OZCOTS 2010

Proceedings of the 7th Australian Conference on Teaching Statistics

Editors:

Helen MacGillivray
Queensland University of Technology

Brian Phillips
Swinburne University

Fremantle, Western Australia

9-10 December, 2010

Satellite to ASC 2010 Australian Statistical Conference
Fremantle, 6-9 December, 2010

SSAI
Statistical Society of Australia Inc.

ISBN: 978-0-9805950-1-7

CONTENTS

PREFACE

OZCOTS 2010

7th Australian Conference on Teaching Statistics

OZCOTS 2010 theme: Building capacity in statistics education.

Statistical education is of vital significance to all statisticians and the statistical profession across the spectrum of educational levels and disciplines. The supply of statisticians is just one aspect of this spectrum, and the education of future consumers, users, producers, developers and researchers of statistics is both challenging and important for a modern information society and hence to the statistical profession. The OZCOTS 2010 program included keynote and contributed papers and forum discussion on topics across the statistical education spectrum of interest to statisticians and the statistical profession.

OZCOTS background

The first OZCOTS was run in 1998 by Brian Phillips with papers by the Australian speakers from the 5th International Conference on Teaching Statistics (ICOTS) which had been held in Singapore earlier in 1998. Its success in bringing together Australians involved in teaching statistics resulted in Brian and his Melbourne colleagues organising annual OZCOTS gatherings from 1999 to 2002. In 2006 Helen MacGillivray was awarded one of the first Australian Learning and Teaching Council's Senior Fellowships, with her fellowship programme to run throughout 2008. As part of her fellowship programme, Helen revived OZCOTS with Brian's help, and ran it as a two-day satellite to the 2008 Australian Statistical Conference (ASC), with a one-day overlap open to all ASC delegates who could also choose to register for the second day. The OZCOTS 2008 invited speakers were all funded as part of Helen's fellowship. OZCOTS 2008 was modelled on the successful International Association for Statistical Education's (IASE) satellite conferences to ISI Conferences, with papers in proceedings and an optional refereeing process offered to authors. The success of OZCOTS 2008 lead to OZCOTS 2010 again being run as a two-day satellite to the 2010 Australian Statistical Conference, with a one-day overlap open to all ASC delegates who could also choose to register for the second day, and with a joint keynote speaker on statistics education on the overlap day.

OZCOTS 2010 acknowledges the support of the Australian Statistical Conference 2010, SAS and the NSW and Vic Branches of SSAI.

OZCOTS 2010 Conference Committee

Helen MacGillivray (joint chair, joint editor), Queensland University of Technology
Brian Phillips (joint chair, joint editor), Swinburne University
Alexandra Bremner (local arrangements), University of Western Australia

OZCOTS 2010 gratefully acknowledges the following referees for their assistance: Michael Bulmer, Rosemary Callingham, Mike Forster, Glenda Francis, Ian Gordon, Eric Sowey, Christine McDonald, Katie Makar, Michael Martin, Peter Martin, Peter Petocz, Matt Regan, Jackie Reid, Richard Wilson, Therese Wilson.

Brief Biographies of Keynote Speakers

Professor Delia North, University of Kwa-Zulu Natal, South Africa
http://statsactsci.ukzn.ac.za/north5844.aspx
Delia has over 25 years experience in teaching statistics across disciplines and levels in universities. Over the years she has become increasingly interested in statistics education, which resulted in her being appointed as the local chair of the 6th International Conference on Teaching Statistics (ICOTS6) which was held in Cape Town, South Africa, in July 2002. She has been a member of the Executive Committee of the South African Statistical Association (SASA) for more than ten years and has been chair of the Education Committee of SASA since 2003. Delia is very actively involved with voluntary work, conducting various outreach activities relating to the introduction of Statistics into the school syllabus for the first time ever in South Africa. She is known internationally for her work in supporting South African teachers.

Professor Chris Wild, Department of Statistics, University of Auckland, New Zealand. http://www.stat.auckland.ac.nz/~wild
Professor of Statistics at the University of Auckland, New Zealand and recognised by Fellowships of the American Statistical Association and the Royal Society of New Zealand, Chris Wild is a member of a rare crossover species. He publishes extensively in statistical methodology, particularly on response-selective and missing data problems, but also works substantively in statistics education. He co-wrote the Wiley books *Nonlinear Regression* (1989) and *Chance Encounters* (2000) with George Seber. His best known statistics education paper is *Statistical Thinking in Empirical Enquiry* with Maxine Pfannkuch (1999, International Statistical Review). Chris's interests in statistics education include curricular revolution at school levels, growing university statistics programmes, and improving the penetration, quality and practical impact of statistics education at all levels. Chris has been a Council member of the International Statistical Institute, President of the International Association for Statistics Education and an Associate Editor of the International Statistical Review, Biometrics, the Statistics Education Research Journal, and ANZJS. 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 teaching award in 2003. His keynote addresses include the Royal Statistical Society, the Statistical Society of Canada, and ICOTS.

Paper Refereeing Process

Papers referred to in the proceedings as refereed publications were reviewed and accepted as meeting the requisite standards by at least two referees selected from a panel of peers approved by the OZCOTS 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 the accepted norms for reporting of research. It was expected that each accepted paper would represent a significant contribution to advancement of statistics education and/or the research processes in statistical education. Authors verified that the refereed published papers for these proceedings were substantially different from papers that have been previously published elsewhere.

Program

Thursday, 9th December

8.00 am	Registration		
9.15 am	**Keynote**	**Chris Wild**	Chair: Helen MacGillivray
		What I see is not quite the way it really is	
10.00 am	Morning Tea		
Contributed papers: Graduates across disciplines			Chair: David Griffiths
10.30 am		Sue Finch	*Statistical consulting with post-graduate students*
10.50 am		Glenys Bishop	*A model for building statistical capacity among Science research students*
11.10 am		Kristen Gibbons	*Training for statistical communication in the workplace*
11.30 am		Sharleen Forbes	*Data visualisation: a new statistical literacy tool for statistical offices*
11.50 am		Kay Lipson	*Comparisons between marketing and psychology students in learning statistics*
12.10 pm	Discussion		
12.30 pm	Lunch		
Contributed papers: Learning strategies			Chair: Richard Wilson
1.30 pm		Ian Gordon	*RealStat: from ideas to data and beyond*
1.50 pm		Graham Barr	*Embedding the teaching of first year statistics in a spreadsheet environment*
2.10 pm		David Griffiths	*Revisiting the misused, misunderstood and unloved stem and leaf plot*
2.30 pm		Michael Bulmer	*Introductory epidemiology with a virtual population*
2.50 pm		Helen MacGillivray	*National statistical curriculum journeys*
3.10 pm	Afternoon Tea		
3.40 pm	ASC Panel		
4.30 pm	ASC Awards		
5.30 pm	ASC Close		
5.35 pm	ASC Happy Hour		
7.00 pm	OZCOTS dinner		

Friday 10th December

8.15 am	Registration		
8.30 am	**Keynote**	**Delia North**	Chair: Brian Phillips
		Transforming statistics education in South Africa	
9.15 am	Forum with panel		Chair: Brian Phillips
		Opportunities and challenges in designing statistics curricula	
10.00 am	Morning Tea		
Contributed papers: Technology and statistics learning			Chair: Ian Gordon
10.30 am		Doug Stirling	*Dynamic diagrams for teaching design and analysis of experiments*
10.50 am		Chris Wild	*A free stand-alone system for statistical data analysis written in R*
11.10 am		James Baglin	*An experimental study comparing strategies of learning how to use statistical software packages in introductory statistics courses*
11.30 am		Martin Gellerstedt	*WWW means win win win in education*
11.50 am		Carey Biggs	*Genstat for Teaching*
12.10 pm	Discussion		
12.30 pm	Lunch	Demonstration of Genstat for Teaching	
Contributed papers: Introductory and transition			Chair: Sue Finch
1.30 pm		Maureen Townley-Jones	*The journey of math diagnostic testing for statistics courses in Australian Universities*
1.50 pm		Isnandar Slamet	*The improvement of learning process of statistical mathematics I subject through students teams achievement divisions (STAD) using English as medium language*
		Maureen Morris	*'She Ca'nt do sums a bit' or can she?: Tracking student learning (Carroll, 2002)*
2.30 pm		Norhayati Baharun	*A learning design to support student learning of statistics within an online learning environment*
2.50 pm		Nazim Khan	*Engaging First Year Students*
3.10 pm	Afternoon Tea		
Contributed papers: Evaluating learning			Chair: Michael Bulmer
3.40 pm		Brian Jersky	*A case study of knowledge of key statistical concepts before and after an introductory statistics class*
4.00 pm		Mitra Jazayeri	*Evidence based students' conception of variation in sampling – comparison of two methods of teaching*
4.20 pm	*National and international networking and planning. Information and discussion*		
5.00 pm	Close		

