is also an industry that recognises the need for specific skills with it being mandatory that companies show that their staff qualified to be in their role both academically and by experience.

There is a recognised need for statistical knowledge for many of the roles within the industry but it also needs to be recognised that staff come from a wide range of backgrounds with varying levels of exposure to statistical knowledge. So, there is a need for a variety of training in statistics including the interdisciplinary courses that currently occur between statistics departments and other science departments, short one- and two-day courses aimed at non-statisticians on a variety of statistical concepts and short courses aimed at statisticians in specific areas.

REFERENCES

Bain & Company press release, 8 December 2003 Has the Pharmaceutical Blockbuster Model Gone Bust?

UNIT STANDARD ASSESSMENT INVOLVING OFFICIAL STATISTICS AN EVALUATION OF FOUR ASSESSMENT TOOLS

Murray Black
Auckland University of Technology
Auckland NZ

INTRODUCTION

In the second half of 2007, a Certificate in Public Sector Services in Official Statistics was introduced in New Zealand as a qualification designed for people working in the State Sector who need to use official statistics as part of their job role. This certificate consists of 40 credits of unit standards with 24 credits being allocated to those standards which involve applications of statistics. These 24 credits are made up by four core unit standards which are assessed separately. They cover the skills and knowledge required for interpreting and using official statistics to form conclusions and make policy recommendations in a State Sector context. The assessments are based on a series of performance criteria within each unit standard. The unit standards are:

- *US 23268: Level 4, 8 credits.* Interpret statistical information to form conclusions for projects in a state sector context**.**
- *US 23269: Level 5, 8 credits*. Evaluate and use statistical information to make policy recommendations in a state sector context.
- *US 23270: Level 4, 4 credits* Assess a sample survey and evaluate inferences in a state sector context.
- *US 23271: Level 5, 4 credits* Resolve ethical and legal issues in the collection and use of data in a state sector context.

A report involving official statistics, for example, the Innovation Report is selected beforehand and used as an exemplar for the assessments. To obtain credit for each standard, the candidate is required to answer all the questions pertaining to the selected report for the assessment correctly. These questions were developed as a means of determining whether or not the standard set by the performance criteria (PC) has been met. In the case of an answer being incorrect or not fully answered, the candidate was offered a re-sit where the candidate would review their answer(s) and resubmit.

Questions were broad ranging and incorporated the following areas;

- assessing the relevance of data
- finding and selecting data relevant to a policy question
- interpreting findings
- making policy recommendations based on the data
- explaining how a particular piece of statistics could be performed with a possible result in the context of the report. For instance, where you would calculate a confidence interval for the difference between means and how the results would be interpreted
- interpreting possible results and designing a data collection to answer a policy or research question , and
- explaining the limitations of a chosen exemplar, for instance stating the omission of a margin of error in the report along with possible consequences.

My paper will evaluate different types of candidate's responses that were required to answer these various types of questions which required the candidate to display statistical reasoning skills.

The objective of this evaluation is to improve these assessment tools for the next cohort of candidates.

LITERATURE REVIEW

Cognitive Domain of Bloom's Taxonomy, Instructional Domains, Literacy, Reasoning and Thinking

This literature review attempts to map a collection of statistical reasoning skills required in the answering of the questions into taxonomy, instructional domain and reasoning frameworks. These questions were not designed originally around these frameworks rather in line with a list of topics that were deduced from the statements contained within the original unit standards.

A domain of educational activity that can be associated with the requirements of these four assessment tools is the cognitive domain as identified by Bloom (1956). This domain involves knowledge and the development of skills and has six major categories (Bloom 1956). We wish to map from five of these six categories, denoting degrees of difficulty, into the instructional domains as defined by (DelMas 2002) which pertain to descriptions of tasks and then into five levels of statistical reasoning as proposed by Garfield (2002). These mappings in hierarchical order are shown in the table below:

Level	**Bloom's Taxonomy Objective**	**Instructional Domains Teaching**	**Reasoning Framework Assessment**
1	**Knowledge** Recall	**Literacy** Identify	**Idiosyncratic** Knows
2	**Comprehension** Meaning	**Literacy** Describes	**Verbal** Defines
3	**Application** Context	**Reasoning** Why?	**Transitional** Partial Understanding
4	**Analysis** Distinguishes	**Reasoning** How?	**Procedure** Application
5	**Synthesis** Contextual Links	**Thinking** Apply	**Integrated Process** Complete Understanding

Table 1 Example of the five levels of statistical reasoning as proposed by Garfield (2002) applied to an assessment

Although not designed to be such, answers in a typical assessment of a unit standard could be used to illustrate these five levels using mean and standard deviation as follows**:**

1. *Idiosyncratic*: Knowing that the mean and standard deviation are used in statistics but not being able to fully appreciate their meanings.
2. *Verbal*: Being able to define both the mean and standard deviation correctly but with no context.
3. *Transitional*: Being able to define both the mean and standard deviation correctly in context to the report.
4. *Procedural*: Being able to explain how both the mean and standard deviation apply to the objective of the report but unable to integrate their meanings into all parts of the report where relevant.

5. *Integrated Process*: Can integrate the meanings of the mean and standard deviation into all parts including the objective of the report where relevant. For instance their application in the use of confidence intervals and margins of error.

EXAMPLE

Segment of five assessment questions categorised by the five levels of statistical reasoning as proposed by Garfield (2002)in hierarchical order.

You have been supplied with a copy of the reports 'Innovation in New Zealand'

<u>Policy question</u>: The Minister for Economic Development has asked:
What proportion of business expenditure is either on research and development or on innovation?

Element 1:
Assess the relevance of the reports (data collections) to the given policy questions.
That is, for the policy question answer the questions below:

1. Identify the population that data is required for? *Level 1*
2. Describe a "margin of error"? *Level 2*
3. If a sampling procedure has been used, describe its main features. OR
   If a sampling procedure has not been used, describe any limits of the data with respect to the objectives! *Level 3*
4. What topics (objectives) does the question relate to? *Level 4*
5. Apply the objectives of the supplied data collection(s) to the policy question(s). Does it fit? Justify your answer. *Level 5*

CHARACTERISTICS OF THE ASSESSMENTS

PURPOSE OF QUESTIONS

The key objective in the various assessments was to put the candidates in the position of having to read and interpret reports bearing in mind the overall objective of the report and how the statistics within the report informed the various answers to policy questions.

A candidate's competency was to be assessed across the following learning areas, all having a state sector context embedded in the following four unit standards:

Standard	**Learning Area**
US 23268(US 68):	Use of descriptive statistics, graphs and measures to describe data
US 23269(US 69):	Use of statistics in constructing policy recommendations
US 23270(US 70):	Use of statistics to make inferences
US 23271(US 71):	Legal and ethical issues

These four standards cover data collection, policy issues, legal and ethical issues, reading and interpreting the statistics all linking to an overall objective used in reports having a state sector context.

The questions were designed originally from the unit standard statements. In the case of unit standards 23268 and 23270 a list of topics was developed then questions were written around those topics to ensure full coverage. For unit standards 23269 and 23271 the questions were written around each performance criteria (PC).

Competency was assessed as either pass or fail for each unit standard individually.

CONTENT OF THE ASSESSMENTS

Two standards were at level 4 and the other two at level 5. These levels represent the final year secondary school and first year undergraduate degree levels respectively. It was difficult to assess concepts that were not in the report exemplars. In some cases candidates were asked to invert examples where the concept could have been used in the report. Some further teaching was required by email to ensure candidates were given a good chance to answer these questions. Eg Confidence intervals for differences between proportions or means. This had the effect of fragmenting the content when it needed to be linked to the other concepts in the body of the report.

Overall the content was well covered with a question pertaining to each performance criteria. All assessments based on the same report(s). Linkages made with examples given with similar reports in the teaching sessions.

TYPE OF QUESTIONS

The tables below categorize the required answer for each question from my assessors point of view using a scale of 1 to 5 (DelMas 2002) depending on the level of reasoning (Garfield 2002) being tested. These levels correspond to table 1.

1 = *Identify* (*pick out answer* from the report).
2 = *Describe* with *no* requirement to answer in *context*
3 = *Why* this? – answer requires *context* to the report in *specified part* only
4 = *How*? – some link to *more than one part of report* is required in *context*
5 = *Apply – all links to relevant parts of report* are required in *context.*

US 23268	Level of Reasoning Category	US 23269	Level of Reasoning Category
1	1	1	5
2	1	2	1
3	1	3	3
4	1	4	1
5	2	5	4
6	3	6	4
7	3	7	3
8	3	8	4
9	3	9	5
10	3	10	3
11	3	11	5
12	2	12	3
13	4	13	3
14	4	14	4
15	4	15	4
16	2	16	3
17	2	17	5
18	2	18	5

Note: In US 68, ordering is affected by last three questions being classified at level 2.

US 23270	Level of Reasoning Category	US 23271	Level of Reasoning Category
1	1	1	2
2	1	2	2
3	3	3	3
4	2	4	3
5	3	5	3
6	3	6	3
7	2	7	3
8	2	8	2
9	3	9	3
10	2	10	3
11	2	11	3
12	2	12	3
13	2	13	4
14	2	14	4
15	4	15	5
16	2	16	5
17	5		
18	3		
19	3		

We can tabulate the number of questions occurring at each level as follows:

LEVEL	US 68	US 69	US 70	US 71
1	4	2	2	0
2	5	0	9	3
3	6	6	6	9
4	3	5	1	2
5	0	5	1	2
TOTAL:	18	18	19	16

If we collapse categories 1 & 2 and 4 & 5 and calculate percentages of the total assessment in each case we get:

LEVELS	US 68	US 69	US 70	US 71
1&2	50%	11%	58%	19%
3	33%	33%	32%	56%
4&5	17%	56%	10%	25%
TOTAL:	100%	100%	100%	100%

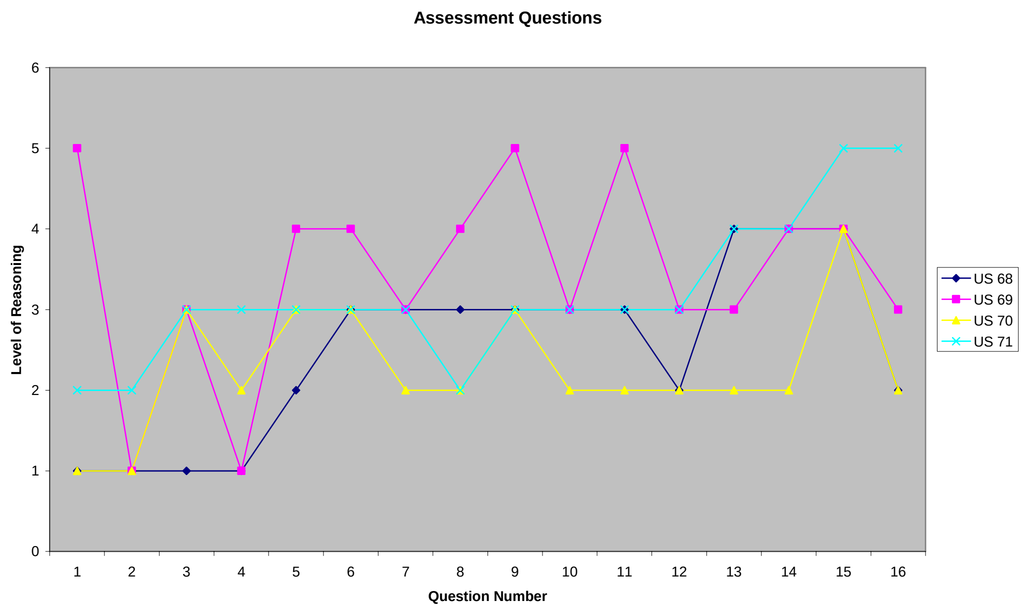
We observe that US 68 and US 70 have considerably more questions in the lowest two levels of reasoning than US 69 and US 71. Conversely US 69 has over half its questions in the highest two levels of reasoning at 56%.

If we run a Chi-Squared test at 5% level of significance we would conclude that level and unit standard are significantly related ($\chi^2 = 17.7 > 12.592$ at $\alpha = 0.05$ and $\nu = 6$). So irrespective of how the questions were designed we have significant differences in the proportions of questions containing the various levels of reasoning over these four unit standards.

ORDER OF QUESTIONS

Originally the assessments were designed so the content was tested in the order it appeared in the performance criteria pertaining to each unit standard.

The graph below gives the levels of reasoning by question order for each unit standard where the line graphs follow the order of questions in the assessment.



Note that the order of questions for assessing US 23271 followed the levels of reasoning the best, low to high, by starting off at 2 and rising to finish at 5.

CANDIDATES RESPONSES TO THE ASSESSMENTS

CANDIDATES ANSWERS TO QUESTIONS

On the whole candidates answered most questions correctly first time however there was a high percentage that required a re-sit on at least one question as the following table shows:

Percentage of Candidates Passing requiring a Re-sit	
US 23268	**77%**
US 23269	**71%**
US23270	**55%**
US23271	**56%**

Re-sit ones – common issues

1. Unable to find example in report e.g. Bivariate analyses,
2. Lack of teaching coverage of some concepts e.g. Cyclical variation in Time Series, performance indicators
3. Questions not answered properly giving what's required e.g. Privacy principles
4. Failure to provide appropriate context when required
5. Answers too brief and not enough detail given e.g. recommendations to management
6. Only one part of question answered e.g. the candidates were asked to distinguish between stratification and clustering but were unable to explain why stratification was preferred in the report.
7. Difficulties in explaining concepts not covered in report E.g. Confidential intervals for differences between proportions. This was largely due to the design where all the concepts had to be related to a maximum of two reports for each assessment.

8. Difficulties more with the two quantitative assessments as opposed to the two qualitative assessments.

All questions and re-sits were handled by email.

DID THE CANDIDATES ACHIEVE WHAT WE WANTED?

There was a group that managed to complete each unit standard in time and keep up. Many asked for extensions from the three weeks which wasn't rigidly enforced by me. There were however some long periods between re-sits and submissions. The table below provides some summary statistics:

UNIT STAND ARD	NO SENT OUT	NO OF RE-SITS	NO OF PASSES	% PASS RATE AT JUNE 11TH 2008	MEDIAN MONTH S	RANGE MONTHS
23270	**13**	**6**	**11**	**85**	**1.25**	**0.25 TO 10.00**
23268	**13**	**10**	**13**	**100**	**1.25**	**0.50 TO 9.00**
23269	**12**	**5**	**7**	**58**	**1**	**0.50 TO 4.25**
23271	**12**	**5**	**9**	**75**	**1.75**	**0.25 TO 6.50**

Note that some of the activity around 23269 and 23271 is current and these standards were presented one month apart.

BETWEEN EACH US ASSESSMENT WAS LEARNING TRANSFERRED?

No mainly because questions didn't require this, initial questions were designed to focus candidates, 69 and 71 were designed to fit under the performance criteria whereas 68 and 70 didn't do that as they focussed on content.

PRIOR STATISTICAL KNOWLEDGE

Trail cohort was really divided into two groups due to background knowledge. Assessment was difficult for the weaker group with re-sit questions required in many instances. In some cases a verbal discussion was more effective in clearing up any outstanding issues.

COMPLETION

Help towards completion was provided by timely feedback re re-sit questions to the point, conversations with candidates re re-sits topics not understood i.e. cyclical variation, index numbers and odds ratios covered with a session with candidates face to face. Feedback was provided to the contact at Stats NZ for tutorials. The identified barriers to completion were:

1. Allocated time of three weeks to complete each assessment not enough
2. Not enough time between the presentation of each unit standard
3. Some concepts in the assessments not being taught fully enough
4. Not enough teaching provided for weaker candidates
5. Reluctance of candidates to request assistance in the form of hints
6. Fragmentation between the four assessment tools
7. Absence of a clear example in the report
8. Difficulties in "making up an example" to illustrate a concept.
9. Work commitments
10. Effect of instruction diminished over time

REDEVELOPMENT OF THE ASSESSMENTS

After the trial cohort we are now into 2008 with a new group of candidates. Efforts are underway to build on past experiences of using these assessment tools and to review and improve them by considering the following:

1. Fewer questions in line with the five stages of statistical reasoning 1to 5 in sequence.
2. No overlaps in questions requiring similar answers between unit standard assessments.
3. Linkages in assessing components of all four standards across a set of questions on the one assessment. Policy question, data collection, descriptive statistics, inferences with confidence intervals then recommendations.
4. Look at assessing some components as part of the teaching. This happened with the pilot in two of the assessments.
5. Production of exemplars showing what's required for a pass for different types of questions.
6. Aim to duplicate workplace environment in conducting these assessments with workplace verification where warranted.
7. Ensure that performance criteria start with the objective of answering a policy question.
8. Introduce a retention component in the assessment to see if learning is maintained over a time period since passing the unit standard.
9. Examine questions that are applicable to the chosen report so not necessary to have 100% coverage of all content to get a pass, however choose reports so coverage of content isn't always the same. Recommend 85% minimum coverage for all assessed unit standards.
10. Remove the requirement of candidates having to invent an answer as there is nothing in the report to observe and comment on for the answer.
11. Explore working in groups however would need to ensure equal contributions.
12. Extend backup and mentoring systems for candidates.

REFERENCES

Bloom, B.S. (1956), 'Taxonomy of Educational Objectives: The Classification of Educational Goals', *Handbook I; Cognitive Domain*, New York, USA: Longmans, Green and Co.

DelMas, R.C. (2002), 'Statistical Literacy, Reasoning and Learning: A Commentary', *Statistical Education*, 10(3).

Garfield, J. (2002), 'The Challenge of Developing Statistical Reasoning', *Journal of Statistics Education*, 10(3).

Contributed Paper (Refereed) – Peter Martin

ASSESSMENT OF PARTICIPANTS IN AN INDUSTRIAL TRAINING PROGRAM

MARTIN, Peter
University of Ballarat

Abstract
The training of industrial or business personnel in using various statistical tools to enhance quality control programs presents a challenge to all concerned. The trainer needs to be familiar with the features underlying adult learning as well as the workplace context within which the training will apply. The training material needs to be relevant to participants' work practices if commitment is to be achieved. In addition, there seems to be an increasing interest in assessing the knowledge and skills of the trainees participating in such programs. As such, the traditional forms of assessment, such as written assignments and examinations, are of little practical use in settings where the primary focus is upon using the tools to improve processes to save dollars. This presentation will describe the procedures used to assess trainees who recently participated in such a training program, specifically with respect to their participation, knowledge gained and application.

INTRODUCTION

The teaching of statistical techniques to people in industry, as part of quality control or process improvement programs, can be a rewarding but often daunting process for academics. The evaluation of such a training program and its outcomes must take into account the objectives of the organisation, as well as the personal needs of all the participants. The new knowledge and the skills imparted by the training program must relate to real workplace needs.

In order to understand how we might assess a training program, it is important, firstly, to understand a little about how a successful training program might be developed. Such programs typically involve training in skills that can be used immediately to optimise some processes or procedures. Industrial training typically implies a distinct end or aim which, in itself, guides the facilitators and instructors.

From an educational perspective, Ramsden (1992) and Biggs and Moore (1993) argue that successful training programs invariably promote a deep approach to learning by relating existing knowledge to a project in hand, or drawing on knowledge from as many sources as possible via project teams. When statistical theory is placed into the realm of work place experience, deep learning will be enhanced. The transformation of theory into practice establishes a meaningful context for the user thereby enabling better understanding and appreciation of a situation.

Werner and Bower (1995) emphasised the need to focus upon participant learning rather than teaching style. As adults we tend to learn best when support is provided for our own personal motivation; when our experiences are valued; when we are encouraged to participate; and when the training material delivered is perceived as being relevant to our daily work (Pretty, et al., 1995). Self-directed learning opportunities and interactive learning environments will shift the focus from the style of the trainer or teacher to the trainee's learning. Therefore it is important to match participant competencies with the needs of the organisation.

THE WARRNAMBOOL PROGRAM

Primary objectives of this program were to build the capability to better utilise statistical tools on a day-to-day basis, to utilise best practice methods in technically based projects and for performing relevant research. The routine use of statistics within the company was recognised as an essential core capability to be systematically developed throughout the company. It was believed that the successful utilisation of statistical tools throughout the company would help secure its future in an increasingly dynamic and competitive national and global industry. The program was to provide staff with a suite of best practice statistical tools for enhancing their functions within the company.

The program was tailored to the needs and roles of individuals within the company. In particular it was to provide the background, skills and tools needed to better use data to improve business performance whether in operations, R&D, marketing, finance, sales, accounting, laboratory, or anywhere else. This was achieved by taking into consideration various aspects relating to organisational needs:

- the range of entry levels into the program in terms of personal capabilities and interest;
- the range of capability levels in terms of level of awareness and applications/use of statistics;
- the need for the training material to be practically based involving parallel application on projects of relevance to an individual's job function in order to enable immediate application;
- the need to have ongoing support by, and access to the trainer for key people to effectively utilise the knowledge and skills developed.

To this end the final program consisted of 15 modules (see Table 1) delivered to 3 groups of participants in half-day sessions of three to four hours. Each session was delivered by the same presenter (the author) and comprised a lecture component, as well as a practical component and group work where necessary. Participants were expected to complete set tasks after each session involving computer worksheets requiring analysis of specific data sets, collection and analysis of workplace data, and/or group presentations of analyses carried out during the training sessions. All participants were provided with a comprehensive set of training notes and web-based reference readings, as well as self-paced computer worksheets involving analysis of work-based data sets using Minitab.

The allocation of staff to these groups was based largely on the perceived extent and depth of statistics to be used by that person in performing their business function. For example, someone in the R&D Group who uses statistics daily was required to attend all training sessions. For those in Group C who only occasionally used statistics for reporting purposes to make, support or question decisions based on data analysis, etc., attendance at the first two sessions only was required. Group A consisted of 8 participants mainly drawn from R&D and laboratories. Group B consisted of 11 participants, mainly comprising process supervisors and people from the production line. Finally, Group C consisted of 9 participants from middle management positions.

Table 1: List of Training Modules and the Groups to Receive the Training

Module	*Groups*		*Module*	*Groups*
Intro & Overview	A+B+C		Correlation & Regression	A
Process Characterisation	A+B+C		Hypothesis Testing & CI's	A
Exploratory Data Analysis	A+B		ANOVA	A
SPC – Basics	A+B		Experimental Design I	A
Gauge R&R	A+B		Experimental Design II	A
Multi-Vari Studies	A+B		SPC – Advanced	A
Process Capability	A+B			

All participants were required to apply their new knowledge and skills, gained throughout the course, to projects drawn from their everyday work in order to secure their understanding and appreciation of statistics in their workplace. These workplace projects were seen as the keystone of the program. The PPDAC cycle (problem, plan, data, analysis, conclusions) developed by Wild

& Pfannkuch (1999) to model statistical thinking was clearly evident. At each step of the cycle participants were required to present a report to the group responsible for overseeing their particular project. This resulted in an ever increasing appreciation of the importance of interpretation and communication of findings for each participant, as noted by Forster, Smith & Wild (2005) on the benefits of getting students to write about statistics. This enhanced appreciation was reflected in some of the participants' comments, as indicated below. These comments also showed an appreciation of the learning of good statistical practice, discussed by Svennson (2007), particularly in the context of the projects in which they were involved.

- *I found it very useful, and feel confident to use the tools to get what we want, particularly GR&R, EDA & Normal Dist. The material was useful & valuable resource, but needs project running alongside.*
- *Value for $; gets to the nuts & bolts; enables you to see differences between issues that do & don't count; useful to apply in project*
- *Changing work behaviour – now see need for planning what data I need to collect; see real need for EDA, GR&R, Capability & ANOVA;*

THE ASSESSMENT PROGRAM

Upon successful completion of the training program and individual projects, Group A and B members were to be assessed, and if successful, awarded two certificates: a certificate of participation and a certificate of competence. This assessment took place 9 months after the completion of the training to allow for the finalisation of projects. The nature and design of the assessment was determined by the trainer in co-operation with the company program organisers.

The aim was for the assessment to be seen as fair, and relevant to the purposes of the training program. It was deemed important that the assessment measure both procedural and conceptual understanding. While it was important for the participants to be able to correctly perform a task, it was considered equally important that they have some knowledge of what was being done, and why. Lipson (2007), and Garfield, delMas & Chance, (2003) have discussed the importance of such aspects in assessing statistical knowledge, admittedly in the context of tertiary students, the same certainly applies for workplace based training involving statistics. In fact, the very applied nature of the workplace context probably makes it easier for this type of assessment to be applied.

A *certificate of participation* was awarded to participants who satisfied the following requirements:

- attendance at 75% of the training sessions;
- satisfactory completion of 80% of the practical worksheets involving computer analysis of specific data sets;
- active involvement in at least one of the projects designed to accompany the training program;
- participation in the preparation and presentation of project reports at various stages throughout the life of the project.

All trainees except one satisfied the above requirements and received certificates of participation signed off by company management and the University of Ballarat.

Assessment of the training material for each topic required participants to submit practical computer worksheets involving analysis of specific data sets using Minitab and/or Excel, and to collect and analyse some data obtained from their particular work environment. This often involved individual and/or group presentations of analyses carried out during the training sessions, or more formal preparations of written reports based upon results of experimental studies carried out at the work-place. Assessment of the worksheets was graded as either satisfactory or not satisfactory, and was included as participation as it was an indicator of involvement and commitment in addition to knowledge gained.

Assessment for the *certificate of competency* was only undertaken by request, and took the form of a 45 minute, one-on-one "interview" with the trainer; in effect an oral examination. Whilst this was recognised as a potentially daunting process for the participants, it was mitigated to some extent in that both trainer and participants had known each other over a period of two to three years, and that each participant was provided with an outline of the types of questions and topics to be covered. Each "interview" was structured to account for the differing levels of experience of the participants by selecting appropriate topics related directly to their work-related responsibilities. The procedure that was followed was the same in each case, and is outlined below. During the "interview" participants were required to:

- speak to a topic of their choosing as well as one selected by the interviewer;
- answer various questions about any of the topics covered;
- interpret computer output in the context of the workplace;
- use Minitab to perform various analytical tasks given some workplace data.

An outline of the types of questions asked and the topics to be covered is provided in Table 2 below. Each participant was provided with a copy of this same outline and was expected to provide a satisfactory response to at least one question from each topic covered during the training program. No access to any references was permitted during the "interview".

Table 2: Sample Questions for Competency Certification

- The following list of topics includes those covered in the training program. Select one and tell me what you know about it.

Levels A&B Topics	Level A Topics
Project Plan	*t-tests*
Process Map	*Paired & Independent t-tests*
Exploratory Data Analysis (EDA)	*Correlation & Regression*
Measurement Systems Analysis (Gauge R&R)	*Chi-Square Test*
Capability Analysis	*ANOVA & Experimental Design*

- What is the purpose of . . .
- Describe how you would do . . .
- Outline a situation at work where it might be appropriate to do . . .

  – *EDA* (A & B)
  – *Gauge R&R* (A & B)
  – *Capability Study* (A & B)
  – *Histogram* (A & B)
  – *Pareto Chart* (A & B)
  – *Time Series Plot* (A & B)

- Give me an example that shows the difference between categorical data and continuous data.
- Give me an example of when it might be appropriate to do a time series analysis

- When would you do . . .
- What are you looking for when you do . . .

  – Paired t-test (A)
  – ANOVA (A)
  – Chi-Square Test (A)
  – Correlation/Regression analysis (A)

- Here is some Minitab output from an analysis of data collected from on-site.
  – What statistical tool is this an example of?
  – What does it tell you?
  – When might you use a tool such as this?

- Here is some data. Use MINITAB to obtain
  - *descriptive stats*
  - *boxplots*
  - *histogram*
  - *normality test*
  - *scatterplot*

The questions asked of a participant were selected according to the designated level of that participant and the results of each "interview" were summarised on a recording sheet. Table 3 below shows an example of the recording sheet used with results for 2 of the participants from each of Groups A and B. A star rating of one-to-three was used to indicate the extent of each participant’s knowledge and understanding of the topics discussed (see Table 3). There was also provision on the recording sheet for comments by the trainer regarding individual performances.

DISCUSSION

From the company’s perspective, evaluation of the overall program was always going to be based primarily upon the successful completion of the projects allocated at the commencement of the training, and in the savings or efficiencies gained by the improvements made therein. While such successes are an important factor in the continuing successful application of the initial training received, they also play a vital role in maintaining a competitive edge in national and global markets. Any researcher will be buoyed by initial success and will most likely be better equipped to continue with the experimental process. Evidence of this continuation for these participants can be seen in the wide-ranging new projects that have been completed in the two years post-training.

Table 3: Recording Summary of Individual Assessment Details:

		Gp A (Initials)		*Gp B (Initials)*	
General Topic	*Comments*	*AD*	*BC*	*CH*	*MS*
Familiar Topic	Self Choice Trainer Choice	*CapEDA*	*GREDA*	*ANVNest*	*GRRRes*
Exploratory Data Analysis	Where is categorical data used?		**		
	Where is continuous data used?				
	What does the stdev measure?	**			***
	What does the mean measure?		***		
	How do we know if the data is normal?	***		***	
	Why do we want to know if the data is normal?		*		***
	What’s a histogram? Boxplot? Bar Graph?	**	**		*** ***
Hypothesis Tests & CI’s	When would you use a single sample t-test?			**	
	When would you use a two-sample t-test?				*
	What does a 95% CI tell you?				**
	What test would I use if …?				
	What is a p-value?			**	
** - needed help; ** - good working knowledge; *** - strong understanding*					
Minitab Tasks	**** - unassisted; ** - needed some help*				
	Descriptive Stats (Milkfill by Machine)	***	***		

	Boxplots of Moisture by Pallet	***	***	***	***
	Pareto (Stoppages)	***	**		
	1-Way ANOVA (MilkFill by Machine)			***+	**
	Extra comments by Trainer	Strong u/standing & appreciation of tools; realises power of applying tools; awkward expression at times.	V.good u/standing & appreciation of tools; quiet confidence in use; good role model/mentor potential.	Good u/standing & appreciation of tools; needs more application egs to increase awareness of application	V.Good u/standing & appreciation of tools; excellent attitude; keen to apply where applicable; excellent mentor

From the participants' perspective, we get some idea of their perception of the success or otherwise of the program from their comments made on evaluation surveys completed at various stages during the program, and also by their continuing involvement and use of the tools in projects post-training. As shown below, not all their comments were positive; and neither have all continued their involvement post-training.

- *Useful to apply in projects, but having no computer made it difficult*
- *Didn't know objectives well enough to find a suitable proj; can see the value, but felt a bit left behind; need a slower pace;*
- *Great value personally; appropriate content, excellent reference when reviewing post training. The A/B break-up was very good & good timetabling of 1/2 day sessions.*

Such comments highlight the importance of the initial planning and structuring of such a program, as well as the relevance of the projects, and the need for provision of adequate facilities for use by the participants. Their critical comments could generally be classified as "constructive", or "destructive" as shown by the two comments below:

- *Project needs to done during sessions;*
- *Don't use tools in my role; not worth the $;*

In all cases the "destructive" comments came from the Group B participants and tended to avoid taking personal responsibility for difficulties encountered during the program (there was a tendency to blame everybody and everything else). The constructive comments came equally from both groups, tended to realise the value of the program and offered suggestions for enhancing the program.

From the trainer's perspective the overall success of the program was evidenced by the pride with which the participants received their awards, the reports of cost-savings and efficiencies resulting from the successful completion of projects, and in continued involvement to this day in post training project work, in both an advisory and consultative capacity.

CONCLUSION

The assessment of the statistical knowledge gained from this program had to be taken outside the traditional format of only testing procedural understanding. Tasks that assessed both procedural and conceptual knowledge were required almost by default for the program to "fit the workplace" in a meaningful and cost effective way. Whilst the assessment program was demanding and time consuming for all, and probably not appropriate for large groups of university students, there were some aspects that might be able to be adopted for small, interest based university classes. Such instances might include, for example, small, one-off classes for research students, or other faculties, occupational health and safety programs, post-graduate programs, etc. It was strongly felt by all concerned that the certificates had been well earned by the participants; the presence of both company and university logos on the certificates lending further credibility to their worth.