Keynote Paper – Chris Wild

The keynote paper by Professor Chris Wild was presented live and was based on a paper read before The Royal Statistical Society on October 20th, 2010, published as: C. J. Wild, M. Pfannkuch, M. Regan and N. J. Horton, *Towards more accessible conceptions of statistical inference* (with discussion), *J. R. Statist. Soc.* A (2011) 174, *Part* 2, *pp.* 247–295 . The full paper with discussion and responses from the authors is available at http://onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.2010.00678.x/full

WHAT I SEE IS NOT QUITE THE WAY IT REALLY IS

WILD, Chris
Department of Statistics, University of Auckland, New Zealand
c.wild@auckland.ac.nz

Abstract
In this talk we gaze through the ripple glass of a bathroom window and wander Alice-like down garden paths through a wonderland where what we see is never quite the way it really is. The paths our odyssey leads us along are conceptual pathways that start with conceptualisations of statistical inference that are intended to be accessible to, and operable by, students mid-way through high school and lead us, via a series of connected trails, all the way to plot annotations that better reveal the stories being told by factor variables in generalised linear models. Along the way, both motivating and suggesting ways forward for all of this, we meet novel visualisations of sampling variation, re-sampling variation and randomisation variation. The talk draws on a paper with Maxine Pfannkuch, Matt Regan and Nicholas Horton entitled, "Towards more accessible conceptions of statistical inference" read to the Royal Statistical Society late in 2010 and on other work on making inference more accessible, particularly via visualisations, with these and other collaborators.

Keynote Paper (refereed) – Delia North

TRANSFORMING STATISTICS EDUCATION IN SOUTH AFRICA

Delia North
University of KwaZulu-Natal
northd@ukzn.ac.za

Abstract
Challenges faced by Statistics Education in South Africa are similar, though often of larger magnitude and of a more critical nature than in more developed countries. This paper will focus on the status of statistics education in South Africa, both at school and tertiary level. The author gives a historical overview, followed by a discussion of the current status, mentioning challenges and successes at the various levels. Special mention will be made of a new school syllabus which has the potential to change the face of statistics education in the country.

INTRODUCTION

There is an ever increasing need to disseminate more data, accurately, in shorter times and in forms desired by users for further analysis (Wallman, 1993). The training of statisticians and the raising of levels of statistical literacy is thus a challenge faced by countries all over the world. However, in South Africa, the challenge is confounded by a legacy of Apartheid, severe shortage of statistics educators, serious lack of adequate professional development of teachers, mixture of cultures and multiple languages in one class room and many more issues that negatively affect statistics capacity building in the country.

This paper gives a historic overview of statistics education (school and tertiary level), followed by a discussion of the current status, outlining challenges and successes at the various levels. The paper further makes special mention of the inclusion of statistics into the new school syllabus and the potential it has for changing the face of statistics education in South Africa

BACKGROUND

South Africa is the 25th largest country in the world, with 47 million citizens of which roughly 79% are Black, 10% are White, 8.5 % are of Mixed Race and 2.5% are Asian (mostly Indian). The country has a very diverse population, with the Black population alone having 9 major ethnic groups. Accordingly, South Africa has 11 official languages, of which English as mother tongue ranks only 5th. Schooling is conducted in all languages at primary school level, but the language of instruction at university is English or Afrikaans (the two official languages during the Apartheid era). English is by far the dominant language of instruction at university level, with all universities switching over to English at senior post graduate level.

During Apartheid in South Africa, education policies were designed to assert white domination and African race inferiority (Badat, 2004), resulting in a school curriculum that systematically ensured that Black school children of that era were prevented from obtaining an education in keeping with the advances of the twentieth century. As a result, the apartheid ideology consciously destroyed a generation of Black mathematics students, thereby depriving them of access to mathematically based disciplines such as statistics.

STATISTICS EDUCATION AT SCHOOL LEVEL: HISTORY AND CURRENT STATUS

When Apartheid was abolished in 1991, education and training in South Africa was restructured to reflect the values and principles of a democratic society, leading to the announcement of a new school curriculum with Outcome Based Education as the fundamental building block. A major difference was that the country now had one school curriculum for all learners, in direct contrast to the racially dividing school curricula that had been in place during Apartheid. This curriculum was further intended to overturn the legacy of Apartheid and catapult South Africa into the

21st century (Chisholm,et al., 2000). The curriculum, known as Curriculum 2005 (DoE, 1997), has been revised and renamed as the National Curriculum Statement (NCS) (DoE, 2003).

Recognition of the cross-curricular need for statistics as an anticipated outcome, led to the collection of data (methods such as interviews and sampling), the application of statistical tools and communication and critical evaluation of finding (North and Zewotir, 2006a) being included in the NCS, under the label "Data Handling". This was in total contrast to what had previously been the case as (1) Statistics had traditionally not been taught at school level in South Africa and (2) mathematical based learning areas, including Statistics, had previously not been in the curriculum of Black school children.

The new outcome-based education system further required that each learner either did mathematics or mathematical literacy in each school year – another major shift from what had been the case prior to the adoption of the new curriculum as it was previously possible to complete schooling without doing any form of mathematics in the last three years of the schooling system. Table 1 below bears evidence of the large increase in the number of school leavers with some form of mathematics under the new schooling system. As can be seen from Table 1, a huge "wave" came through in 2008 (first year when NCS was fully implemented) when a total of 537 271 students registered for the final school exam in mathematics/mathematical literacy as compared to a total of 347 000 in 2007(last school graduates under the previous system).

Year	Type of Mathematics	Registrated
2007	Higher Grade (Main stream)	41 000
	Standard Grade (subsidiary)	306 000
2008	Mathematics (Main stream)	287 487
	Mathematical Literacy (subsidiary)	249 784

Table 1: Number of Registered Students for Mathematics Exam in Grade 12 in 2007 and 2008

In 1998, the Human Sciences Research Council in South Africa conducted a study under the auspices of the International Association for the Evaluation of Educational Achievement. A total of 225 secondary schools were randomly selected from the 9 provinces, resulting in more than 8000 learners, 350 teachers and 190 principals responding to questionnaires and interviews to give an indication of the status of mathematics training in the country at that time. The average score of all participants was 275 points out of 800 points, well below the international average of 487 points. The South African pupils' performance was relatively low in every mathematics topic (from 37% for algebra to 45% for data representation, analysis and probability). The average score for data representation, analysis and probability is the highest in that the score of 356 points (out of 800 points) makes this the scale with the smallest difference from the international average. The results were surprising as Data Handling (Statistics) had not yet been taught at school level at that time!

Challenges:

The Data Handling component of the NCS aims to ensure that each school leaver is statistically literate. The challenges in achieving this aim are to ensure that (1) the school syllabus has the desired content; (2) teachers have the skill and confidence required to promote basic data handling and interpretation skills in the class room.

In-service teachers have generally had no training in Statistics, as this was previously not part of the school syllabus, but the real barrier to meeting the second challenge is the sheer magnitude of the problem. As a legacy of Apartheid there are simply not enough mathematics teachers to meet the demand. This shortage is most pronounced in rural areas, resulting in non-specialist training of mathematics in many schools (Mail and Guardian Online, 2008). All school children now require a level of mathematics (hence Statistics) training, but during Apartheid over 80% of the nation did not receive any mathematics training, leading to the severe shortage of mathematics teachers today.

The Study conducted by the HSRC in 1998 showed that school children in South Africa achieved relatively better scores in Data Handling when it was not part of their school curriculum, than they did in the other learning areas of Mathematics. It is thus a challenge for teachers to build on this natural instinct that the children have for Statistical principles, so that their proficiency and interest in Statistics leaves a lasting legacy when they enter further education and training. The challenge is thus ultimately to have developmental training programs for pre- and in-service teachers in line with this objective.

Successes:

Before the new school syllabus was introduced, school leavers generally had no statistics training at all, so the inclusion of statistics in all grades at school may be regarded as a success.

The Statistics component of the NCS was initially developed by the Department of Education (DoE), taking a formula driven approach, much like the first few chapters of a typical classic university Statistics text book. This prompted intervention from the South Africa Statistical Association (SASA), the mouthpiece of the majority of professional and practicing statisticians in South Africa, and resulted in the successful rewriting of the school curriculum by the DoE (North and Ottaviani, 2002).