REFERENCES

Biggs, J. B., & Moore, P. J. (1993). Process of Learning, 3rd edition. Sydney: Prentice Hall.

Forster, M., Smith, D.P., & Wild, C.J. (2005). Teaching students to write about statistics. In L. Weldon and B. Phillips (Eds) Proceedings of the IASE/ISI Satellite Conference on Statistics Education, CD-ROM and online http://www.stat.auckland.ac.nz/~iase/publications.php

Garfield, J., delMas, R., & Chance, B. (2003). Assessment Resource Tools for Improving Statistical Thinking . Retrieved from https://app.gen.umn.edu/artist/articles/AERA_2003.pdf 21st April 2008.

Lipson, K. (2007). Assessing understanding in statistics. In B. Phillips and L. Weldon (Eds) Proceedings of the IASE/ISI Satellite Conference on Assessing Student Learning in Statistics. CD-ROM and online at http://www.swinburne.edu.au/lss/statistics/IASE/CD_Assessment

Morgan, C.J., & Deutschmann, P.W. (2003). An Evolving Model For Training And Education In Resource-Poor Settings: Teaching Health Workers To Fish. eMJA 2003 178 (1): 21-25

Pretty J.N., Gujit I., Scoones I., & Thompson J. (1995). A Trainer's Guide For Participatory Learning And Action. London: International Institute for Environment and Development

Ramsden, P. (1992). Learning To Teach In Higher Education. London: Routledge.

Svennson, E. (2007). Assessing learning by students' own examination tasks. Experiences from research courses in biostatistics. In B. Phillips and L. Weldon (Eds) Proceedings of the IASE/ISI Satellite Conference on Assessing Student Learning in Statistics. CD-ROM and online at http://www.swinburne.edu.au/lss/statistics/IASE/CD_Assessment

Werner D., & Bower W. (1995). Helping Health Workers Learn. Hesperian Foundation, Berkeley, California.

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

STATISTICS AS BREADTH: THE MELBOURNE EXPERIMENT. I: CONTENT AND DELIVERY

Ian Gordon, Sue Finch, Robert Maillardet
University of Melbourne, AUSTRALIA

Abstract
In 2008 The University of Melbourne introduced the `Melbourne Model' - a significant reform of its degree structure. Students enrol in one of six new degrees; 25% of their degree points must be taken as "breadth" material outside their core degree. This requirement can be met by enrolling in a "University Breadth Subject" which is available to all students and has no pre-requisites. We developed a subject called "Critical thinking with data". It has the bold intention of teaching important elements of statistical science, with minimal mathematics. We present our approaches to content and delivery of the subject. We made extensive use of visual and other media, integrating case studies from the press and elsewhere with the pedagogical content. Much of the background information is available via our learning management system. Three eminent guest lecturers provided inspiration from fields in which critical thinking about data is integral.

BACKGROUND

The opportunity to develop broad, no-prerequisite subjects with a strongly inter-disciplinary focus at The University of Melbourne (UM) arose out of the new Melbourne Model. This model is a radical departure from the previous Melbourne system, although recognisable as similar to structures elsewhere. There are only six undergraduate degrees: Arts, Biomedicine, Commerce, Environments, Music and Science. All students must take one quarter of their degree as "breadth", defined as subjects not part of their core degree. The Melbourne Model had its first intake this year (2008).

Students can meet the breadth requirement either by taking standard subjects that would not normally be taken within their degree (e.g. taking chemistry in a Commerce degree) or by taking a "University Breadth Subject". This new type of subject is broad in the sense of taking perspectives from many disciplines on a topic, involving teaching from different faculties and teaching students generic skills as well as subject specific content. We developed *Critical thinking with data* (CTWD) as this type of breadth subject; it aimed to teach first year students some fundamentals of statistical science.

The University requires breadth subjects like CTWD to be available to all students. At first year level there can be no pre-requisites, and the subject must be able to be taken with any other first year subject – the implication being that the content of CTWD must be sufficiently different from any other (statistics related) first year subject. These requirements also meant that we could not rely on students having a mathematical background.

In the last decade or so, introductory statistical education has become more applied, with more emphases on design and aspects of statistical thinking. This is exemplified in the newer generation of texts, starting with Moore and McCabe (1999). There are also liberal arts style introductions to statistical thinking which have less emphasis on mathematics; *Lies, damned lies and statistics*, a course taught at the University of Auckland, and Utts' (1994) book *Seeing through statistics* reflect this approach. A subject like CTWD was unprecedented at UM, and perhaps long overdue – we are surely well and truly into the period HG Wells envisaged when he wrote the famous words: "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write." Utts (2003) articulated seven important topics that the statistically educated citizen of today should know; this kind of general education is not traditionally part of an Australian undergraduate curriculum. Best (2005) argues that although we have long recognised the need to improve statistical literacy, "no discipline stepped up and took responsibility for teaching critical thinking" (p.214), and claims that "statisticians are likely to consider teaching courses in statistical literacy as beneath their talents." We suggest that the challenges of teaching statistical literacy require talents and orientations that are not always found among mainstream academic statisticians. Further, the educational task is formidable, but offers exciting potential for effective innovation and creativity.

OUR PHILOSOPHY

We wanted to convince the diverse audience for our course of its broad relevance to their everyday lives and to their academic aspirations. This meant that the materials we used needed to be relevant to young adults and should come, at least in part, from familiar sources. Materials needed to be rich, interesting and diverse – reflecting the application of the logic and procedures of data-based investigations in medicine, commerce, land and food science, ecology, biology, environmental science, and so on. The statistical content of CTWD would be enriched by the applied context in which it was presented; students would meet (and learn) more than just the statistical content. Examples – case studies – would be presented in some depth, and students would need to know about some of this content.

The two main developers of the content of CTWD work in a strongly applied and diverse environment – the Statistical Consulting Centre at UM. This provides the opportunity to see statistical applications across the full gamut of complexity, from the banal to the profound, and in all areas of quantitative research, and to learn about how researchers think about designing and analysing data. Consulting is approached as an educative process, and provides direct insight into the common understandings and misconceptions. Consulting requires a statistical disposition that is enquiring – we need to ask questions about the purpose and logic of data-based investigations, and about the processes involved in conducting such investigations; we need to question and interrogate data-based arguments. Consultants need good oral and written skills in communicating statistical knowledge, and data-based critiques and arguments. Our aim was to develop some of the dispositions of a statistical consultant in our students.

ACHIEVING BREADTH IN CONTENT AND CONTEXT

CTWD needed breadth in both statistical content and in its applications. It would be far from sufficient to take a ‘standard’ set of first year topics and present them with less mathematical content and more applied contexts. A CTWD student also taking a first year introductory statistics course would need to meet ‘new’ concepts and approaches to understanding and interpreting data-based investigations in CTWD. We did several things to develop broad statistical content.

We consulted an advisory committee of academics from all areas of the new generation degrees. We also consulted Professor Chris Wild, given his role in the Auckland successes in statistical education, to learn about the range of courses taught there. Three eminent guest lecturers each contributed two “bookend” lectures, bracketing a section of the subject. They provided a statistical orientation from different disciplines as well as engaging and realistic examples. The three disciplines involved were epidemiology, actuarial science and environmental science.

Breadth in applications – the examples and case studies we used extensively – was, of course, introduced by the guest lecturers. In the early stages of development of the subject, we put a great deal of effort into collecting materials describing data-based investigations, preferring those with general media representations (newspaper articles etc.) and “rich media” (video clips, images, substantial information on the web). We also saw much educational potential in the use of the media as a motivating source of material (see Watson, 1997). Applications in some fields tend to attract media attention more than others; medical examples are very common. We chose examples to ensure that we covered interest of students from all new generation degrees.

STRUCTURE AND BREADTH IN CONTENT

The structure of CTWD was informed by our own view of the central themes of statistical thinking and by liberal arts approaches to statistics (e.g. Utts, 1994). In statistical consulting especially, it is important to focus on the “big picture” questions, such as “where do the data come from”? These kinds of issues are along the lines of the “worry questions” of Wild and Pfannkuch

(1998), featured in the University of Auckland subject *Lies, Damned Lies and Statistics* and the 7 critical components described by Utts (2004).

The subject covered four "themes", and 15 topics, as shown below.

Topic	Theme 1: Finding data as evidence	
1	Data quality	Anecdotes, intuition or evidence
2	Context	Data – a number with social value
3	Variation	Embracing the way we vary
4	Sampling	Sampling matters
5	Designed experiments	Evidence – by design
	Theme 2: Examining evidence in data	
6	Graphics	Good pictures paint a thousand words
7	Summaries	Understanding relationships
8	Observational data	Relationships can be deceptive
9	Statistical models	Bell curves & other interesting models
	Theme 3: Understanding uncertainty in data	
10	Probability	Probability – objective and subjective
11	Risk	Understanding risk
12	Psychological influences on probability	Psychology and probability
	Theme 4: Drawing conclusions from evidence in data	
13	Confidence intervals	How certain can we be?
14	P-values	How significant is our evidence?
15	Meta-analysis	Resolving inconsistencies and accumulating knowledge

The topics for CTWD that particularly reflected breadth and divergence from 'standard' content were context, designed experiments and observational studies, risk, psychological influences on probability, and meta-analysis. The lectures on psychological influences on probability discussed, for example, empirical research characterising lay and expert judgements about chance and interpretations about probabilistic statements. Meta-analysis was included to deal with two common concerns introductory students often have – how to deal with seemingly contradictory findings from different studies, and how to move beyond the particular context in which a study is conducted. We believe an exposure to meta-analysis at an introductory level really helps students put another important aspect of experimental variability into an important broad context.

Breadth was also achieved in the depth and nature of the ways topics were covered. Graphics, for example, on face value might appear to be a topic common to CTWD and a standard introductory course. In a standard course, students might be taught about a variety of different types of graphical representations and how to construct them. The course might have some discussion about some important features of graphs – ensuring titles and axis labels are included, and so on.

In CTWD, there were three lectures on graphics. The lectures presented and discussed poor graphics, and demonstrated the way good communication and clear data-based arguments can be made with good graphics. An underlying theme was the pursuit of graphical excellence, exemplified by Cleveland's (1994) empirical work and insights from optics, and Tufte's (1983) work on visualising data – both of which have led to guiding principles for the construction of quantitative graphics.

The lectures discussed the importance of clear titles, axis labels and so on, but the central message was that good graphs follow five important principles: show the data clearly, use good alignment on a common scale for quantities to be compared, use simplicity in design, keep the visual encoding transparent, and prefer standard forms demonstrated to be effective. The standard forms recommended and discussed in detail were time series plots, bar charts, scatter plots, dotplots, histograms and boxplots. Panel displays as powerful tools for displaying more complex data were covered. The lectures on graphics and the examples used also provided an opportunity to discuss the identification and interpretation of unusual values, and to illustrate the role of unusual observations in explaining unusual outcomes or events. Students were also introduced to dynamic graphics exemplified by the Gapminder software of Hans Rosling.

The lectures on graphics used about 20 different case studies and examples; sources included media, Internet and research publications. Two case studies were presented in some detail – the Challenger disaster is a widely known example. In its simplest form, the Challenger case study illustrates the importance of choosing the appropriate kind of graphic. However, a richer presentation reveals messages about how the data chosen to inform the questions was wrong, how important contextual information was missing in the debate to launch, and how poor data representation can impede clear thinking.

The second main case study for graphics was based on a graphic showing greenhouse gas emissions by source in Australia, published in Royal Auto in March 2008 (Negus & Mikedis, 2008), shown below.

Error! Objects cannot be created from editing field codes.

Source: Royal Auto, March 2008, p. 12.

The Royal Auto graph is a stunning example of poor graphics; the proportions represented in the graph do not even preserve the order of the data in a table provided with the article. However, in investigating the background to the data presented in the graphic, we found lessons about the importance of context (what were the categories that the icons were intended to represent?), clarity in communication of the results of data-based investigations, and the need for attention to detail in reporting. The richness of case studies like these allowed us to integrate content across topics and reiterate important lessons already presented.

CASE STUDIES AND EXAMPLES

CTWD used a very large number of case studies and examples. Case studies were data-based investigations that had richness and some complexity, and were often relevant to more than one topic. The case studies actually fall into three broad categories, although we did not choose them with this taxonomy in mind. Of course, all case studies were primarily selected with the educational goal of illustrating relevant content.

1. There were case studies that are important historically, or have become so due to ways that educators have used them. These include such stories as John Snow and the Broad Street pump, the Challenger disaster, the very early randomised trials in England (streptomycin for TB, pertussis vaccine for whooping cough), the mice experiments for penicillin, the Literary Digest debacle, the Salk vaccine trials, smoking and lung cancer.
2. Some case studies were chosen because they were from an ‘unexpected’ source, while still illustrating an important topic. These were chosen to communicate the wide effectiveness of statistical science, and to engage students who may have interests in those areas. These case studies were identified through news sources – paper or radio. They were “good news value” – and hence had good potential to engage students. An example of this is Ben Olken’s randomised trial to investigate corruption control measures in rural Indonesia; one would not necessarily expect to see such a study design in development studies (Olken, 2007). Lucy King’s study of the possible deterrent effect of bee sounds on elephants was used: quirky, but directly illustrative of several important concepts in experimental design (King, 2007). We used Stanley Coren’s study of dogs and earthquakes to illustrate outliers and issues in observational studies (Coren, 2006).

3. Third, we used "current" case studies, sometimes integrating material directly into lectures only days after its media release. In general, current case studies had global significance if they were international (the surveys estimating deaths in Iraq, for example), or had strong local media or other impact if they were Australasian or Victorian. Examples of this type included the breast cancer cluster at the ABC studios in Toowong, Queensland and the NZ schoolgirls' exposure of Ribena's vitamin C content.

DELIVERY

CTWD has conventional components of delivery, specifically, lectures (36) and tutorials. There are no laboratory or practice classes; at the time of planning the course, the potential enrolments were difficult to estimate. If enrolments had been very large, practical classes would have been difficult to manage. Hence the delivery of CTWD was largely via lectures.

The lectures generally followed the sequence of topics above. The lectures made extensive use of "rich media". Examples of video clips we were able to use included George Bush dismissing the findings of the first Iraq survey, elephants reacting to bee sounds in Lucy King's experiment, Hans Rosling demonstrating the Gapminder software showing several dimensions simultaneously, and several stakeholders commenting on the Toowong breast cancer cluster. As our Royal Auto example suggests, we attempted to model good statistical thinking and practice in the lectures – to exemplify the disposition of a statistical consultant. We looked into the details behind the examples and case studies we found to question and interrogate data-based arguments and representations. Where possible, we constructed graphical representations of data that adhered to the principles of graphics we taught, rather than relying on (possibly mediocre) graphs from original sources. We collected data from students in a student-designed survey and as part of a regular departmental survey; the results were quickly processed and integrated into lecture material.

Our goal in lectures was to deliver rich statistical content without relying on mathematics. As we have described, we used rich context and interesting case studies to motivation and illustrate the important lessons. We used many examples, rather than few. We also used visual representations wherever possible – pictures, diagrams, graphics and graphs. StatPlay (Cumming & Thomason, 1998) was used to present simulations. We tried to consider the many different ways an idea, concept or principle might be represented.

Tutorials were designed to supplement lectures but did not aim to present new content to students. Tutorial activities gave students the opportunity to look at data-based investigations in detail, and to practice applying their developing statistical dispositions. As discussed in our companion paper on assessment, students worked on a large assignment in groups during tutorial time; one part of this large assignment was completed in a tutorial devoted to a poster-display and presentation session.