The training of teachers to engage with the new NCS was supposed to be done by the subject advisors of the DoE, but they generally had not received any training in Statistics either! In recognition of this dilemma SASA stepped in and is now very actively involved in the teaching of Statistics in schools via its Education Committee, a subcommittee with the specific brief of furthering statistics education at tertiary and preparatory (school) level. This committee has built up a collection of games, projects, newspaper articles, etc. which are used to encourage mathematics teachers to teach statistics in a more meaningful and stimulating way. .

Accordingly, the SASA Education Committee initiated an awareness of the dilemma of statistics education of teachers by giving various talks at local conferences, holding workshops and helping with teacher training. It was however only when Statistics South Africa (Stats SA), the national statistics office, launched the maths4stats campaign, that the human capacity and finances were available to address the dilemma on a national basis. The maths4stats project addresses the dilemma with a roll-out plan to provide statistics training to roughly 10 000 mathematics educators (grades 10 – 12), from 2 800 schools. The objective of the maths4stats campaign is to create a specialized body of educators with a passion for mathematics, and to instil love and interest for mathematics and statistics in educators and learners. Details of the project can be found in North & Scheiber (2008). The long term aim of this project is to strengthen the expertise in Statistics at all levels so that ordinary people have trust in the information they receive from Stats SA (Lehohla, 2002).

The author has headed the statistics outreach programs of Stats SA since inception and has had particular success with training teachers who are not qualified to teach mathematics, but are being forced to do so due to the lack of mathematics teachers. These teachers in particular have a very positive reaction to Statistics as they find it more practical and easier to teach than the advanced mathematical concepts that they often have not mastered themselves.

Modules in Educational Statistics, including research based pedagogical content, are now part and parcel of training of pre-service teachers (Wessels, 2008) to promote the statistical literacy, thinking and reasoning abilities called for in the NCS.

STATISTICS EDUCATION AT UNIVERSITY LEVEL: HISTORY AND CURRENT STATUS

Statistics was first introduced at South African universities in the 1930’s and had a very theoretical focus, deliberately shying away from applied statistics (De Wet, 1998). Statistics historically started at second year level (in a three-year program leading to a Bachelor's degree in Statistics) as one needed a high-level of mathematics to follow the calculus - based approach to teaching statistics that was in place at that time. Historically statistics training was thus geared towards furthering the discipline in a theoretical way. It must however be noted that a few individuals did manage to excel in the applications of statistics through their own self-interest rather than through skills acquired from their tertiary level statistics training.

In recent times, a more balanced view of theory and applications has become apparent in statistics training at South African Universities. Currently 12 universities in South Africa have statistics departments that typically offer a three-year program leading to a Bachelor's degree in Statistics, a one year honours degree, one- or two - year Masters program and four-year PhD programs. Masters and PhD programs are mainly by research dissertation only.

The current structure of the statistics courses in the various South African universities is very similar. Two courses (one per semester) at first year level, two courses at second year level and four courses at third year level are generally the required courses for the BSc program. The two courses (one per semester) in the first year offer an introductory approach to the theory, principles and applications of statistics. The two courses in second year mainly deal with distribution theories, estimation procedures and inference. The third year courses are a mix of methods and applications with a theoretical basis. Relatively intermediate level advanced statistical theories and methods with computer practicals are offered at the honours level. A common feature of honours programs amongst the South African universities is the statistics project, where independent research under guidance of faculty is a key factor in spotting talent for subsequent post graduate studies.

Statistics service courses (non-specialists) generally are very big classes as at least one module in statistics is an essential part of almost all programs at South African universities. Classes can be as large as 500 – 700, making it very difficult to engage in active learning, so pivotal to conceptual understanding. The result is that many students find Statistics courses difficult with a resulting poor pass rate in such courses (North and Zewotir, 2006b). For instance, the average pass rate for Engineering students in a Statistics service course offered at the University of KwaZulu Natal for the period 1997 – 2005 is 73.82% (Zewotir and North, 2007).

In 2006 a survey was conducted by SASA, to determine graduation numbers at the various levels for statistics modules. Table 2 bears evidence of the heavy load that service courses put on statistics departments as well as the small number of graduates in Statistics at the higher post graduate levels.

BSc program			Honours	MSc	PhD	Service Courses
1st year	2nd year	3rd year				
4243	1185	945	190	90	22	16 246

Table 2: Number of passes in statistics modules in the South Africa (2006)

Challenges:

Statistics training at tertiary level in South Africa has challenges embedded in South Africa's Apartheid history. The first challenge is the fact that the majority of incoming students have poor or insufficient mathematical foundation for the theoretically orientated first year statistics courses. This is the direct result of inferior mathematics teaching at school level, generally the result of not having enough adequately trained mathematics school teachers. Instruction at the lower levels is further characterized by large classes where the majority of students typically have not had exposure to computers, have instruction in a language other than their home language and have financial constraints that result in text books not being bought. It is thus not uncommon to have high failure rates in statistics modules, particularly at level 1 (Steffens 1998).

The second challenge is a shortage of post graduate students and statistics lecturers. Statistics education at the higher levels is characterized by very small numbers, which is a great concern as BSc and Honours graduates in statistics get lucrative offers from industry, business and government, luring them away from academia. Accordingly, very few South African statistics graduates opt to pursue postgraduate studies. South African universities on average have at least 15% of their Statistics posts vacant, whilst 20% of posts at Statistics South Africa, the national statistics office, are vacant. The International Review Panel Report on the Review of Mathematical Sciences Research at South African Higher Education Institutions (18.12.2008) concluded that "the shortage of academic statisticians is so critical that the field is in danger of disappearing through lack of academic capacity", further noting that "the closure of academic departments is a real possibility". The shortage of Statistics lecturers at universities and increasing numbers of first year students results in ever increasing teaching loads, which has a negative effect on research output, clearly a concern.

Successes:

A number of universities have introduced access programs (extended length), supplementary instruction, mentorship programs, and hot seats (private lessons by post graduate students) in first year Statistics modules to deal with the problems discussed earlier. The use of local textbooks, or more effectively, the writing of course packs to replace foreign textbooks do, to a certain extent, overcome the problem of presenting examples that the students can relate to.

Most Statistics departments are characterized by low research output compared to other sciences. This is clearly an area of grave concern for SASA and resulted in the SASA education committee organizing a workshop where four senior academic statisticians were invited to jointly present a workshop on growing a research profile in Statistics and improving supervision of postgraduate students in Statistics, with a subsequent panel discussion. The panel discussion highlighted deficiencies in the undergraduate and postgraduate programs, lack of collaborations between the senior and young statisticians, and lack of motivational incentives for further studies as the main issues. SASA and Stats SA are jointly negotiating ways to address the problem on an ongoing basis.

CONCLUSION

The building of statistics capacity to meet the ever increasing demand from business, industry and government has been a hot topic at conferences all over the world throughout the past decade.
South Africa has a critical shortage of statistics skills, resulting in an urgent need to produce more statistics graduates. The abolishment of Apartheid and recent introduction of Statistics into the school syllabus in South Africa has given the country the opportunity of exposing all school children to basic statistical principles, in direct contrast to what had been the case historically. The extent, to which this exposure will result in school leavers with an increased appreciation for, and interest in Statistics, will only be known in years to come. The recent joint efforts between Stats SA and SASA to work collaboratively with the DoE to promote statistics education at school level are commendable as the probability that the introduction of statistics at school level will filter through to tertiary level and increase the number of statistics graduates in the country is clearly positively influenced by the level of success the country has in training mathematics school teachers to be confident and empowered to teach statistics meaningfully in the class room.

REFERENCES

Badat, S (2004).Transforming South African Higher Education, 1990–2003: Goals, Policy Initiatives and Critical Challenges and Issues. In N. Cloete, P. Pillay, S. Badat & T. Moja (Ed.), *National Policy and a Regional Response in South African Higher Education*. Published in association with Partnership for Higher Education in Africa, James Currey (Oxford) and David Philip (Cape Town).

Chisholm, L., Volmink, J., Ndhlovu, T., Potenza, E., Mahomed, H., Muller, J., Lubisi, C., Vinjevold, P., Ngozi, L., Malan, B., & Mphahlele, L. (2000, May 31). A South African curriculum for the twenty-first century. (Report of the review committee on Curriculum 2005). Pretoria: South Africa.

Department of Education [DoE] (1997). Senior phase policy document. Pretoria: Department of Education.

Department of Education [DoE]. (2003). NCS Mathematics Grades 10 - 12, Department of Education, Pretoria, South Africa,
http://www.education.gov.za/Curriculum/SUBSTATEMENTS/Mathematics.pdf

De Wet, J.I. (1998). Teaching of Statistics to Historically Disadvantaged students: The South African Experience. *Proceedings of the Fifth International Conference on Teaching Statistics.* Voorburg, Netherlands: International Statistical Institute. Available from
http://www.stat.auckland.ac.nz/~iase/publications/2/Topic5e.pdf

Howie, S. & Plomp, T. (2002). Mathematical Literacy of school leaving pupils in South Africa. *International Journal of Educational Development.* **22**(6), 603-615.

Lehohla, P. (2002). Promoting Statistical literacy: a South African perspective. In B. Phillips (Ed.). *Proceedings of the sixth International Conference on Teaching Statistics*, Cape Town, South Africa. International Statistical Institute and International Association for statistics Education. Online: www.stat.auckland.ac.nz/~iase/publications.

Mail and Gaurdian Online. (2008). Teachers Flunk Maths. Available at http://www.mg.co.za/article/2008-08-03-teachers-flunk-maths

North, D & Ottaviani, G. (2002). Statistics at Foundation School Level in South Africa – The Way Forward. *Proceedings of the Sixth International Conference on Teaching Statistics.* Voorburg, Netherlands: International Statistical Institute. Available from http://www.stat.auckland.ac.nz/~iase/ publications/1/2d2_nort.pdf.

North, D & Scheiber J (2008). Introducing Statistics at School level in South Africa. The crucial role played by the national statistics office in training in-service teachers. *Proceedings of ICMI Study and IASE Roundtable, Mexico*. Available from http://www.ugr.es/~icmi/iase_study/ .

North, D & Zewotir T (2006a). Introducing Statistics at school level in South Africa. In D. Ben-Zvi & J. Garfield (Ed.). *Proceedings of the Seventh International Conference on Teaching Statistics*, Cape Town, South Africa. International Statistical Institute and International Association for statistics Education. Online: www.stat.auckland.ac.nz/~iase/publications.

North, D & Zewotir T (2006b). Teaching Statistics to Social Science Students: Making It Valuable. *South African Journal of Higher Education* **20**(4), 503-514.

Wallman, K. K. (1993). Enhancing Statistical Literacy: Enriching our Society. *Journal of the American Statistical Association*, **88** (421), 1 – 8.

Wessels, H. (2008). Statistics in the South African School Content: assessment and teacher training. In C. Batanero, G. Burrill, C. Reading & A. Rossman (eds), Joint IcMI/IASE study: Teaching Statistics in School Mathematics Challenges for teaching and teacher Education. *Proceeding of the ICMI study 18 and 2008 IASE round Table conference*. Online: http://www.ugr.es/~icmi/iase_study/Files/Topic1/T1P3_Wessels.pdf

Zewotir T. and North, D. (2007). Focus on the Statistical Education of Prospective Engineers in South Africa. *Pythagoras* **65**: 18-23.

Contributed Papers

Topic: Graduates across disciplines

Contributed paper – Sue Finch and Ian Gordon

STATISTICAL CONSULTING WITH POST-GRADUATE STUDENTS

FINCH, Sue and GORDON, Ian
University of Melbourne, Australia
Sfinch@unimelb.edu.au

Abstract
The Statistical Consulting Centre at The University of Melbourne provides a post-graduate consulting service to higher degree research students across all disciplines from medicine to traditionally non-quantitative fields like development studies. Students can obtain advice about any stage of the research cycle from refining their research question and developing a suitable design to presenting and communicating findings. Some students have very little knowledge or experience in statistics, except in collecting their research data. Others are competent but need particular advice about specific problems. We took the opportunity to survey consulting sessions with post-graduate student, in order to reflect on the educational needs of these clients. We report on our survey and present some case studies to characterise aspects of graduate education that are applied and contextually relevant.

Contributed paper – Glenys Bishop and Emlyn Williams

A MODEL FOR BUILDING STATISTICAL CAPACITY AMONG SCIENCE RESEARCH STUDENTS

BISHOP, Glenys and WILLIAMS, Emlyn
Statistical Consulting Unit, The Australian National University, Australia
glenys.bishop@anu.edu.au

Abstract
The Statistical Consulting Unit at the Australian National University provides support to Honours and graduate research students, in addition to research staff. Students may consult one of the statisticians in the unit about design of the data collection phase such as design of an experiment or survey. They may also seek assistance with analysis, to the point where they are sufficiently adept to apply the recommended techniques themselves. In addition to the benefits of a better thesis outcome, students have also learnt the value of collaboration and been exposed to a range of design and analysis possibilities that may not have been available to them before consulting the unit.

However, to make the consulting process more efficient, it is important that students have enough statistical knowledge to be able to communicate effectively with statistical consultants. The Statistical Consulting Unit, at the request of the Colleges of Science, established a project to investigate the most practical and effective ways to transfer knowledge about statistical practice and statistical thinking to ANU Science students. After extensive consultation with stakeholders, a number of options, including their advantages and disadvantages, were proposed. An online learning model was adopted as the main method to build statistical capacity among Science research students. This can be supplemented by short courses addressing specific needs of some groups within the university.

This paper will report on a number of issues which had to be considered in arriving at this model, including the availability of existing online courses, the experiences of other institutions, whether to make an online course compulsory or voluntary, how to encourage students to take a voluntary online course, how many and which topics to include, ease of implementation and constraints of the university environment. These will be illustrated where possible using examples from the implementation.

In H. MacGillivray and B. Phillips (Eds.). Proceedings of the Seventh Australian Conference on Teaching Statistics, December, 2010, Fremantle, Australia.

Contributed paper – Kristen Gibbons and Helen MacGillivray

TRAINING FOR STATISTICAL COMMUNICATION IN THE WORKPLACE

GIBBONS, Kristen[1] and MacGILLIVRAY, Helen[2]
[1] Clinical Research Support Unit, Mater Medical Research Institute
[2] School of Mathematical Sciences, Queensland University of Technology
kgibbons@mmri.mater.org.au

Abstract
Transitioning from university student to a statistician working as a collaborative researcher, consultant and educator of doctors, nurses, scientists and allied health staff in a medical research environment is a very daunting and challenging task, particularly when explaining about the use and misuse of statistics. While a sound undergraduate statistics and mathematics background provides the necessary foundation for ongoing learning and to competently understand and perform statistical analyses, it does not necessarily enable effective communication of the concepts and results of these analyses to busy professionals with limited or no prior statistical experience. This paper describes how experience gained in a developmental and mentored program in tutoring statistics has proven to be invaluable in working as a statistician in a large tertiary hospital that also incorporates a basic science medical research institute.

INTRODUCTION

Preparing for the workplace after completing an undergraduate degree in any field is a daunting prospect. This is made easier in many disciplines knowing that you will be communicating with like-minded colleagues, who have had similar training and on a day-to-day basis are performing comparable tasks. However, the possible employment opportunities after completing a mathematics degree are endless, and as a result, very varied. This is particularly true if the mathematics degree includes major studies in statistics, and even more so if it is partnered with an information technology degree. Although all disciplines need communication skills in their graduate capabilities, they are especially important for mathematics and statistics graduates, because of both the nature of these disciplines and the very diverse workplaces and careers open to such graduates. Graduate statisticians are employed in a diverse range of positions, with a varied statistical support base, and as such require the skills to not only be able to communicate with other statisticians, but more importantly, with professionals in their work environment who are not familiar with statistical concepts and methodology. Statisticians also need to understand, model and tackle problems in unfamiliar contexts, and to plan, implement, analyse and interpret investigations in collaboration with those who are familiar with the contexts of interest. Cameron (2009) notes how many statisticians have commented over the years on the importance of such skills in training statisticians. Communication capabilities are integral to all these skills, are complementary to a sound knowledge of statistical methodology, and should be a vital part of any statistics degree.

Statistical consulting is demanding of such skills, and working as a collaborative researcher, consultant and educator of doctors, nurses, scientists and allied health staff in a medical research environment is especially demanding. The training of statisticians for such roles must incorporate the skills required for not only performing statistical analyses, but also for consulting with researchers with limited statistical knowledge, effectively teaching researchers both the concepts and the use of statistical software, as well as being able to ensure that the correct statistical advice is not lost in the power struggle of competitive research. In a standard undergraduate program, there is often little opportunity for students to learn the skills necessary to be able to communicate with non-statisticians in the 'real world'. While opportunities such as working in groups to complete assessment tasks and making presentations to fellow students are encouraged, this does little to foster the skills required to communicate concepts to others with little or no statistical experience. Without these skills, graduates in an environment with the requirement to consult with non-statistician's could be left with a lack of skills to competently both apply the appropriate statistical methodology, and communicate the methods and results to their colleagues.