CTWD made extensive use of UM's Learning Management System (LMS). Lectures were recorded with Lectopia, and access was provided via the LMS; course content was always available to students. CTWD did not have a set text or a reading pack of materials. We used the UM's Digital Repository to give students access to multimedia and reading materials; this could be accessed via the LMS or the UM website. The Digital Repository for CTWD currently contains 129 items – newspaper articles, video clips, links to websites and academic papers. There is material directly relevant to the statistical content as well as material related to the case studies. Students are required to read or examine a small proportion of the material available on the Digital Repository. However, background material on case studies and supplementary material on statistical content is available so that students can follow-up on content that particularly interests or challenges them.

CONCLUSION

From the perspective of a professionally-trained statistician with considerable consulting experience the content of this course might appear to be common sense. The evidence of our teaching experience suggests otherwise. Intelligent but otherwise untrained minds can falter on many a misconception or preconception, and there is no shortage of either of these when it comes to the universally important task of thinking critically with data. Our building block concepts are quite abstract, so we taught them initially through appealing and concrete case studies drawn from students' everyday life and experience. Once students grasped the ideas in one context we then helped them to generalise these concepts by repeated exposure to manageable but open-ended interpretative tasks using on-line and standard assessment tools, again drawing from everyday life and the media. We believe that our new course has been an important first step towards helping a much broader audience take its own first – albeit faltering – steps towards developing genuine statistical insight.

REFERENCES

Best, J. (2005) Lies, calculations and constructions: Beyond how to lie with statistics. Statistical Science, 20:210-214.

Cleveland, W.S. (1994) *The elements of graphing data.* Monterey Park, CA: Wadsworth.

Coren, S. (2006). Can dogs predict earthquakes? A possible auditory answer. (Paper delivered at the Psychonomic Society meetings, Houston, Texas, November 2006.)

Cumming, G., & Thomason, N. (1998). StatPlay: Multimedia for statistical understanding. In L. Pereira-Mendoza, L. Kea, T. Kee, & W-K. Wong (Eds.) Statistical education: Expanding the network. Proceedings of the Fifth International Conference on Teaching of Statistics (pp. 947-952). Voorburg, The Netherlands: International Statistical Institute.

King, L.E., Douglas-Hamilton, I., Vollrath, F. (2007). African elephants run from the sound of disturbed bees. *Current Biology*; 17: R832-833.

Moore, D.S. & McCabe, G. (1999) *Introduction to the practice of statistics*. New York: W.H. Freeman.

Negus, B., & Mikedis, S. (2008) The road to greener motoring. *Royal Auto*, March: 12-13.

Olken, B.A. (2007). Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy*; 115: 200-249.

Tufte, E. (1983) *The visual display of quantitative information*. Graphics Press.

Utts, J. (2003). What education citizens should know about probability and statistics. *American Statistician*, 57:74-79.

Utts, J. (2004). *Seeing through statistics*. Belmont, CA: Thomson Brooks/Cole.

Watson, J.M. (1997). Assessing statistical thinking using the media. In I.Gal & J. Garfield (Eds.), *The Assessment Challenge in Statistics Education*. (pp. 107-121). Amsterdam, the Netherlands: IOS Press.

Wild, C. & Pfannkuch, M. (1998). What is statistical thinking? In L. Pereira-Mendoza, L. Kea, T. Kee & W. Wong (Eds.) Proceedings of The Fifth International Conference on Teaching Statistics 21-26 June 1998, Vol. 1 Voorburg, ISI, (pp. 335-341).

STATISTICS AS BREADTH: THE MELBOURNE EXPERIMENT. II: ASSESSMENT

Sue FINCH, Ian GORDON, Robert MAILLARDET
University of Melbourne, AUSTRALIA

Abstract
"Critical thinking with data" is a new "University Breadth Subject" developed for first year students under The University of Melbourne's "Melbourne Model". It aims to teach important elements of statistical science, with minimal mathematics, and was taught in first semester 2008. We present our approaches to assessment of the subject. This has required the use of approaches that are quite distinct from mainstream statistical subjects, since students are not really being taught to do statistical work. They are required to make astute judgments of material with quantitative information, including such texts as a short article about some research in the newspaper. We have used a variety of forms of assessment, including weekly quizzes, (very) short assignments, and a larger project. The style of assessment is more consistent with that used in humanities subjects, and therefore has some important challenges for staff involved in marking, for example.

WHAT IS CRITICAL THINKING WITH DATA?

Critical thinking with data (CTWD) is a new subject available to all first-year University of Melbourne students in 2008. Students learn about important aspects of statistical science. As a "University Breadth Subject" it draws on perspectives from a very broad range of disciplines and covers stages of data-based investigations from asking a research question to interpreting and communicating findings. Our goal is to help students develop statistical dispositions and an appreciation of how core concepts of probability and statistics inform real-world arguments based on empirical data. Gal and Garfield (1997) characterise statistical dispositions as involving an appreciation of chance and randomness, of the use of data-based methods in making decisions in uncertain situations, and of the value of data-based investigations over anecdote and subjective opinion; statistical dispositions prompt us to question and interrogate data-based arguments.

With statistical dispositions, students can learn to critically assess data-based arguments in media reports and academic materials, and construct sound data-based arguments. They should have an awareness of the potential flaws in reasoning about data and the likelihood of events, and in communicating information about data-based investigations.

A 'breadth' subject at The University of Melbourne is broad in many senses; it draws on perspectives and knowledge from many disciplines, and provides students with knowledge and skills that have broad application across disciplines. Breadth subjects also teach students more generic skills including those relating to presentation, written communication, and teamwork.

The content of *CTWD* covers the purpose and logic of data-based investigations, the processes involved in conducting such investigations including collecting, summarising and drawing conclusions from data. Students learn about the variety of approaches that arise across disciplines, and the different processes involved. The course includes some fundamental ideas of probability, and material about how people understand probability. There is virtually no teaching of procedural and computational skills in carrying out statistical analysis in *CTWD*. However a well-tuned statistical disposition requires the ability to question and interrogate numerical information and to make sense of summary reports of data; this requires some simple computational skills. Our companion paper – Statistics as breadth: the Melbourne experiment. II: content and delivery – provides more detail about the rationale for the subject.

Student learning is often driven by assessment. There are important messages to students in the content we choose to assess, and in the way we ask questions and set projects. The assessment and exam in one semester can set the agenda for students in subsequent semesters. The framework briefly outlined above required assessment tasks that allowed students to demonstrate their developing statistical dispositions as well as some generic skills. Assessment materials needed to be strongly grounded in real-world problems; richness without ambiguity or too much statistical complexity would be ideal.

OUR STRATEGY

A team of staff in consultation with academics from a range of disciplines developed the course materials for *CTWD*; the course was developed from the ground up. A framework of lecture topics with some details of content was outlined. We then described the specific learning outcomes that we wanted to achieve for each topic. By considering the learning outcomes at this stage, we ensured that lecture material was well-tailored to the types of knowledge, skills and dispositions we wished to assess.

For example, one of the topics *CTWD* covered was graphics. The lecture content provided illustrations of good and bad graphical practices, and discussed important features of good graphs and research on interpretation of graphs. Lectures gave examples of media reports including graphical displays and modelled how to critique and improve graphical displays. Five principles for good graphics were presented, and a number of standard forms were recommended. The principles were developed from the work of Tufte (1983) and Cleveland (1994).

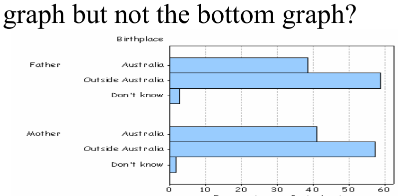
The learning outcomes required students to develop limited skills in producing graphs for small data sets. However, they did not use software. Rather, students should understand:

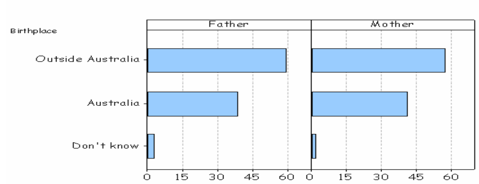
Good graphical presentation of data rarely happens automatically.
Good graphs are simple in design.
Graphs should, above all else, show the data clearly.
Good graphs have transparent visual encoding.
Good graphs have accurate titles and well-labelled axes with measurement units defined.
Axes should generally maintain constant measurement scales.
Good graphs identify the source of data.
Pie charts and pictograms are poor choices for representing data.
Aligning data along a common horizontal axis facilitates accurate comparisons.
Standard forms have been developed for particular purposes.
Unusual values should be investigated carefully for information they may provide; they are not necessarily mistakes.

A limited search for some very simple graphics software capable of supporting the production of well-designed graphics was unsuccessful. Some commonly available software was avoided because it encouraged the unthinking production of some very poor designs and non-standard forms.

The assessment tasks we designed evaluated students' capacity to apply this knowledge appropriately to real-world problems. Hence we asked questions that related to recognising the instantiation of the principles we wished students to learn rather than requiring the production of 'statistical output' according to these principles. For example:

> Consider the two graphs below. They show responses to the questions about the birthplace of parents from a 'Getting to know you' survey designed collaboratively with students in the course. Which important graphical design principle described in lectures is used in the top graph but not the bottom graph?





OUR CHALLENGES

There were many challenges in developing and choosing methods of assessment and appropriate materials for *CTWD*. A very tailored approach was needed; existing assessment resources, including exercises and activities, could not really contribute to the course. The assessment tasks should require the application of generic skills and data-related thinking skills; traditionally assessment in statistical courses had focussed more on the data-related skills.

We chose to use examples and materials that students themselves might have easily come across, in order to emphasize the relevance of the course to material met in the media as well as academic contexts. There are many beautiful historical examples and case studies that are ideal lecture material, some of which we used, including stories of the trials of the Salk vaccines, the decision to launch leading to the Challenger disaster, and John Snow's mapping of the 1854 cholera epidemic. Assessment materials needed to be contemporary and have a strong local flavour. Local material we used included the story of the breast cancer cluster in the ABC studios in Toowong, research on cannabis use in Australian adolescents and a Herald-Sun survey of Victoria Police.

Traditional statistics courses often rely on good examples of the methods they are trying to teach. We needed to find examples with a mix of good and bad features, and with the right degree of complexity to allow students to apply their critical enquiry skills. Simple flawed examples of data-based investigations can assess an important principle but have limited scope for assessing varying depths of students' understanding; examples that are too complex can include features that are not relevant to the assessment task but may mislead students to focus on them.

The student cohort in *CTWD* is mixed; in Semester 1 2008, 28% were Arts students, 27% were Science, 18% and 15% were from Commerce and Biomedicine respectively. The capacity of this mix of students to deal with the course content and assessment was unknown. Interestingly, the course attracted a broader spread of students than many other University breadth subjects launched at the same time. The majority of students in many of these other subjects were from just one or two faculties corresponding to a specific discipline interest relevant to their main course of study.

Traditional statistics courses often rely on a mathematical approach to teaching and assessing many fundamental statistical ideas. *CTWD* included topics such as variation and statistical modelling, but without formal mathematical treatment. Developing appropriate lecture content required innovation and finesse; setting appropriate assessment tasks was even more challenging.

TYPES OF ASSESSMENT

Assessment needed to be continuous for both course developers and the students. Early in the course, students expressed uncertainty about what they were supposed to be learning from lectures, or indeed what the subject was "all about". Course developers needed feedback about students' capacity to deal with the scope of material presented. *CTWD* used four types of assessment.

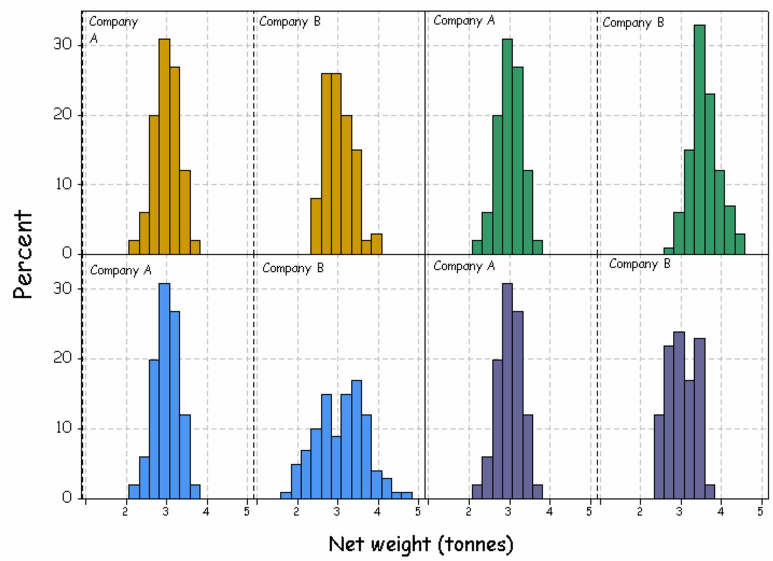
Weekly quick quizzes of 8 to 10 questions could be attempted repeatedly (until the closing date). These quizzes were worth 5% of the total marks; they were important for revision of lecture content and concepts. There were 5 short assignments worth 4% each and totalling 1000 words. A major assignment was worth 25%, and had two components – group work and an individual write up. The final exam was worth 50% and was modelled on the other assessment components. The quizzes, short assignments and major project are described in detail below. We also provide an example from the exam.

QUICK QUIZZES

Students completed the weekly quick quizzes via the University's Learning Management System. This allowed a variety of different types of questions to be asked including multiple choice, multiple answers, exact numerical, matching, ordering, fill-in-the-blank and 'hot spot'; all of these types of questions were used in *CTWD*. A 'hot spot' question gives a visual representation, and students need to click on the area of the representation that gives the answer to the question.

The use of visual representations was one approach used to present and assess concepts and principles traditionally given a formal mathematical treatment. For example, this hot spot question examined the idea of modelling location differences in relation to background variation:

> There are four pairs of histograms below. Each compares the net weight of garbage dumped by two companies. Click on the pair that shows the strongest evidence of mean differences between the two companies, relative to the background variation within companies.



SHORT ASSIGNMENTS

The length of the five short assignments was deliberately very constrained – students could only write 200 words on each. The due dates for the assignments were spaced throughout the semester. Broadly, the assignments asked for a critical evaluation of a data-based argument, report or representation; in each case, students were guided in terms of which aspects to focus on. For example:

> This assignment is about a small item in The Age's "Diary" column, which is published on the back of the front section of The Age newspaper. The item appeared on 29 November 2007 as follows:
>
> *Yeah, right*
>
> … And how about No Idea Magazine's latest poll discovery: 'one in five' women in Australia felt threatened by violence in their homes. Check the bottom line and you find the poll quizzed just 1500 of the nation's 20.4 million people, thereby smearing at least 2 million blokes in a poll covering just 0.000073% of the population.

Students were asked to write two emails: one to The Age diarist pointing out the misconceptions and errors in the short quote, and a second to the magazine involved (New Idea, derisively referred to as No Idea) to ask for important information to assist in an evaluation of the survey.

This was the second short assignment; it assessed students' understanding that it is the sample size rather than the sampled proportion of the whole population that is important to the statistical reliability of a study. Lectures had discussed the 1936 Literary Digest poll extensively. Students needed to check the percentage reported from the figures provided in the article in order to identify a computational error. Most students took the calculation at face value; they did not expect that it required checking. This is an example of the kind of checking and simple computational skill needed.

Another short assignment asked students to evaluate a graphical representation found on a website describing the capacity in Melbourne's water storages over time. Students were provided with a number of alternative representations of the data, and asked to choose the one best suited to a particular purpose and to write a description based on this.

The short assignments assessed critical thinking skills in an applied context, and also required good reading comprehension and writing skills. The assignments gave very clear guidance about the nature of the task, the content to consider, and the points to be addressed in the word limit. Some students still failed to evaluate the materials within these guidelines.

MAJOR ASSIGNMENT

The major assignment gave students the opportunity to make a detailed review of a single research study. The lectures illustrated how media reports often only tell part of the story, and how many important and interesting questions are raised when the background material for the study is examined. Students worked on one of five case studies; each case study had a newspaper report and a published article.