There has been increasing emphasis in statistics education on inclusion of experiential learning of the whole process of the statistical data investigation cycle (Wild and Pfannkuch,

1999, MacGillivray, 1998, Forster and MacGillivray, 2010) and on communicating statistics (see, for example, Lipson and Kokonis, 2005, Forster et al, 2005). As an undergraduate, the first author was fortunate to have these learning experiences, but a further facet that has received little attention in statistics education has proved to be significantly advantageous. The experience gained by the first author during her undergraduate degree in a developmental and mentored program in tutoring statistics has proven to be invaluable in subsequent employment in a consultative medical research environment, dominated by professionals with little or no statistical experience. Through tutoring different aspects of statistics, including the experiential learning by other students of data investigations and problem-solving, as well as different disciplines of students, the principles of communicating statistics were learnt, and developed over a number of years. Without these tools the role of the statistician in a workplace where the majority of staff have a minimal statistics background would be near impossible.

After describing the program, the tutoring experiences, and reviewing some of the skills learnt, the requirements of the statistician in the medical environment are outlined and some case studies illustrating the training provided by the developmental and mentored tutoring program described. The paper concludes with a recommendation for similar programs in training statisticians.

THE TUTOR DEVELOPMENT PROGRAM

The first author completed a dual degree in Bachelor of Mathematics (Honours) and Bachelor of Information Technology from the Queensland University of Technology (QUT) in Brisbane, Australia, graduating in 2006. For the mathematics component of the study, approximately one third of units were in the statistical stream, and the focus of the honours project was also statistical, applying statistical inference methodology with computational statistics.

The first part of the developmental tutoring program was working as a mentored volunteer duty tutor in the drop-in facility of the QUT Maths Access Centre (QUTMAC), of which the second author is Director. The drop-in facility is a collaborative student workroom with resources and duty staff (volunteer or QUTMAC staff) on a roster that covers approximately 15 hours per week. Undergraduates with an appropriate balance of academic background and level of achievement are encouraged to apply for these positions. For example, a student who has completed the first year mathematics and statistics courses requires a high grade point average, but this may not be as necessary for a student with third year mathematics and statistics. Other criteria relevant to tutoring are used, but high achieving and committed students to whom such volunteer work appeals are generally given every encouragement. Volunteer drop-in duty tutors generally work for an hour a week for at least one semester, interact with fellow duty tutors and QUTMAC staff, consult the QUTMAC director for advice, and can refer any problems or difficulties to the QUTMAC director.

Mentored volunteer drop-in duty tutoring provides an excellent introduction to the development of communication and tutoring skills, with mostly one-to-one assistance in an informal friendly environment with no time pressures or formal obligations, and with any concerns or difficulties referred to the QUTMAC director. The first author had the opportunity to do drop-in volunteer duty tutoring in her second year. Since then, the concept has been consolidated and developed, with more integrated and systematic preparation, mentoring and feedback. Feedback has been increasingly and consistently highly favourable, with almost all QUTMAC volunteer duty tutors praising the scheme as invaluable in providing them with a base for moving into tutoring. The volunteers most suited for tutoring greatly enjoy the experience, and some love it so much that they continue as volunteer duty tutors after they are appointed as sessional paid tutors in formal tutorials. A few discover they do not enjoy it and are able to withdraw. After completing at least one semester of QUTMAC duty tutoring, the volunteers are invited to participate in a two-day tutor training program run by the QUTMAC director. They are then included on the list of potential sessional tutors in mathematics and/or statistics.

After completing a semester of volunteer duty tutoring, the first author participated in two days of tutor training, which had the following general program:

- Day 1:
  - The nature of the tutorial
  - Duties and responsibilities
  - Working with the lecturer/discipline co-ordinator
  - Planning and preparation
  - Connecting with students
  - Understanding students' school and other backgrounds
  - Overview of ways people learn mathematics and statistics
  - Consideration of examples to select for trainees' demonstrations
  - Group discussion about reflection and articulation of good tutorial practice
- Day 2:
  - Presentations of examples and subsequent feedback from peers and staff, with discussion
  - Administration matters
  - Training in marking using marking schemes and exemplars

For the majority of students completing the QUTMAC duty tutoring and tutor training, this was the first exposure to being on the other side of the teaching environment; teaching, not being taught. Although all students had achieved highly in their own studies and were very confident in the material that would be taught, it was important, and this was recognised by all, to attempt to learn additional skills so that the material could be explained in ways that would assist students to comprehend the statistics being taught. While these days proved to be daunting at times (no matter how well you know how to answer or work through a question, it is still nerve-wracking doing it in front of your peers for the first time!), at the close of the training program were all armed with a lot more understanding about the challenges of helping others to understand.

After completing the above program, the first author was appointed as a sessional tutor in her third year of study, and continued as a sessional tutor in the third and fourth year of undergraduate studies, during the honours year, and even a small amount during the early part of her full-time employment after graduation. Tutoring was in three different statistics courses to cohorts ranging over all of science, engineering, mathematics, together with some education, surveying, health and other disciplines. All tutoring involved some preparation of material to be presented in the tutorial, some presentations, working with students in small groups and one-on-one during the tutorial session, and marking subsequent assessment.

THE TUTORING EXPERIENCE AND SKILLS LEARNT

A wide range of skills were learnt over the course of the QUTMAC volunteer duty tutoring, the two-day training program, and the extended experience of tutoring, however some have proved to be more relevant to life as a consultant statistician than others. As mentioned already, communication is a key factor, but not only is this far more than merely being knowledgeable, articulate and able to present and explain clearly, it is also an integral part of the full spectrum of skills gained from tutoring, provided, of course, that the tutor is fully engaged in, and dedicated to, facilitating learning. A good tutor learns to listen and observe, and thus, through reflection and discussion with the lecturer and other tutors, learns skills in communicating and working with a wide range of

- different groups of people;
- skills and knowledge background;
- motivations and personalities; and
- learning and working styles.

In addition, explaining concepts, techniques and their applications to other people provides great depth to the tutor's own understanding of statistics and statistical thinking. As one experienced and expert professor in statistics commented during interviews (MacGillivray, 2008), "*I learnt understanding of statistics through teaching service courses*".

However, it must be emphasized that the value in the tutoring experience was because of the nature of the tutorials, the pedagogies of the courses, and the collaborative mentoring of the lecturer. The focus of the tutorials was on working with the students, facilitating their learning, and not on merely standing in front of a class providing instruction. The pedagogical approach of the courses was similar, providing experiential learning of the full data investigation cycle and problem-solving in probability and distributional modelling, in an environment of collaborative and supportive learning. Tutoring that involves no more than instruction by the tutor in courses with out-of-date instructional dictatorial pedagogy provides no more opportunities to learn than reading a book by oneself.

There were numerous practical experiences in both the QUTMAC duty tutoring and sessional tutoring experiences that have proven to be invaluable in working life. One of the key components of the QUTMAC duty tutor is to work one-on-one with a student to help them discover how to progress through specific problems. In particular, working through the cycle of understanding the problem, explaining the necessary theory and assisting the student with applying the theory has proven to relate very closely to the consultative work in graduate life, and this was an invaluable skill to have been exposed to. An aspect of the sessional tutoring that has been particularly useful is the interaction with groups of students throughout the semester on their free-choice projects exploring the data investigation cycle. This would include helping with forming an interesting question, developing methods to collect the data, determining the appropriate data analysis, and assisting with analysis and interpretation of data after collection. Assisting with these projects built on the first author's experiences as students carrying out such investigations. Again, this is very closely related to the work as a consulting statistician, and the experiential nature of this experience not only assisted us as students, but also then as tutors in providing wonderful experience for future graduate roles.

PROFESSIONAL SETTING

Since graduating, the first author has been employed by Mater Health Services, in the Mater Research Support Centre, which has since been integrated with the Mater Medical Research Institute (a biomedical research facility on campus) to form the Clinical Research Support Unit. Mater Health Services is a tertiary facility comprising seven hospitals (both public and private, across two campuses), a research institute and additional clinical services such as Refugee Health Queensland, and has well over 7000 staff.

The Clinical Research Support Unit aims to support high-quality research activities on campus across all disciplines and hospitals and for any staff member. In addition to research projects, quality assurance activities are also supported. Services offered include protocol design and development, statistical advice and analysis, assistance with data management and clinical research co-ordination. The Unit assists with over 160 projects a year, ranging from simple audits through to complex clinical trials.