There were two parts to the major project. Students worked in groups on the first part; groups were formed within tutorials to prepare for a (tutorial) poster display and presentation session. This session was modelled on a conference poster session; each group gave a brief presentation of their poster for 3 minutes, and then assessors and fellow students circulated the room to learn more about the research study. Both the verbal and poster presentation were assessed. One aim of this task was to help students understand the research study they were reviewing quickly and efficiently. We chose studies that were relatively straightforward in their design, but the complexity of the statistical analysis was quite varied. We did not expect students to understand all of the details of the analysis in the papers provided; in the tutorial time provided for establishing the group work, students were given guidance about this.

The second and individual part of the major project asked students provide a 1200 word critical assessment of the reporting of the study in the news item and in the published article, and of other aspects of the design, implementation and analysis of the study. This critique required description and explanation of both the strengths and weaknesses of the study. Students could comment on the analysis of some of the studies, for example, by critiquing the graphical presentation of the results.

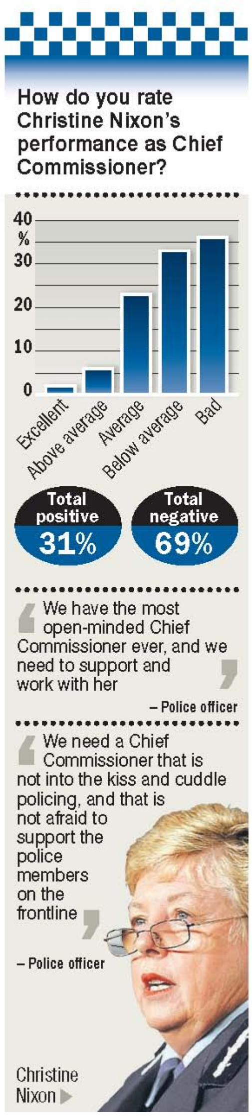
The major project assessed students' capacity to work well in groups as well as their oral presentation and written communication skills.

EXAMINATION

The examination included quick quiz type questions, short answer questions and longer questions similar to those used on the short assignment. Here is an example of a longer question:

> In April 2008, The Herald Sun published a series of articles over several days about a poll of Victorian police officers. The headline on the first day was "POLL: POLICE FACE CRISIS". There had been heightened tensions between the Police Commissioner, Christine Nixon, and the secretary of the Police Association, Senior Sergeant Paul Mullett.
> The Herald Sun published the following information about the survey:
> "The Herald Sun wrote to more than 11,000 Victoria Police officers and invited them to complete the Herald Sun survey – 3459 responded. That means 30 per cent of the force's sworn officers took part in the unprecedented poll.
> The survey sample of 3459 is much larger than those commonly used by pollsters. Galaxy Research vetted the questions to ensure they conformed to accepted standards of conducting research in Australia.
> The Herald Sun provided each police officer with a unique password, which expired after it was used once. That ensured they could not submit multiple entries.

The completed surveys were analysed by an independent market research company, which provided the results to the Herald Sun for publication."



Unless otherwise specified, use the extracted information from the Herald Sun article to answer the following questions.

a) What was the sample frame for this survey?

b) The day the survey was released, Chief Commissioner Christine Nixon was critical of it: "I don't intend to go anywhere," Ms Nixon said this morning after a landmark survey raised a raft of serious issues about the state of the force. Speaking on radio 3AW this morning, Ms Nixon attacked the results because just 30 per cent of sworn officers – 3459 – had responded to the survey. "That is despite the sample being much larger than that commonly used by pollsters." (Herald Sun)

What was the basis of Christine Nixon's criticism?

c) What argument is the Herald Sun using in its comment: "That is despite the sample being much larger than that commonly used by pollsters."?

Provide an analysis of the argument.

d) What strengths and weaknesses of the survey are apparent from the Herald Sun's description of the survey? Comment on any features you have not discussed in (b) and (c) above.

Parts (e) to (h) refer to the extract from the Herald Sun shown at the left.

e) What type of information is in the two quotes supplied at the bottom? What weight should be attached to these comments as evidence?

f) Examine the bar chart at the top of the extract. What type of data is represented?

g) Comment on the descriptions of the categories of response ("Excellent", "Above average" … "Bad") and the effect they could have had on the responses.

h) How were the two simpler categories – "Total positive" and "Total negative" constructed? Comment on this construction.

This relatively simple example allowed questions to be asked about many aspects of the processes involved in a data-based investigation as well as to assess examples of statistical reasoning. In the examination, questions about this type of material were quite structured give the time constraints within which students had to work. In short assignments there was less explicit guidance.

MARKING

Tutorial staff from three different departments supported CTWD. Just as the students were required to make astute judgements about the material provided, tutors were required to make astute judgements about student arguments. The style of assessment is more consistent with that used in humanities subjects. Constraints on short assignments were set to encourage students to write clearly and succinctly, but also to limit marking loads. Second markers attended poster display and presentation sessions to assist with on-the-spot marking of presentations and student responses to questioning.

STUDENT FEEDBACK

Formal feedback about assessment and other aspects of the course was sought in an online survey during the fourth week of the course; 93 students (63% response rate) responded. At this time, over 50% of students indicated that the short assignments were beyond the right level of challenge. About half of the students found the word limit too restrictive. One student gave this perspective: “Short assignments are perfect I think, and the 200 word limit really forces you to develop concise yet relevant points.” Informal feedback towards the end of the course suggested that students did appreciate the challenge and constraints of short assignments, and they felt the assignments had helped them develop their writing skills. We are yet to receive results from students’ evaluations in the final week of the course.

LESSONS

The development of any new subject is time consuming. The development of *CTWD* had unusual demands of needing specialised materials and innovative approaches to teaching and assessment. The contribution of staff working in applied fields meant that a great deal of potential contemporary local material could be identified for lecture material and interested assessment tasks. Often our investigations of the details and background behind the stories we found revealed richer material that could also be used in teaching and assessment. In two cases the analysis work done to prepare a case study for presentation in the subject led to important discoveries about the published research, which were communicated back to grateful original authors.

The challenge to teach a mixed and different cohort of students forced us to consider the many different ways in which important statistical concepts, principles and ideas can be represented and explained – through examples, counter-examples, visual representations, simulations, and (very occasionally) with formulae. The need to consider the broad relevance of statistical dispositions to many disciplines without relying on mathematics brought new perspectives on teaching and assessment – perspectives that will inform and enhance our teaching in a broad range of courses. Our own assessment of the approaches we used to assessment in *CTWD* has only just begun.

REFERENCES

Cleveland, W.S. (1994) *The elements of graphing data.* Monterey Park, CA: Wadsworth.

Gal, I., & Garfield, J. (1997) Curricular goals and assessment challenges in statistics education. In Gal, I., & Garfield, J. (Eds.) *The assessment challenge in statistics education*. Amsterdam: International Statistical Institute.

Tufte, E. (1983) *The visual display of quantitative information*. Graphics Press.

TWO PROFILES OF STATISTICS LEARNERS: A DISCRIMINANT ANALYSIS OF ATTITUDES TOWARD STATISTICS

Zamalia Mahmud
University Teknologi MARA, Malaysia

Abstract

This study attempts to construct the profiles of two types of statistics learners: namely those with a positive and a negative attitude towards statistics. The contribution of this work lies in its attempt to characterize each profile of learner by relating to his/her perceived attitudes toward statistics, types of learners, mode of study, program structure, age, gender and learners' evaluation towards the statistics course. These variables are used as predictors that discriminate learners with positive and negative attitudes toward statistics. The results indicate that learners with positive attitudes can be reliably distinguished from learners with negative attitudes toward statistics. This then can assist instructors to optimize the teaching and learning of statistics in the classroom.

INTRODUCTION

What distinguishes a statistics learner with a positive attitude towards statistics from a learner with a negative attitude towards statistics? Does each type of learner have a different profile of attributes? This study attempts to construct profiles of two types of statistics learners - those with a positive attitude towards statistics and those with a negative attitude towards statistics. The process involves identifying the predictors that discriminate between learners with a positive attitude and learners with negative attitudes toward statistics. This study attempts to characterize the profile of each learner by looking into their perceived attitudes toward statistics based on Schau's Attitude Towards Statistics (ATS) instrument (Schau et. al., 1995) comprised four dimensions (Affect, Cognitive Competence, Value, Difficulty), types of learners, mode of study, program structure, area of study, age, gender and the learner's evaluation towards the statistics course.

ATTITUDES TOWARD STATISTICS

Studies on attitude towards statistics have been conducted in various parts of the world and different aspects of attitude surveys were reported (Schau *et al*., 1995; Wisenbaker *et al*., 1995). However, most studies were confined within their own respective courses. The challenge in conducting such studies is the ability to measure the students' attitude across several disciplines prior to their enrolment in any introductory statistics course.

Many statistics educators and most statistics students believe that attitudes toward statistics are important in the learning process. Schau (2003) discovered that students attributed their positive change towards the learning of statistics to the attitudes of their instructors/teachers. They attributed their negative attitudes at the beginning of the statistics classes to poor teaching that eventually led to poor achievement in mathematics.

In another related study, Wisenbaker *et al.* (1995) opined that attitudes toward statistics and course achievement causally impact each other. Schau (2003) defined Prior Attitudes and Prior Achievement as exogenous variables where students who enter classes already possess attitudes toward statistics and learning that will impact their course performances. Attitudes and Course Achievement are endogenous variables that impact each other throughout the course and are impacted by both Prior Attitude and Prior Achievement.

There is also a growing interest among statistical education researchers on the extent of the relationship between the attitude dimensions (Affect, Cognitive Competence, Value and Difficulty) and the students' profile such as age, gender, mathematics and statistics achievements (Cashin & Elmore, 2005; Hilton *et al*., 2004). It is also of concern to several researchers to determine which dimensions of attitude subscales are expected to impact on the statistics course performances (Carmona, 2002; Kottke, 2000).

Results from Mill's study (2004) revealed that undergraduate students who enrolled in an introductory undergraduate statistics course at a large southeastern university in the College of Business have more positive attitudes toward statistics, a finding that coincides with several prior researches (Perney & Ravid, 1990; Waters et al., 1989). However, to a certain extent a small number have indicated less than positive attitudes, i.e. students agreed that they get frustrated over statistics tests in class, that statistics is a complicated subject, that it requires a great deal of discipline, that it is highly technical, and that it is not a subject quickly learned by most people.

An examination of the cross tabulations of the gender variable provided the most interesting results. It was depicted that males were more likely than females to report that they were not scared of statistics, that they can learn statistics, and they felt confident mastering statistics material. Similar results on females' negative attitudes have been discussed (Fullerton & Umphrey, 2001) but others have reported no differences between males and females (Faghihi & Rakow, 1995; Waters et al., 1989). The results of another study (Ware and Chastain, 1989) revealed that further attention may be required to improving female attitudes toward statistics particularly if their academic performance also suffers.

METHODS AND INSTRUMENT

The study was conducted on two profiles of participants – government officers attending a compulsory course in statistics and data analysis as part of the requirement for securing a scholarship for further studies and postgraduate students attending a statistics and data analysis course as part of the postgraduate studies requirement. A sample of 200 out of 240 course participants responded to the questionnaire which addresses several issues. The respondents were asked to answer a number of questions which included background and demographic information, personal characteristics, course evaluation and more importantly their perceived attitudes toward statistics (ATS) constructs across four dimensions – Affect, Cognitive Competence, Value and Difficulty. The ATS constructs were adapted from Schau *et al.* (1995) and used in the study. The ATS is a 28-item instrument with a 7-point, Likert-type response format, with higher ratings indicating more positive attitudes after recoding the 19 negatively keyed items. The instrument incorporates four subscales, including the 6-item Affect subscale, the 6-item Cognitive Competence subscale, the 9-item Value subscale, and the 7-item Difficulty subscale. Examples of items/constructs on the Affect subscale are "I like statistics" and "I feel insecure when I have to do statistics problems"; on the Cognitive Competence subscale – "I make a lot of math errors in statistics" and "I can learn statistics"; on the Value subscale – "Statistics is worthless" and "I use statistics in my everyday life"; and on the Difficulty subscale – "Statistics is a complicated subject" and "Learning statistics requires a great deal of discipline".

ANALYSIS AND RESULTS

This study uses discriminant analysis, a method used to assess whether or not a set of variables discriminates between two groups of participants. Discriminant analysis produces discriminant function coefficients for each predicating variable, which indicates the importance of each variable. This study also uses means to compare the differences in the perceived attitudes of learners between the profile and characteristics of the participants. Learners' attitudes toward statistics were investigated in order to identify the categories of attitude - positive and negative based on their perceived attitudes toward statistics across the four dimensions namely, Affect, Cognitive Competence, Value, and Difficulty. Based on the variable distribution (ranging from 1 to 7), positive attitude was determined as equal to or greater than 4.50 and negative attitudes was set as equal to or smaller than 3.50. Consequently, 163 respondents (82% of the total number of respondents) were included in the analysis with 133 (81.6%) learners having positive attitude and 30 (18.5%) learners displaying negative attitudes toward statistics. The rest, 37 respondents (18.5%) were in the mid-range of the scale, where attitudes were considered to be neither positive nor negative. Of the learners with positive attitudes, 79% were government officers, 21% were postgraduate research students; 48% were males, and 52% were females; of the learners with negative attitudes 50% were government officers and 50% were postgraduate students. About 57% were males, and 43% were females. Both groups attended the statistics and data analysis course conducted at different point of time.

In the following stage, statistical differences were tested between learners with positive and negative attitudes in relation to the predictors. Table 1 depicts the results for the test for equality of group means. Based on these tests, it was determined which of the variables discriminated between learners with positive and negative attitudes. The variables that showed differences between positive and negative learners were types of learners, mode of study, perceived attitudes toward statistics based on the Affect, Cognitive Competence, Value and Difficulty subscales, mode of study and learners' evaluation towards the statistics course. Gender was excluded since there was no significant difference. Finally, a discriminant analysis was conducted to predict group membership from a set of the statistically significant predictors. Table 2 presents the results of the discriminant analysis model. It shows that the variable with the largest effect on attitudes is course evaluation followed by mode of study, value, types of learners, cognitive competence, affect and difficulty. Box's M test results in Table 3 indicate that the data do not differ significantly from the multivariate normal (sig. $p = 0.168$). Discriminant analysis maximizes the between-groups differences on discriminant scores and minimizes the within-groups differences. The eigenvalue is one statistics for evaluating the magnitude of a discriminant analysis. In Table 3, the eigenvalue was 5.913 with a canonical correlation of 0.925. Squaring the canonical function equals 0.855 which indicates that 85.6% of the variability of the scores for the discriminant funtion is accounted for by the differences between the two groups of learners. Here the eigenvalue is high which implies that the between-groups differences are much greater than the within-group differences. Wilks' λ indicates how good the discriminating power of the model is. Wilk's λ which equals 0.145 indicates that differences between the two groups of learners account for 100% of the variance in predicting the variables. The significance of the χ^2 implies that the discriminant functions discriminate learners with positive and negative attitudes toward statistics well. The discriminant analysis also reveals that for both positive and negative learners, 100% of the original cases are correctly classified.

Table 1: Tests of Equality of Group Mean

Predictor variables	Wilks' λ	F	df1	df2	Sig.
Types of Learner	.948	8.658	1	158	.004
Gender	.998	.312	1	158	.577
Age	.991	1.490	1	158	.224
Mode of Study	.974	4.230	1	158	.041
Program Structure	1.000	.031	1	158	.860
Affect	.713	63.619	1	158	.000
Cognitive Competence	.741	55.284	1	158	.000
Value	.740	55.594	1	158	.000
Difficulty	.739	55.684	1	158	.000
Course Evaluation	.219	564.729	1	158	.000

Table 2: Discriminant analysis of learners'attitudes toward statistics

Predictor Variables	Canonical Discriminant Function
Types of Learner	0.698
Mode of Study	0.973
Affect	0.512
Cognitive Competence	0.523

Value	0.900
Difficulty	0.442
Course Evaluation	-3.698
Constant	3.111

Table 3: Eigenvalues and Wilks' Lambda Test

Function	Eigenvalue	% of Variance	Cum %	Canonical Correlation	Wilks' λ	χ^2	df	p
1	5.913(a)	100.0	100.0	.925	.145	294.837	11	.000

The differences between learners with positive and negative attitudes toward statistics with regard to the predicting variables that were found to be statistically significant are also described in Tables 4 and 5. The results revealed two different profiles of learners with positive and negative attitudes toward statistics where learners who were classified as having positive attitudes were government officers enrolled in a full time masters by course of study and learners who were classified as having negative attitudes were comparable between postgraduates and government officers, also enrolled in a full-time masters by course of study. There is no significant difference of attitudes between the male and female respondents. This result is consistent with the findings of Faghihi & Rakow (1995), and Waters *et al*., (1989). In Table 5, all ATS dimensions (Value, Cognitive Competence, Affect, Difficulty) were significantly different between the two profiles of learners. Respondents with positive attitude scored higher on the ATS subscales compared to those with negative attitudes.

Table 4: Demographic variables by learners' attitudes toward statistics

Predictor	Categories	Positive (%)	Negative (%)	χ^2	p-value
Types of learners	Postgraduate students Government officers	21.1 78.9	50.0 50.0	10.562	.001**
Mode of study	Full time Part time	91.7 8.3	78.6 21.4	4.172	.041*
Program Structure	PhD Masters by research Masters by coursework	37.6 8.3 54.1	33.3 26.7 40.0	8.176	.017*
Gender	Male Female	56.7 4303	48.1 51.9	0.715	.398

*p<0.05; **p<0.01

Table 5: ATS dimensions by learners'attitudes toward statistics

Dimensions of ATS	Reliability[a]	Positive Mean (SD)	Negative Mean (SD)	t-statistics	p-value
Value	0.807	5.45 (.70)	4.36 (.69)	7.727	.000**
Cognitive Competence	0.805	5.18 (.66)	4.01 (.85)	8.290	.000**
Affect	0.800	5.17 (.83)	3.81 (.63)	8.414	.000**
Difficulty	0.697	3.76 (.74)	2.64 (.49)	10.143	.000**
Course Evaluation	0.749	5.41 (.537)	2.90 (.305)	24.637	.000**

[a]Cronbach's α. *p<0.05; **p<0.01

DISCUSSION AND IMPLICATIONS

Based on the canonical discriminant equation of Z = 3.111-3.698(course evaluation) + 0.973 (mode of study) + 0.900(value) + 0.698(type of learner) + 0.523 (cognitive competence) + 0.512 (affect) + 0.442 (difficulty), future predictions can be made on the profile of learners' with regard to their attitudes toward statistics. The study reveals two profiles of statistics learners, one group with a positive attitude towards statistics, and another group with negative attitudes toward statistics. The groups can be distinguished by learners' perceived attitudes toward statistics across the four ATS dimensions (Value, Cognitive Competence, Affect, Difficulty), mode of study, types of learner, program structure and course evaluation.

The findings have important implications for both learners and instructors of statistics. Even though the percentage of those with negative attitudes is small, instructors must be aware of the effect of their methods and approach of teaching statistics on learners' attitudes toward statistics. Instructors should therefore be attentive to the various components of teaching methodology and the effective delivery of statistics contents rather than just focusing on rote learning. Learners who evaluate high on the course tend to perceive their attitudes toward statistics positively than those who evaluate low on the course. Those who enroll in full time programs also tend to perceive their attitudes toward statistics positively than those who enroll in the part time programs. This should also be of concern to instructors, as different approaches may need to be adopted in handling between full-time and part-time learners. Based on the scores of each ATS dimensions, learners who scored high on the Value, Cognitive Competence, Affect and Difficulty subscales tend to perceive their attitudes toward statistics positively.

In this study, the discriminant analysis prediction will help course instructors to distinguish the group of learners and identify factors that predict learners' attitudes toward statistics. Knowing the profile of learners would enable instructors' to diversify their course contents and develop more innovative methods of teaching statistics. Perhaps a more practical and worked example approach plus remedial classes provided for learners with negative attitudes toward statistics can encourage active participation in the classroom and spark more interest in learning statistics.

REFERENCES

Carmona, J. (2002). Mathematical background and attitudes toward statistics in a sample of undergraduate students. [On-Line] www.stat.auckland.ac.nz/~iase/publications/11/Carmona.doc

Cashin, S.E & Elmore, P.B. (2005). The Survey of Attitudes Toward Statistics Scale: A construct validity study. *Educational and Psychological Measurement, 65(3),* 509-524.

Faghihi, F., & Rakow, E. A. (1995). The relationship of instructional methods with student responses to the survey of attitudes toward statistics. (ERIC Document Reproduction Service No. ED 392 358).

Fullerton, J.A., & Umphrey, D. (2001). An analysis of attitudes toward statistics: Gender differences among advertising majors. ERIC Document Reproduction Service No. ED 456 479.

Hilton,S.C., Schau, C., & Olsen, J.A. (2004). Survey of Attitudes Toward Statistics: Factor structure invariance by gender and by administration time. Structural Equation Modelling, 11(1), 92-109.

Kottke, J.L. (2000). Mathematical proficiency, statistics knowledge, attitudes toward statistics, and measurement course performance. *College Student Journal, 34(3),* 334-347.

Mills, J.D.(2004). Students' Attitudes Toward Statistics: Implications for the Future. *College Student Journal*, Vol. 38.

Perney, J., & Ravid, R. (1990). The relationship between attitudes toward statistics, math self-concept, test anxiety and graduate students' achievement in an introductory course. (ERIC Document Reproduction Service No. ED 318 607)

Schau, C. (2003). *Students' attitudes: The "other" important outcome in statistics education.* Paper presented at Joint Statistical Meeting – Section on Statistical Education.

Schau, C., Stevens, J., Dauphinee, T.L. & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement, 55,* 868-875.

Wisenbaker, J. & Scott, J. (1995). Attitudes about statistics and achievement among students taking introductory statistics courses. Paper presented at the meeting of the American Educational Research Associates, San Francisco, CA.

Waters, L.K., Martelli, T.A., Zakrajsek, T., & Popovich, P.M. (1988). Attitudes toward statistics: *An Evaluation of Multiple Measures. Educational and Psychological Measurement,* 48, 513-516.

A COMPARISON OF "STUDENT EVALUATION OF TEACHING" WITH STUDENT PERFORMANCE

H. Short, R. Boyle, R. Braithwaite, M. Brookes, J. Mustard, and D. Saundage
Deakin University, Australia

Abstract
This study measures the evaluation of teaching given by students against their final outcomes in a subject. The subject in question had an enrolment across four campuses of 1073 students at the time of the evaluation and is a statistics subject that is core (i.e. compulsory) to several undergraduate business degrees. This study is based on the 373 students (34.8%) who responded to the survey, and their final results. The evaluations were open for a period of six weeks leading up to and just after the final exam. The study matches the responses to the question "This unit was well taught" to final outcomes, in an attempt to ascertain whether there is a link between student evaluation of teaching and performance. The analysis showed that for the students who self-selected to complete the survey:

- *Students who perform well in the subject generally give higher scores than lower performing students.*
- *The same general pattern prevailed when other secondary factors were taken into account, such as, when the evaluation was completed, campus and gender.*
- *The timing of when a student completes the evaluation seems the most important of these secondary variables.*
- *In general, students who submitted their evaluations after the exam gave higher ratings if they eventually obtained a pass grade or better, and lower grades if they failed.*

INTRODUCTION

This paper addresses the particular issue as to whether students who are performing well in a unit, or perceive themselves to be performing well, are more likely to give higher scores in student evaluations, while those who are performing poorly, or perceive themselves to be performing poorly, are more likely to give low scores.

Those who administer the student evaluation system at the university, and those who use the output from the system without question, implicitly assume that the students who respond to the survey for a given unit do it with (a) an equal interpretation of each question and (b) are able and willing to answer each question dispassionately and honestly, regardless of any personal issues, beliefs or biases.

This research made use of various sources of data pertaining to a first year business statistics subject that is core to a commerce undergraduate degree at an Australian university. The data relates to students undertaking the subject during semester 2 (July – November) 2007.

The study is based on the analysis of a database with over 80 variables relating to 1,073 students. The study was able to use clear measures of student performance, in particular, assignment and exam results, and final marks and grades. Further, as the survey period was conducted over six weeks, with 24 days prior to the exam and 15 post, it was possible to check for extra indications of students' perceived performance in the evaluation scores.

LITERATURE REVIEW

There has been an enormous amount of literature based around the evaluation of post secondary teaching in general, and in particular, student evaluations of teaching. Wachtel (1998) provides an excellent overview of the history of student evaluation literature dating back to the 1920's. The authors differ greatly in their opinions and are usually polarised into those who believe student evaluations provide a reliable and valid measure of teaching performance (see Wachtel 1998; Aleamoni 1974; and Marsh 1980) and another group who have a variety of concerns including self selection bias, extraneous factors that influence ratings (Felton et al 2004; Germain and Scandura 2005; Davies et al 2007) and that students are not necessarily the best source of data to evaluate teaching performance (Simpson & Siguaw 2000),.

In the Australian context, the first responsibility for assuring the quality of its courses and programs resides with a university itself. In-house student evaluations of teaching and learning form a key part of this process and some universities have a long history of conducting such surveys. In relatively recent times, external checks have emerged which have in effect made such surveys mandatory, in particular, the Australian Graduate Survey (conducted by Graduate Careers Australia - http://www.graduatecareers.com.au/content/view/full/24 [Accessed 5 May 2008]), the Learning and Teaching Performance Fund announced in 2003 by the then Federal Government (http://www.dest.gov.au/sectors/higher_education/policy_issues_reviews/key_issues/learning_teaching/ltpf/ [Accessed 5 May 2008]) and AUQA, the Australian Universities Quality Agency, (http://www.auqa.edu.au/aboutauqa/ [Accessed 5 May 2008]) which conducts external audits on universities on a cyclical basis.

As a consequence of this, as well as the publication of results by Australian universities, a greater emphasis is being placed on student evaluations by university administrations. In some universities, performance by academics in student evaluations can be used in tenure, promotion and annual performance reviews. In theory, units or courses could be cancelled as a result of negative feedback from students. Thus, it is imperative that in any student evaluation procedures put in place by universities, the data collected are valid and representative of actual performance of academics. In particular, the influence of extraneous factors and procedural factors should always be considered.

A large proportion of the literature on student evaluations is generally favourable towards student evaluations (for example Aleamoni 1974 and Marsh 1980). Wachtel (1998) claims the '*majority of researchers believe that student ratings are a valid, reliable and worthwhile means of evaluating teaching*.'

There are, however, many authors who still question the validity and usefulness of such ratings. Of all the criticisms presented, it is the correlation between the evaluation a student gives and their performance that has the most support. A particular concern, expressed by a number of researchers is the potential for 'grade inflation'. Krautman and Sander (1999) opine that if *evaluations can be increased by giving higher grades, then they are a flawed instrument for the evaluation of teaching*. Various studies have demonstrated a positive correlation between evaluation ratings and performance. In a study of students completing a third year / fourth year compulsory security analysis subject in a finance major, Worthington (2002) found that there was a higher probability that students expecting a higher grade in the subject will assign a higher rating.

Aleamoni (1974), and references therein, provide counter-arguments, arguing that the link between performance and student evaluation ratings is only weak since any positive correlations found by researchers do not exceed 0.30. Marsh (1980) examined the relationship between student evaluations and a variety of characteristics (including perceived performance) for a data set involving over 500 courses. He concluded that although expected grade had a statistically significant effect on student ratings, the effect was not large and could further be reduced by controlling for other variables.

In the studies described above and elsewhere, the types of characteristics considered as potential influences or biases to student evaluations are typically course related (eg. class size, course level, on campus or distance mode, etc.) or student related and obtained during the evaluation survey (eg. gender, prior interest, perceived difficulty, expected performance, etc). See, for example, Marsh (1980) and Wachtel (1998). Where there does seem to be a gap in the literature is considering other characteristics which are not as easily captured as the evaluation surveys tend to be anonymous. For instance, actual performance of a student, class attendance, participation and when the survey was completed by a student, are usually not available for analysis as part of the student evaluation data. It is factors such as these, coupled with the other characteristics listed above that will be the focus of this paper.

METHODOLOGY

To capture an exhaustive list of characteristics described at the end of the literature review, we have chosen to focus on an individual case study involving just a single subject. Most contributors to the literature make use of data across several or many subjects. The reason for our choice is two-fold: firstly for a single subject is easier to manage the actual data collection allowing a greater depth of data to be collected; secondly, it also enables us to control for certain factors such as different content, teaching styles, assessment, etc.

The chosen subject for this case study is a first year business statistics subject that is a core component of a commerce undergraduate degree at an Australian university. The subject is taught across three different campuses (metropolitan city campus, regional campus and a smaller country campus) and is also available via distance education. The subject is traditionally taught in two semesters each year. However, this study only concentrates on the data related to students completing the subject in semester 2, 2007.

The student evaluation system under consideration collects student responses over a six week period from the last two weeks of the semester up to one week after the end of the final exam period. Students are notified about the survey via email as well as pop-up windows on the online teaching environment each time it is accessed during the survey period. Additional online announcements and email reminders continue on a weekly basis and persist if a student does not submit their evaluation for all the subjects in which they are currently enrolled (this can amount to five subjects in any given semester). In addition, students are offered incentives such as the chance to win book voucher prizes for submitting all their evaluations.

The survey consisted of nine questions that relate to various aspects such as teaching, quality of course, assessment, library resources, use of online technologies, plus additional questions on individual teaching staff (typically, students were taught by a lecturer and a tutor). There was also the option to provide written comments. For each question, students were asked to choose a response on a scale of 1 (strongly disagree) to 5 (strongly agree). A ‘not applicable’ (NA) option was also available for each question. Note that for this paper we will only consider the question: *This unit was well taught.*

The unique part of this research is the coupling of data from multiple sources. This is usually not possible as most student evaluations are conducted anonymously. In this particular case, permission was obtained to first match the various data sets before stripping out any identifying information. The data sources were:

- Student marks and grades for all assignment work and the final examination
- On-line conferencing activity during the semester
- Marks for individual exam questions (for the City campus students only) and some tutorial attendance data for the City campus students
- Student demographics from the university student database system (including age, fee status, citizen status, and previous study attempts)
- Survey response data for those who completed the survey (including day of submission, and scores from 1 to 5 given for each question)

After adjustments for a number of factors, including removal of four students who had completed the survey after they had dis-enrolled and the exclusion of students who were enrolled at partner institutions but not eligible for the survey, the end result was a database with over 80 variables that related to 1,073 students of which 373, or 34.8%, completed the survey.

RESULTS

The analysis is not based on a random sample, but on the group of students who self-selected to complete the survey, hence, the results are in effect just a descriptive analysis of the data for those students who self-selected.

Following are the key findings from the student evaluations of teaching to the question:

This unit was well taught

As mentioned in the previous section, students rated teaching on a five point Likert scale from Strongly Disagree (1) to Strongly Agree (5). As part of the analysis, the mean response was calculated for the entire sample as well as for various subgroups. We acknowledge that the use of the mean can be problematic with Likert scales, but the mean is the summary measure used by the University. Nevertheless we have found the mean to be an effective way to analyse the data set.

Average student evaluation rating vs performance

The mean score for all 369 valid responses from students was 3.60 with standard deviation of 1.07. However, when considering the average response across the different grades achieved by students, a clear pattern emerges (see Table 1). Generally, students who perform better are on average rating the teaching higher than the lower performing students. The mean ranges from approximately 4.2 to 3.3 across the different grades.

Grade	High Distinct.	Distinct.	Credit	Pass	Fail, not passed	Failure, no assess.	Total
Mean	4.21	3.66	3.57	3.50	3.27	3.33	3.60
N	42	77	84	107	56	3	369
SD	0.750	1.008	0.948	1.127	1.272	0.577	1.072

Table 1. Means table for Student Evaluation score vs Grade

Average student evaluation rating vs performance across different sub-categories

We now consider if the relationship between the student evaluations and performance holds across the various subgroups in the student population. For brevity, of the approximately 80 factors contained in the data set, only the most significant variables or those suggested by the literature are given here.