Initially, the role undertaken by the first author was purely that of a statistician, but it has since grown to encompass the supervision of research-related data management activities. Specifically, the following duties are undertaken:

- Randomisation services
- Statistical consultations with researchers in the planning phase, including sample size calculations
- Statistical analyses
- Co-ordination of development and maintenance of research databases
- Review reporting of statistical results
- Data and Safety Monitoring Committee statistical support
- Statistics education activities

Both the clientele and the projects supported by the author are extremely varied. While the majority of projects are carried out by medical staff who undergo some element of statistical education in their training, the range of statistical experience is enormous. A substantial portion of the support services provided are to medical trainees, many who are required to complete a research project to fulfil their discipline-specific training. The majority of these doctors are willing to learn basic research techniques such as literature searching and study design, however there is generally a lack of desire to both understand and carry out the statistical aspects of their projects. At the other end of the spectrum are senior clinicians, many of whom have exemplary research track records and are both willing and eager to further educate themselves on the statistical techniques available and how they can be applied to their research. Assistance is also provided to allied health, nursing and midwifery staff, many of whom have no formal statistical education, and most have no research training either. Given this diverse clientele base, the ability to communicate statistical concepts to a wide range of experience levels is vital.

In addition to the primary role of consultation, there is also a very important focus on education of staff within the Unit. An annual two-day introductory statistics course is run for hospital staff, aimed at those researchers who have little to no background in statistics, and emphasising the interpretation of statistics, rather than the underlying theory. In addition, smaller focused sessions are held regularly for individual groups of staff, for example the obstetrics and gynaecology registrars. As a large teaching hospital, it is imperative to ensure that trainees have the required skills in both clinical areas, as well as research.

SOME CASE STUDIES ILLUSTRATING SKILLS UTILISED

Over the four years of employment as a consultant statistician, there have been many times when the skills described above have been utilised. Three such examples will be described, highlighting the use of communication, knowledge of the data investigation cycle and consultative experiences. The three examples consider the educational, consultative and research aspects of the role.

As noted above, education is a key component of the services provided by the Clinical Research Support Unit. There are a number of opportunities for the first author to educate staff about various aspects of statistics, ranging from spontaneous unstructured sessions (generally consultations), short seminars (either at a trainees session, campus wide grand rounds or invited presentations at research meetings) to the structured two day education course. The wide range of experiences from the tutoring program have helped in all aspects of education.

The two day introductory statistics course is run each year and is aimed at novice researchers on campus, covering topics such as basic probability theory, differences between two groups, correlation, regression, graphical representation of data and misuse of statistics. The first author is the primary organiser and presenter of this course, and the course is limited to 30 participants. Given the target audience, the course is designed to provide participants with the basic tools necessary to be able to read a scientific article and critically appraise the methods used. As such, the course involves many examples of how to calculate basic statistics, as well as learning to interpret statistical output from computer programs as well as understanding statistical methods and results in journal articles. The style of teaching is not dissimilar to that experienced as an undergraduate tutor, particularly of service courses, as the participants have little to no background in statistics, and are often only participating to obtain relevant professional development accreditation. At the end of each course, an evaluation survey is conducted to ascertain whether the participants found the course useful and whether the style of presentation should be changed. It has been noted on evaluations that the approachability of the presenter, use of practical examples and information on other aspects of research projects (for example, data collection and management) were all strengths of the course, demonstrating the practical application of skills learnt from the tutoring experience.

The primary function performed by the first author is to consult with researchers from across the hospital campus to assist with their quality assurance and/or research projects. The level of consultation can range from a brief chat to reassuring a researcher that the statistical methods that they have employed are appropriate, to formulating a research plan and educating the researcher on what statistical methods are available and should be used for their project. The confidence that the first author gained through tutoring has translated to the ability to spend time educating the researchers and encouraging them to attempt to write their own statistical methods and perform their own analyses, rather than write the relevant sections of a statistical plan and conduct the analyses for them. While this method does not appeal to all researchers (particularly trainees who have no interest in their research project apart from completing it to satisfy training requirements), numerous comments have been made that the research clients appreciate being taught the statistics, and being able to converse with a statistician, compared with previous negative experiences of consultant statisticians who had few communication skills and would not attempt to educate the researchers.

For the majority of research projects the aim is to be involved from beginning to end; to ensure that an appropriate sample size is calculated, the data are collected in a fashion suitable for data analysis, analyses performed properly and results interpreted correctly. The experience of learning and then tutoring the data investigation cycle have played an integral role for the statistical consultant in assisting researchers to design their research projects, and be able to offer advice not only for statistics, but potential pitfalls in data collection as well as the appropriate steps to take when developing the research protocol. The usefulness of the input from the first author is evident through the increasing number of referrals to the service through existing researchers, as well as clients who have previously only used the services of the statistician at the end of their project, now asking for statistician input at the beginning of the project.

CONCLUSION

Sound knowledge of statistical concepts and methodology and high achievement in formal assessments are no longer enough for a graduate statistician to succeed in a consultative environment interacting with professionals with little or no statistical knowledge. Preparation of graduate statisticians for life after university should include the opportunity during their degree to foster the skills required to communicate not only statistical concepts, but also to work with professionals to develop and implement projects. These skills can be developed or enhanced by universities by implementing an undergraduate mentoring program, including a training program, where students are provided with the opportunity to assist, mentor and tutor other students.

It is important to note that simply tutoring a class or participating in a training program is not enough; the courses themselves must genuinely provide experiential learning of statistics that reflects how statisticians work in the workplace and how they tackle problems. The added skills of listening, observing and helping a wide range of personalities, capabilities, backgrounds and motivations enhances the communication skills learnt. Additionally, the tutoring experience must involve interaction with students (as opposed to merely demonstrating from the front of the class), regular meetings and involvement with the course co-ordinator (to foster the mentoring relationship) and tutoring courses that have a large component of experiential learning to ensure the most is gained out of this fantastic opportunity.

The integration of these aspects into an undergraduate statistics degree will result in a new generation of statisticians, capable of smoothly transitioning into working life, with the skills necessary to consult, collaborate with, and educate a wide range of professionals with little or no statistical knowledge.

REFERENCES

Cameron, M. (2009). Training statisticians for a research organisation. *Proceedings of the International Statistical Institute 57th Session.* Durban, South Africa. ISI.

Forster, M., Smith, D., & Wild, C. (2005). Teaching students to write about statistics. In L. Weldon & B. Phillips (Eds.), *Proceedings of the ISI / IASE satellite on statistics education and the communication of statistics*. Voorburg, The Netherlands: ISI.

Forster, M. and MacGillivray, H.L. (2010). *Student discovery projects in data analysis*. In Reading, C. (ed) The Proceedings IASE/ISI 8th International Conference on Teaching Statistics, Ljubljana: ISI, Voorburg, The Netherlands.

Lipson, K., & Kokonis, S. (2005). The implications of introducing report writing into an introductory statistics subject In L. Weldon & B. Phillips (Eds.), *Proceedings of the ISI/IASE satellite on statistics education and the communication of statistics*. Voorburg, The Netherlands: ISI.

MacGillivray, H.L.(1998). Developing and synthesizing statistical skills for real situations through student projects, *Proceedings 5th International Conference on Teaching Statistics*, 1149-1155, ISI: Voorburg, Netherlands.

MacGillivray, H.L. (2008). *Learning and developing higher education teaching in statistics*, CETL-MSOR Conference, Lancaster, UK, September 2008.

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

Contributed paper – Sharleen Forbes

DATA VISUALISATION: A NEW STATISTICAL LITERACY TOOL FOR STATISTICAL OFFICES

FORBES, Sharleen

Victoria University of Wellington
P.O.Box 600
Wellington
New Zealand

Statistics New Zealand
P.O.Box 2922
Wellington
New Zealand

sharleen.forbes@vuw.ac.nz

Abstract
The ability to harness massively increased computing power together with the availability of new free graphical tools easy to download from the Internet has led to a burst of creativity in new dynamic and interactive data visualisations. Some of this activity has taken place in national statistics offices. Official statistics have traditionally been released in simple and standard tables and graphs but these statistics provide the evidence base for much of government policy and new static and dynamic graphs and maps that allow users to interrogate and interact with data in new ways have been developed to increase the usefulness and understanding of these statistics. In this presentation a Consumers Price Index kaleidoscope is used to investigate the structure of index numbers, multidimensional scatterplots to teach conceptual understanding of multiple regression, and dynamic population pyramids, commuter flows and integrated graphs and maps to show the multi-disciplinary nature of statistics in the real world.