Table 2 shows the mean responses across the four factors: when evaluation was completed, campus, gender and nationality of student. The timing of when the evaluation was completed is particularly important to this study as it provides another mechanism for checking if a student's perceived performance affects their rating of teaching. Note that the complete Table 2 is shown in the Appendix.

		High Distinct.	Distinct.	Credit	Pass	Fail, not passed	Failure, no assess.	Total
Evaluation Completed	**Pre-Exam**	4.22	3.60	3.55	3.44	3.39	3.50	3.57
	Post-Exam	4.20	3.93	3.65	3.72	2.83	3.00	3.69
Campus	**City**	4.00	3.44	3.45	3.24	2.96	3.50	3.39
	Regional	4.50	4.28	4.10	4.03	3.43	3.00	4.06
	Country	5.00	4.75	4.67	4.33	5.00		4.70
	Distance	4.00	3.10	2.33	3.07	3.00		3.04
Gender	**Female**	4.39	3.52	3.56	3.48	3.25	3.00	3.56
	Male	4.08	3.77	3.59	3.53	3.28	4.00	3.63
Nationality	**Domestic**	4.26	3.70	3.52	3.47	3.09	3.33	3.57
	International	3.75	3.00	3.80	3.77	4.00		3.77

Table 2 (Brief). Means table for Student Evaluation score vs Grade vs Other Categories

Some important observations for the different factors:

- With some exceptions, the same general pattern detected in Table 1 prevails across all the factors.
- The timing of when a student completes the evaluation seems the most important of these secondary variables. In general, scores post-exam are higher on average than scores pre-exam – perhaps indicating that students felt they had done well enough. The notable exception is the Fail group, where there evaluations scores were lower on average – perhaps indicating that they felt they had not done well.
- There are a number of differences across the various campuses. Higher performing distance students tend to give on average lower scores than those given by on campus students. The country campus has high scores across all grades (but this may be accounted for by other factors such as class size).
- Gender does not seem to be a significant factor, with males and females giving similar ratings.
- Domestic high performing students, on average, give a higher score than international students. The situation is reversed for lower performing students with international students generally giving higher ratings than domestic students.

The effect on when student completes the survey

One of the unique parts of this case study is that students can complete the evaluation survey over a six week period which falls either side of when the final examination is sat but before results are released. Of those who submitted, 281 (76.2%) did so in the weeks before the exam. The remaining 88 (23.8%) students, having completed the survey after the examination, may have had a stronger perception of how they performed overall in the unit. Table 3 demonstrates that this seemed to have occurred. In this section, we delve further to see if the effect is common across other factors as well.

Table 3 shows the effect of the timing of the evaluation across the campuses. On the regional campus, students who achieved a Pass grade gave an average of 4.44 (post exam) compared to 3.86 (pre-exam). This may suggest a relief factor from students who perhaps were not expecting to pass going into the exam. Also noteworthy on the City campus was the Fail group who gave a lower average of 2.6 (post exam) versus 3.05 (pre exam).

	PRE EXAM	High Distinct.	Distinct.	Credit	Pass	Fail, not passed	Failure, no assess.	Total
Campus	City	4.00	3.42	3.39	3.20	3.05	3.50	3.36
	Regional	4.50	4.15	4.23	3.86	3.30		4.00
	Country	5.00	4.75	5.00	4.75	5.00		4.88
	Distance	4.00	3.00	2.60	3.09	3.38		3.09
	Total	4.22	3.60	3.55	3.44	3.39	3.50	3.57
	POST EXAM	**High Distinct.**	**Distinct.**	**Credit**	**Pass**	**Fail, not passed**	**Failure, no assess.**	**Total**
Campus	City	4.00	3.56	3.73	3.40	2.60		3.52
	Regional	4.50	4.60	3.86	4.44	3.75	3.00	4.20
	Country	5.00		4.00	3.50			4.00
	Distance	4.00	4.00	1.00	3.00	2.00		2.92
	Total	4.20	3.93	3.65	3.72	2.83	3.00	3.69

Table 3. Means table for Student Evaluation across different campuses, grades and pre/post exam.

In Table 4, gender is shown for pre and post exam. For the cohort of students who failed, those completing the evaluation pre-exam gave on average higher evaluations than those post-exam for both males and females. The average female rating fell from 3.33 to 2.67 and Males from 3.43 to 2.89. Male students who achieved a Pass grade and completed the survey post-exam gave an average of 4.00 (which is the same as the males who submitted post-exam and who achieved a HD).

	PRE EXAM	**High Distinct.**	**Distinct.**	**Credit**	**Pass**	**Fail, not passed**	**Failure, no assess.**	**Total**
Gender	**Female**	4.44	3.41	3.59	3.48	3.33	3.00	3.53
	Male	4.11	3.74	3.48	3.38	3.43	4.00	3.60
	Total	4.22	3.60	3.55	3.44	3.39	3.50	3.57
	POST EXAM	**High Distinct.**	**Distinct.**	**Credit**	**Pass**	**Fail, not passed**	**Failure, no assess.**	**Total**
Gender	**Female**	4.33	4.00	3.33	3.50	2.67	3.00	3.67
	Male	4.00	3.89	3.79	4.00	2.89		3.71
	Total	4.20	3.93	3.65	3.72	2.83	3.00	3.69

Table 4. Means table for Student Evaluation score across gender, grades and pre/post exam.

Lastly Table 5 compares domestic and international students. The most interesting result is for domestic students in the Fail category: they gave a lower on average rating of 2.70 (post-exam) compared to 3.20 (pre-exam)

	Pre Exam	**High Distinct.**	**Distinct.**	**Credit**	**Pass**	**Fail,not passed**	**Fail, no assess**	**Total**
Nationality	**Domestic**	4.25	3.64	3.52	3.44	3.20	3.50	3.55
	International	4.00	3.00	3.67	3.40	4.11		3.66
	Total	4.22	3.60	3.55	3.44	3.39	3.50	3.57
	Post Exam	**High Distinct.**	**Distinct.**	**Credit**	**Pass**	**Fail,not passed**	**Fail, no assess.**	**Total**
Nationality	**Domestic**	4.29	3.93	3.53	3.55	2.70	3.00	3.63
	International	3.00		4.33	5.00	3.50		4.22
	Total	4.20	3.93	3.65	3.72	2.83	3.00	3.69

Table 5. Means table for Student Evaluation score across domestic and international students.

Regression analysis

Given that performance appears to be a factor, we undertook regression analysis to determine if there were non-performance factors that could also provide further explanation of variation in teaching evaluation scores (see Table 6). We used ordinary least squares and constructed numerous alternative models across the 80 factors contained in the data set. The model below was chosen because of its simplicity, R^2, and also because some of the variables were suggested by the literature. We note however, that regression analysis on this type of data, based on self-selection is questionable and thus should be treated with caution.

Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.523(a)	.274	.262	.925

a Predictors: (Constant), Units failed:Pre S207, Country_Campus, Regional_Campus, Ass 1: 100, Years at University, Mark Final:100
b Dependent Variable: Q.1 Teaching

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	115.583	6	19.264	22.522	.000(a)
	Residual	306.214	358	.855		
	Total	421.797	364			

a Predictors: (Constant), Units failed:Pre S207, Country_Campus, Regional_Campus, Ass 1: 100, Years at University, Mark Final:100
b Dependent Variable: Q.1 Teaching

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	1.915	.255		7.500	.000		
	Regional_Campus	.714	.112	.293	6.391	.000	.963	1.038
	Country_Campus	1.256	.219	.266	5.734	.000	.943	1.061
	Ass 1: 100	.010	.004	.154	2.559	.011	.561	1.782
	Mark Final:100	.015	.004	.211	3.507	.001	.561	1.784
	Years at University	-.201	.049	-.211	-4.130	.000	.780	1.282
	Units failed:Pre S207	.081	.030	.151	2.745	.006	.672	1.487

a Dependent Variable: Q.1 Teaching

Table 6. Regression analysis of teaching evaluation vs other variables.

The overall model is significant $F(6,358) = 22.522$, $p < .001$. 27.4% of the variation in teaching scores can be explained by the linear model.

In the model, Regional and Country campuses are the most important predictors of Teaching Scores (Beta = .293 and .266 respectively) indicating that non-performance variables may also explain variation in evaluation scores.

As expected, Assignment 1 Mark and Final Mark (both measures of performance) also have significant positive relationships with Teaching Scores.

CONCLUSION

Past literature on student evaluations has indicated that student perceived performance has an important, positive effect on student evaluations. This study has taken the measure of performance further by using measures of actual performance as well as indications of perceived performance.

The analysis is not based on a random sample, but on the group of students who self-selected to complete the survey. The matter of self-selection versus random selection is an area of study in its own right, and this paper in effect provides just a descriptive analysis of the data for those students who self-selected. This paper found that students who are performing well in a unit, or perceive themselves to be performing well, are more likely to give higher scores in student evaluations, while those who are performing poorly, or perceive themselves to be performing poorly, are more likely to give low scores.

This positive relationship (which has been observed in a number of previous studies) is also evident across a number of background variables, with few exceptions. Further, the study found that, in general, students who submitted their evaluations after sitting the exam gave different ratings to those who submitted before.

In concluding, what does this relationship between teaching evaluation and performance mean for the validity of these types of student evaluations in general? We cannot answer this question directly, however we can speculate on the reasons for the relationship:

1. Students may not have the same interpretation of the question. This may be because, as some authors suggest, students do not have the relevant background/experience to make such judgements and thus may not necessarily be the best to critic teaching. (Alternatively the questionnaire itself may be ambiguous and more guidelines for students are needed).
2. Students do not answer each question dispassionately and honestly, regardless of any personal issues, beliefs or biases.
3. Tht students have accurately judged when he/she has understood well (and performed well) and hence recognised that this was attributable to good teaching. Likewise, for poor learning and poor teaching.

If one of the first two scenarios is indeed true then this should be of great concern to university administrators who conduct and make decisions based on these types of surveys. If the third outcome is true, then it gives some support that the results of such surveys are valid and meaningful. From this study, we cannot identify which if any of these three scenarios is likely to be true. In reality, we suspect, that the picture is not clear cut and aspects of all three likely to apply. Indeed, it is our intention in future research to investigate this particular aspect. We also intend to extend our analysis across different subjects (and hopefully also different universities).

ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for their feedback. Particularly for the comments about the subtlety of the relationship between performance and evaluation.

REFERENCES

Aleamoni, L. M. (1974). Typical faculty concerns about student evaluation of instruction. *Symposium on methods of improving university teaching at the Technion*, Israel Institute of Technology, Haifa, Israel 24 March 1974

Davies, W. M., Hirschberg, J., Lye, J. Johnston, C. and McDonald, I. (2007) Systemic Influences on Teaching Evaluations: The Case for Caution. *Australian Economic Papers*, 46(1) 18-38

Felton, J., Mitchell, J., and Stinson, M. (2004) Web-based student evaluation of professors: the relations between perceived quality, easiness and sexiness, Assessment and Evaluation in Higher Education, 29(1), 91-108

Germain, M. and Scandura T. A. (2005) Grade Inflation and student individual differences as systematic bias in faculty evaluations. *Journal of Instructional Psychology*, Vol. 32

Krautmann, A. C. and Sander, W. Grades and student evaluations of teachers. *Economics of Education Review* 18 (1999) 59–63

Marsh, H. W. (1980) The Influence of Student, Course, and Instructor Characteristics in Evaluations of University Teaching. *American Educational Research Journal* 17(2), 219-237

Nelson, B. (2004), Promoting Excellence in University Teaching, *Dept. of Education Science and Training* http://www.dest.gov.au/Ministers/Media/Nelson/2004/04/nmin675060404.asp (accessed 6 May 2008)

Simpson, P. M. and J. A. Siguaw (2000), Student evaluations of teaching: an exploratory study of the faculty response, *Journal of Marketing Education, 22*(3), 199-213

Wachtel, H. K. (1998). Student Evaluation of College Teaching Effectiveness: A brief review. *Assessment and Evaluation*, 23 (2), 191-211

Worthington (2002) The Impact of Student Perceptions and Characteristics on Teaching Evaluations: a case study in finance education, *Assessment and Evaluation of Higher Education*, 27(1) 49-64

APPENDIX

		Grade	High Distinct.	Distinct.	Credit	Pass	Fail, not passed	Failure, no assess.	Total
When Evaluation Completed	Pre-Exam Evaluation	Mean	4.22	3.60	3.55	3.44	3.39	3.50	3.57
		N	27	62	64	82	44	2	281
		SD	0.751	1.063	0.942	1.123	1.262	0.707	1.077
	Post-Exam Evaluation	Mean	4.20	3.93	3.65	3.72	2.83	3.00	3.69
		N	15	15	20	25	12	1	88
		SD	0.775	0.704	0.988	1.137	1.267	.	1.054
Campus	City Campus	Mean	4.00	3.44	3.45	3.24	2.96	3.50	3.39
		N	22	45	55	55	26	2	205
		SD	0.816	0.918	0.812	1.138	1.216	0.707	1.012
	Regional Campus	Mean	4.50	4.28	4.10	4.03	3.43	3.00	4.06
		N	14	18	20	31	14	1	98
		SD	0.519	0.575	0.788	0.836	0.938	.	0.810
	Country Campus	Mean	5.00	4.75	4.67	4.33	5.00		4.70
		N	2	4	3	6	5		20
		SD	0.000	0.500	0.577	0.816	0.000		0.571
	Distance Student	Mean	4.00	3.10	2.33	3.07	3.00		3.04
		N	4	10	6	15	11		46
		SD	0.816	1.370	1.211	1.223	1.483		1.299
Gender	Female	Mean	4.39	3.52	3.56	3.48	3.25	3.00	3.56
		N	18	33	43	62	24	2	182
		SD	0.698	1.121	1.007	1.112	1.359	0.000	1.115
	Male	Mean	4.08	3.77	3.59	3.53	3.28	4.00	3.63
		N	24	44	41	45	32	1	187
		SD	0.776	0.912	0.894	1.160	1.224	.	1.031
Nationality	Domestic	Mean	4.26	3.70	3.52	3.47	3.09	3.33	3.57
		N	38	73	69	94	45	3	322
		SD	0.724	1.009	0.964	1.133	1.328	0.577	1.095
	International	Mean	3.75	3.00	3.80	3.77	4.00		3.77
		N	4	4	15	13	11		47
		SD	0.957	0.816	0.862	1.092	0.632		0.890

Table 2 (Complete). Means table for Student Evaluation score vs Grade vs Other Categories

INVESTIGATING SAMPLE BIAS IN STUDENT EVALUATIONS OF TEACHING

M. Brookes, R. Boyle, R. Braithwaite, J. Mustard, D. Saundage and H. Short
Deakin University, Australia

Abstract
Student evaluations of teaching have increased in importance to universities in Australia over recent years due to changes in government policy. There has been significant debate in the literature as to the validity and usefulness of such evaluations and as to whether students who respond to the evaluations are indeed representative of the student population. A potential invalidating issue is self selection in the evaluation process. In this paper, we consider student evaluations of a large first year business statistics subject that had 1073 eligible students enrolled across four campuses at the time of the evaluation. The study is based on the 373 students (34.8%) who responded to the survey, and their final results. The evaluations were open for a period of six weeks leading up to and just after the final exam. The study looks in detail at the student population identifying such attributes as gender; home campus; course of study; domestic/international; Commonwealth Supported Place/full fee paying, etc. and then mapping these results to those of the students who responded to the survey.

INTRODUCTION

The rationale for this paper stems from concerns by academics about potential defects in the university's student evaluation system. This paper addresses the particular concern that students who self-select to complete the survey for a given subject, do not form a representative group with respect to the overall student cohort for that subject, either in regard to demographics, and/or in regard to their views about the quality of teaching and learning.

Those who administer the student evaluation system at the university and those who use the output from the system without question, implicitly assume that the students who respond to the survey for a given view are representative in terms of demographics and views and opinions.

The research team arranged access to sources of data pertaining to a first year business statistics subject that is a core component of a commerce undergraduate degree at an Australian university. The subject is traditionally taught in two semesters each year. However, this study only concentrates on the offerings in semester 2, 2007 when it was taught in three campus locations in face-to-face mode, plus another cohort off-campus (ie. distance education). (The unit is also taught at a number of partner institutions in Australia and overseas, however this survey does not include these partner institutions.)