INTRODUCTION:

Tucker (2010, p459) stated that '*government agencies are interested in statisticians who know how to approach real-world problems analytically and use their statistical training to solve them. Writing and communication skills also are emphasized'*. He appears to be making a case for government statisticians to be statistically literate in the sense defined by Gal (2002). That is, as having *'the ability to interpret, critically evaluate, and communicate about statistical information and messages'*. There is already a large body of research of methods for developing these skills some of which, such as having activity-based courses, small groups and building on the ideas and material that students bring to the classroom, are summarized in Garfield (1995). Wild & Pfannkuch (1999) suggest that *'The cornerstone of teaching in any area is the development of a theoretical structure with which to make sense of experience, to learn from it and transfer insights to others*' (p224). Visual tools provide one way to students to gain experience of statistical concepts. Paparistodemou & Meletiou-Mavrotheris (2008) used the Tinkerplots (Konold & Miller, 2005) dynamic statistics software to enhance the learning of statistical inference by young school students. Wickham (2010) focuses '*more on visual exploration and less on quantitative modelling*' when teaching statistics to heterogeneous university students and Dominguez-Dominguez & Dominguez-Lopez (2010) contend that the visual approach is *'learning by playing and do-it-yourself'*.

Examples that I have used in my own teaching are portraying the mean as the 'balancing point' or centre of gravity of a set of data by placing small (lego) blocks and the standard deviation (as a measure of the thickness of 'spread') visually showing changes as points are altered from being clustered about a mean to having a wider spread. There is a wide body of literature on the development and use of visual displays for communication and analysing statistics (e.g. Tufte, 1983; ten Bosch & de Jonge, 2008; Hidaglo, 2010) and there has been somewhat of an explosion in the availability of data visualization software that provide new methods for access to and interpretation of data (Forbes et al, draft). Many of these tools can be used to develop conceptual understanding when teaching statistics and several examples of the potential use of official statistics visualizations are given below.

USING DATA VISUALIZATIONS TO TEACH STATISTICAL CONCEPTS:

Stepping from simple to multiple regression:

Both conceptually and educationally there is a large step from simple bivariate regression to multivariate models that may or may not incorporate interactive terms. It is often difficult for students to understand the impact of interaction (of one variable acting as a modifier on another). A visual aid to assist with this is a three-dimensional (3-D) display (in this case a pin-graph where the heads of the pins form a 3-D scatterplot). Figures 1a and 1b were created using the open-source software package 'R' (www.r-project/org). These graphs enable students to see and discuss the degree of interaction present and decide whether to just include main effects or to add interaction terms in a multiple regression model by looking at the consistency of pattern (for example, across highest education classes in figure 1a). An extension from three- to four-dimensions is achieved by the use of colour in figure 1b.which could be used to discuss whether separate models should be fitted when the pattern of work is different for different groups (females dominating part-time and males full-time work). Other ways of increasing dimensionality are by changing the size, or colour intensity, of data points or by adding a dynamic feature across overlaid static graphs as used by Hans Rosling (2007) in the creation of his Gapminder graphs.

Figure 1a: Weekly Income by Highest Educational Qualification and Hours Worked

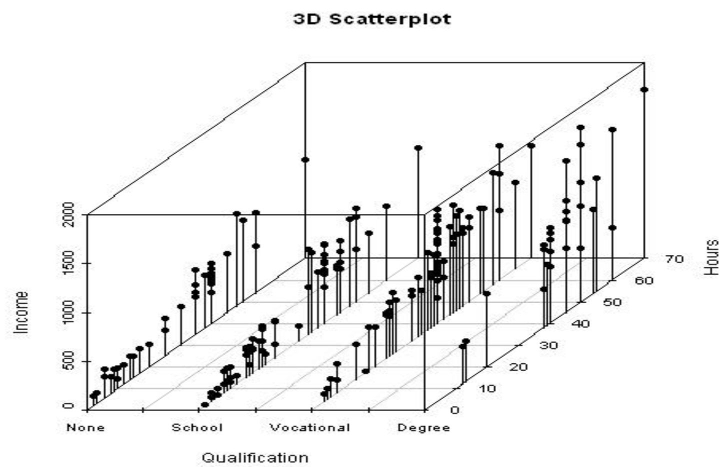
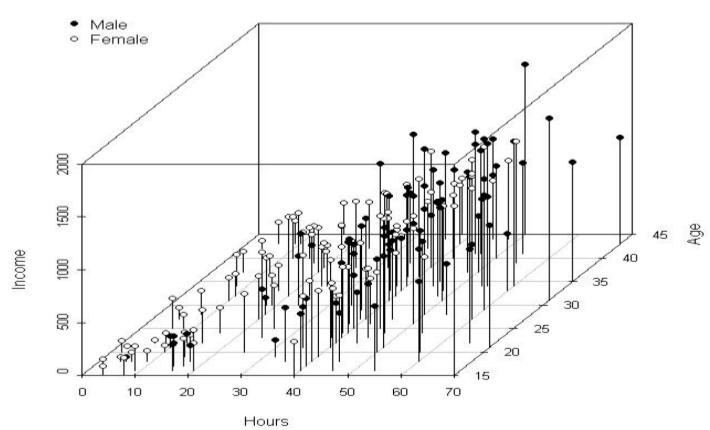


Figure 1b: Weekly Income by Sex, age and Hours Worked



Understanding Price Indices:

Consumer's Price Indices (CPIs) measure the rate of price change of representative goods and services (the basket of goods) purchased by households. In New Zealand there are about 690 goods and services in 107 classes categorized up into 45 subgroups then 11 groups (Statistics New Zealand, 2008). Each good or service is assigned an (expenditure) weight representing its relative importance in household spending patterns (as determined by the Household Expenditure Survey). The CPI is often used by the media to report the changes in prices that households face, but it is also used in New Zealand by the Reserve Bank to set monetary policy, by the government to adjust benefit rates and by employers and employees in wage negotiations. While it is relatively easy to explain the calculation of a single item price index as

$$\frac{current\ price}{reference\ price} \times index\ reference$$

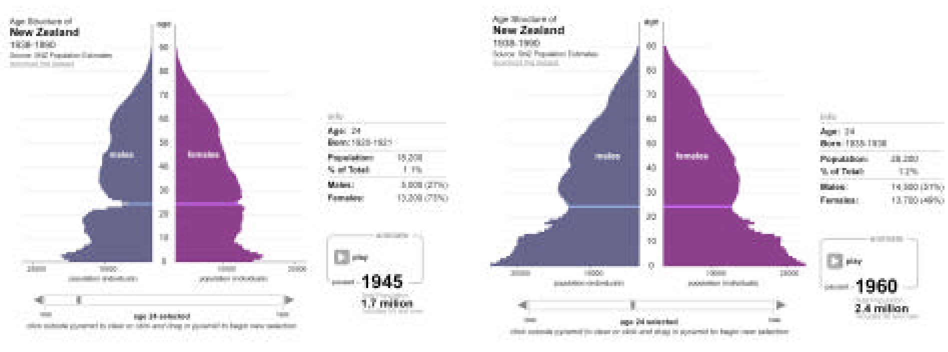
it is more difficult for the whole CPI, even when the relatively simple Laspeyres formula[1] (Statistics New Zealand, 1999) is used. Both subgroups and weights can be displayed visually with a tool such as the Price Kaleidoscope produced by the Federal Statistical Office of Germany (http://www.destatis.de/Voronoi/PriceKaleidoscope.svg). In this tool the CPI is represented by a circle and within this each group (and subgroup) has an area proportional to its weight in the CPI. Clicking on a area displays both the weight and quarterly change in that subgroup allowing students to explore how groups are formed and what the impact of the weights is. They quickly see that while there may be an overall small percentage increase in the overall CPI some subgroups may have large positive or even negative changes and that the effect of this on the total CPI is determined by its weight (area).

Demography simplified:

One of the ways of viewing the changes in large data sets over time is to overlap graphs (or maps) taken at different time points and then use an animation tool to 'play' these over time. An example is the dynamic population pyramids now in relatively common usage by national statistics agencies. As shown in figures 2a – 2d these dearly indicate changes in a population structure as it 'ages'. By playing these graphs demographics effects such as momentum (either that of population growth resulting from a youthful age structure or that that of population decline resulting from an older age structure) can be viewed. As Jackson (2001) says these '*two trends are often on a seemingly unavoidable collision course*'. Dynamic population pyramids provide a graphic of this process in action, leading up to and beyond the point (different for each country) where natural increase (growth) shifts to become natural decrease (decline). They can also be used to demonstrate the difference between historical analysis of population change and the 'what if' analyses of population projections.