LITERATURE REVIEW

In the higher education sector, there has been an increasing pressure from both within Universities (Hester, 2008) as well as governments (Nelson, 2004) to conduct student evaluations of teaching. Generally there are three reasons for universities to conduct student evaluation surveys:

- to improve the quality of teaching through constructive feedback,
- using the results in staff appraisals (ie. for promotion, decisions concerning tenure and contract renewal) (Hester, 2008), and
- explicit requirements (such as government mandated reporting systems linked to funding) or an implicit obligation felt by university administrations (Davies et al, 2007; Kember et al 2002; Nelson 2004).

However, there are potential methodological problems in the conduct of some student evaluation surveys primarily due to the nature of the survey responses being based on self selection rather than a random sample (Heckman, 1979). Heckman also discusses how self selection in surveys leads to a bias in the sample.

A further confounding factor is the belief that students are not the ideal source of data to evaluate teaching performance (Simpson & Siguaw 2000). The validity of the student rating now has been well established (Marsh 1984; Arubayi 1987) but the focus has shifted to understanding the background characteristic factors that are beyond the instructor (Worthington 2002).

Demographic data of the student profile (eg. gender, age, ethnicity, etc.) and the course profile (eg. entry qualification, assessment results, units completed so far) helps in determining how far student perceptions and experiences vary between cohorts of students and helps establish whether responses are representative of the student population (Brennan & Williams 2004; Davies et al. 2007). This study attempts to cross-check the respondents against the key demographics.

Whilst student evaluations have received a great deal of attention in the literature, this study differs in two key respects. First, the majority of the studies have been conducted in the United States (Davies et al. 2007) and second, the approach for this research was a case study about a single subject rather than across many subjects and, accordingly, a great deal of depth in the data was achieved.

METHODOLOGY

The student evaluation system under consideration collects student responses over a six week period from the last two weeks of the semester up to one week after the end of the final exam period. Students are notified about the survey via email as well as pop-up windows on the online teaching environment each time it is accessed during the survey period. Additional online announcements and email reminders continue on a weekly basis and persist if a student does not submit their evaluation for all the subjects in which they are currently enrolled (this can amount to five subjects in any given semester). In addition, students are offered incentives such as the chance to win book voucher prizes for submitting all their evaluations.

The survey consisted of nine questions that relate to the subject, plus additional questions on individual teaching staff (typically, students were taught by a lecturer and a tutor). There was also the option to provide written comments. For each question, students were asked to choose a response on a scale of 1 (strongly disagree) to 5 (strongly agree). A 'not applicable' (NA) option was also available for each question.

At the end of the survey period, the results are analysed by the university and then published. The published reports show mean responses for each question, standard deviations, and response rates.

For this study, the research team compiled data from the following sources:

- Student marks and grades for all assignment work and the final examination
- On-line conferencing activity during the semester
- Marks for individual exam questions (for the city campus students only) and some tutorial attendance data for the city campus students
- Student demographics from the university student database system (including age, fee status, citizen status, and previous study attempts)
- Survey response data for those who completed the survey (including day of submission, and scores from 1 to 5 given for each question)

The survey response data were merged with data provided by the university planning department who returned a final database to the researchers with all student identifying data (such as IDs) removed.

After adjustments for a number of factors, including removal of four students who had completed the survey after they had disenrolled and the exclusion of students who were enrolled at partner institutions but not eligible for the survey, the end result was a database with over 80 variables that related to 1,073 students of which 373, or 34.8%, completed the survey.

RESULTS

There are a number of attributes that could be examined to determine whether or not the collected sample adequately represents the underlying student population in the student evaluation survey that we are analysing. For our study, we have concentrated on student gender, campus of study, faculty, final grade, age and whether the student is a domestic or international student. The analysis is based on cross tabulations of the attributes in question against the response numbers (either *yes* or *no* to whether they responded to the survey). The cross tabulation is then subject to a chi-square test of independence between each attribute and the response proportions. For each attribute a further cross tabulation and chi-square test against the students' results for the survey question on teaching quality is used to evaluate whether the lack of balance in the representation results in a significant bias in the overall results.

A. GENDER

As shown in table 1, the gender balance of the students enrolled in the subject and eligible to respond (ie. the population to be sampled) is approximately 60% male and 40% female. If valid random sampling techniques were used, then we would expect that the gender balance of the sample (ie. the group of students that responded) would be close to that of the population and, by implication, also be similar to the gender balance of the group of students that did not respond. We found that the gender balance of those that did participate in the student evaluation was approximately 51% male and 49% female. The difference between sample proportions and population proportions suggests that the sample is not truly representative with respect to actual gender balance.

Overall, the response rate for female students was 42%, whereas the response rate for males was only 30%. Once again, this suggests a disparity between the response rates of males and female students. Female students are more likely to respond yet the student population in question contains a higher proportion of males.

Frequency Count:	**Responded to Survey**			**Unit was well taught**			
Gender -	**Yes**	**No**	**Total**	**Disagree**	**Neutral**	**Agree**	**Total**
F	182	249	431	30	44	108	182
M	191	451	642	25	44	118	187
Total	373	700	1073	55	88	226	369

Table 1: Two variable cross tabulation of gender against response count and evaluation of teaching quality

A chi-square test of independence (at a significance level of 5%) between gender and response proportions shows that the response rate is dependent on gender (p-value was 0.00003). Therefore the conclusion can be drawn that the sample is not representative of the population from a gender perspective.

Of the students that responded to the survey, 4 did not submit a response to the specific question on teaching quality that we have analysed hence the total shown in table 1 for whether the unit was well taught is less than the total number of students that responded to the survey. While the response rate is dependent on gender, a chi-square test of independence (at a significance level of 5%) between gender and the evaluation of teaching quality shows that the teaching quality result is not dependent on gender (p-value was 0.66053), thus no significant bias will result from females being over represented in the sample.

B. CAMPUS COHORT

The survey included four separate, geographically dispersed campuses of which one is classified as off campus (or distance education). The number of responses and non-responses for each campus is shown in table 2. The campuses vary significantly in size with the city campus accounting for 60% of the students, the regional campus 26%, the distance education campus 11% and the country campus 3%.

Frequency Count:	Responded to Survey			Unit was well taught			
Campus	**Yes**	**No**	**Total**	**Disagree**	**Neutral**	**Agree**	**Total**
City	206	439	645	39	55	111	205
Regional	100	176	276	3	17	78	98
Distance	47	68	115	13	15	18	46
Country	20	17	37	0	1	19	20
Total	373	700	1073	55	88	226	369

Table 2: Two variable cross tabulation of campus against response count and evaluation of teaching quality

While the overall response rate was 35%, the city rate was 32%, regional was 36%, distance was 41% and country was 54%. One potential explanatory variable that is worthy of further analysis in the future is the size of the cohort on each campus as it appears as if the response rate increases as the cohort size decreases.

A chi-square test of independence between the response rates for each campus (at a significance level of 5%) shows that the response rate was dependent on the cohort or campus (p-value was 0.01479). Since the campus of study has a marked impact on the response rate, the implication is that higher response rates from campuses with smaller cohorts will bias the overall results.

Given that the response rate is dependent on campus, a chi-square test of independence (at a significance level of 5%) between campus and the evaluation of teaching quality shows that the teaching quality result was dependent on campus (p-value was 0.00000), thus inferring that the potential for bias due to campus differences in response rate is significant.

C. FACULTY

Students enrolled in the subject in question come from undergraduate degrees that originate from four faculties plus a very small number from outside the university (see table 3). Two of these faculties account for 95% of the students.

The response rate across the faculties is reasonably consistent with the overall response rate and this is supported by a chi-square test of independence between the response rates for each faculty (at a significance level of 5%) which indicates that the response rate is indeed not dependent on the faculty (p-value was 0.84819).

Frequency Count:	Responded to Survey		
Faculty	**Yes**	**No**	**Total**
A	49	100	149
B	302	568	870
C	9	12	21
D	11	18	29
No Faculty	2	2	4
Total	373	700	1073

Table 3: Two variable cross tabulation of faculty against response count

D. GRADE ACHIEVED

Table 4 lists the response numbers against final grades achieved by the students. The overall response rate to the survey was 35%, however, the response rate for students that passed the subject was 42% compared with only 19% for students that failed. This discrepancy is further highlighted when we further analyse the responses by final grade received. Of the students who received a High Distinction, 69% responded to the survey. Of the students who received a Distinction, 57% responded. The response rates for Credit and Pass grades were 39% and 32% respectively.

Frequency Count:	**Responded to Survey**			**Unit was well taught**			
Final Grade	**Yes**	**No**	**Total**	**Disagree**	**Neutral**	**Agree**	**Total**
High Distinction	43	19	62	0	8	34	42
Distinction	79	59	138	12	14	51	77
Credit	84	129	213	12	21	51	84
Pass	108	234	342	18	29	60	107
Fail	56	195	251	13	14	29	56
Did not sit exam	3	64	67	0	2	1	3
Total	373	700	1073	55	88	226	369

Table 4: Two variable cross tabulation of final grade against response count and evaluation of teaching quality

A chi-square test of independence between the response rates for each final grade (at a significance level of 5%) shows that response rate is dependent on level of performance (p-value was 0.00000). When final grade is broken down into its components (ie. the assignment and exam marks upon which the final grade is based), the same dependence on level of performance exists for each component although to varying degrees.

The conclusion that can be drawn here is that the higher the final grade awarded, the more likely a student will respond to the survey. This indicates that the sample is not representative of the population when the student's level of achievement in the subject is taken into consideration.

Given that the response rate is dependent on grade achieved, a chi-square test of independence (at a significance level of 5%) between grade and the evaluation of teaching quality shows that the teaching quality result was dependent on grade (p-value was 0.04623). This infers that the potential for a student to achieve a high grade did influence how that student evaluated the teaching quality and thus if higher grades are over represented this also introduces bias into the evaluation of teaching quality. Students that achieve higher grades are more likely to respond and also more likely to rate teaching quality highly thus potentially inflating the outcome for teaching quality evaluation.

E. AGE

For analysis purpose, student ages were grouped into three ranges; 17 to less than 27, 27 to less than 37 and 37 and above (see table 5). These ranges were chosen because they represented decade splits starting from the youngest student's age. The top two decades were then combined because of the low frequencies involved. Of the students less than 27 years of age, only 31% participated in the survey. For those students aged between 27 and less than 37, this rate increased to 46%, while for students of 37 years and older, 79% completed the survey.

Frequency Count:	Responded to Survey			Unit was well taught			
Age	**Yes**	**No**	**Total**	**Disagree**	**Neutral**	**Agree**	**Total**
17 to < 27	333	663	996	46	74	209	329
27 to < 37	29	34	63	6	10	13	29
37 to 57	11	3	14	3	4	4	11
Total	373	700	1073	55	88	226	369

Table 5: Two variable cross tabulation of age against response count and evaluation of teaching quality

A chi-square test of independence between age and response rate (at a significance level of 5%) shows that the student evaluation response rate is dependent on age (p-value was 0.00031). The implication here is that the two subsets of the population (respondents versus non-respondents) are not homogenous. Participation in the student evaluation is biased towards older students and thus the sample is not representative of the population since the older the student, the more likely they are to respond to the survey.

While the response rate is dependent on age, a chi-square test of independence (at a significance level of 5%) between age group and the evaluation of teaching quality shows that the teaching quality result is not dependent on age (p-value was 0.13658), thus no significant bias will result due to older students being more likely to respond.

F. DOMESTIC/INTERNATIONAL

The response rate comparison between domestic and international students only shows a marginal difference between the two, with 36% of domestic students responding compared to 29% of international students (see table 6).

Frequency Count:	Responded to Survey		
Domicile status	**Yes**	**No**	**Total**
Domestic	326	587	913
International	47	113	160
Total	373	700	1073

Table 6: Two variable cross tabulation of domicile status against response count

A chi-square test of independence between domestic/international status and response rate (at a significance level of 5%) shows that the response rate is independent of whether a respondent is a domestic or international student (p-value was 0.12083). This indicates that, while there appears to be a slight relationship, it is insufficient to conclude that being a domestic or international student has any effect on the response rate.

CONCLUSION

Past literature on student evaluations has indicated that the results are an important part of university management and decision making. Recent government edicts in Australia also link such issues as funding, staff tenure and promotion to student evaluation outcomes hence accuracy and accountability are of paramount importance. This study has examined the demographics of a sample derived through student self selection and compared it with the demographics of the underlying student population that it is trying to characterise, to see if the sample is truly representative.

Analysis of a database that contains over 80 variables relating to 1,073 students showed:

- The sample group was not representative of the demographics of the student group.

- Higher performing students, female students and students from smaller campuses were over-represented.
- The views expressed by students at two of the campuses were out of phase when compared to the overall result despite the use of the same basic course materials and assessment methods.
- The diverse range of responses also raises the question as to whether or not students really understand each question (eg. what does "Well taught" mean?).

Our analysis of the dataset in question suggests that the sample is clearly not representative. The relationships that we have found are strong and raise concerns about the administration, procedural aspects and interpretation of student evaluations of teaching if similar relationships exist in other samples.

Of the factors considered in this study, campus and grade achieved each showed relationships with both response rate and the result for evaluation of teaching quality thus introducing significant bias into the assessment of teaching quality. However, neither gender nor age showed any relationship with the evaluation of teaching quality hence neither are likely to introduce significant bias.

If such relationships do exist in other samples, then our finding also raises concerns about how worthwhile student evaluations are in measuring teaching effectiveness if the sample is not representative of the population and hence potentially introduces biases into inferences based on the evaluation scores. This then raises questions as to whether the results of a student evaluation of a subject can be trusted for important decision making purposes. University administrators need to be made aware of potential systematic bias in student evaluation outcomes and procedures for administration and interpretation need to be evaluated, and perhaps changed, in order to address imbalances in the demographics of the student responses.

In addition to the potential existence of bias in the sample, we have also discovered other potential issues that require further analysis including the possibility of *satisficing behaviour patterns* in survey responses, the potential for misinterpretation of the broadly worded questions that are posed in the surveys, and the impact of when the survey response takes place given that students can respond either before or after the final examination.

ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for their feedback, and especially their comments on the likelihood of generating an inaccurate picture of the actual quality of teaching.

REFERENCES

Arubayi, Eric A. (1987). Improvement of Instruction and Teacher Effectiveness: Are Student Ratings Reliable and Valid? Higher Education 16: 267-78

Brennan, J. and Williams, R., (2004) Collecting and using student feedback. A guide to good practice. *Learning and Teaching Support Network*, Centre for Higher Education Research and Information

Davies, W. M., Hirschberg, J., Lye, J. Johnston, C. and McDonald, I. (2007) Systemic Influences on Teaching Evaluations: The Case for Caution. *Australian Economic Papers*, 46(1) 18-38

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* 47(1) 153-161

Hester, E. J. (2008). Student evaluations of advising: Moving beyond the mean. *College Teaching* Winter 2008 56(1) 35-38

Kember, D., D. Leung and K. Kwan, (2002), Does the Use of Student Feedback Questionnaires Improve the Overall Quality of Teaching?, *Assessment & Evaluation in Higher Education*, (27)5, 411-425

Marsh, H. W. (1980) The Influence of Student, Course, and Instructor Characteristics in Evaluations of University Teaching. *American Educational Research Journal* 17(2), 219-237

Nelson, B. (2004), Promoting Excellence in University Teaching, *Dept. of Education Science and Training* http://www.dest.gov.au/Ministers/Media/Nelson/2004/04/nmin675060404.asp (accessed 6 May 2008)

Simpson, P. M. and J. A. Siguaw (2000), Student evaluations of teaching: an exploratory study of the faculty response, *Journal of Marketing Education*, *22*(3), 199-213

Worthington (2002) The Impact of Student Perceptions and Characteristics on Teaching Evaluations: a case study in finance education, *Assessment and Evaluation of Higher Education*, 27(1) 49-64