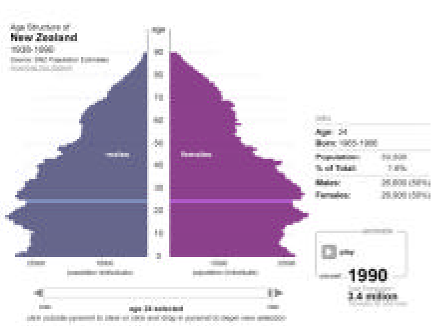
Figure 2: Snapshots of the Statistics New Zealand population pyramids

(a) At year 1945 **(b) Fifteen years later - 1960**

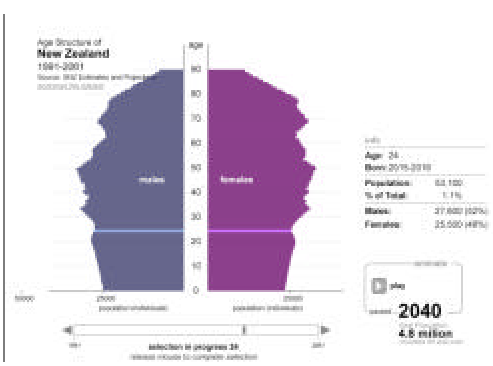


[1] The Laspeyres formula calculates the index for period t on base period o by: $Index = \frac{\sum P_t Q_o}{\sum P_o Q_o} \times 1000$ where Q is the quantity (weight) and P the price of the item.

(b) Another thirty years on – 1990



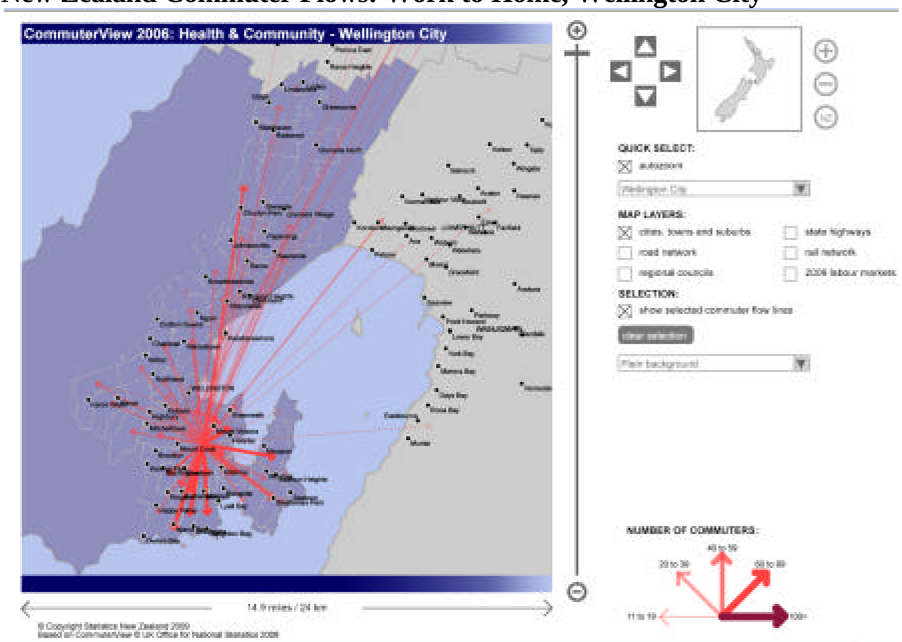
(d) Another fifty years on – 2040?



Geography and statistics:

Many statistics, including official statistics, have an associated geography and new ways of demonstrating this are being explored. Geo-visualisation is '*conceptualised as a means of visually representing spatial data to better explore and understand patterns and relationships in the underlying information*' (de Róiste et al , 2009). Public domain geo-visualisation tools (such as Google Maps and Google Earth) are only the beginning. An example of the use of a mapping tool to reduce, make manageable and present very large and complex data sets of data reduction is the interactive mapping tool "Commuterview" (based on a product from the UK Office for National Statistics) which was released on DVD and distributed on request by Statistics New Zealand. This tool displays linked information from two 2006 New Zealand Census of Population and Dwellings questions 'Where do you live?' and 'Where do you work?'. The 'spider' graphs in figure 2 show the home locations of staff employed by the central public hospital in Wellington city and indicate that a number live over hills or at some distance from the hospital which could be an issue in disaster mitigation planning as the city lies along a major fault line.

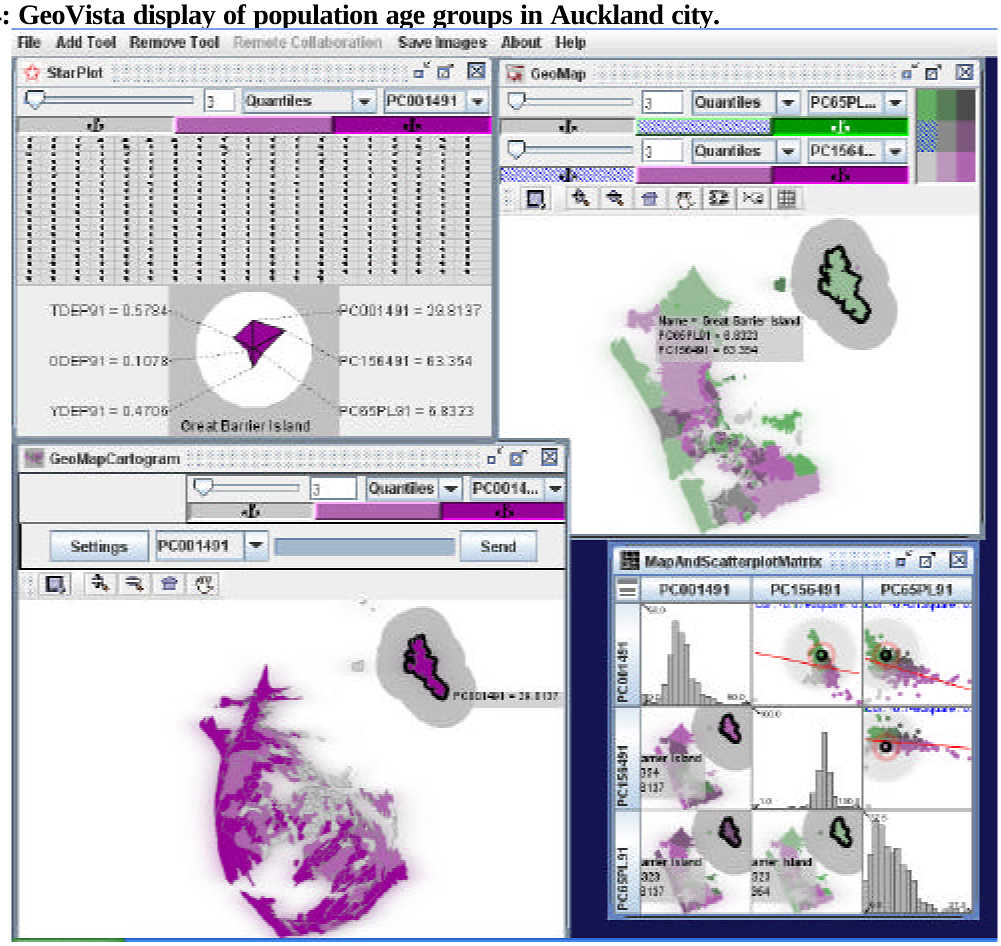
Figure 3: New Zealand Commuter Flows: Work to Home, Wellington City



Commuter flows for selected ethnic groups, modes of transport, industry and occupation can be shown and used by students to explore a variety of social issues, such as public transport planning, demonstrating the multi-disciplinary nature of statistics in the real world. There are exciting trends in open source GIS systems in addition to Google Maps, such as Quantum GIS (http://www.qgis.org), SAGA GIS (http://www.saga-gis.org) and GRASS GIS (http://grass.osgeo.org) that remove barriers to access to sophisticated technology.

The latest visualization tools combine graphs and maps as well as data analysis. One innovative product that integrates maps and graphs is GeoVista, a Java based open-source software product created at Penn State University (http://www.geovista.psu.edu/grants/cdcesda). Tools such as this could be viewed as primarily for specialist researchers or academics but they also provide a way for policy analysts to explore the multivariate characteristics of complex official statistics datasets and to visually link these characteristics to their underlying geographic patterns. The identification of geographic clusters is important when policy interventions are being designed, and many issues can be explored using the tool. Figure 4 displays, using 2006 Census data, the proportions of the population of each area in Auckland City that are children (aged 0-14), working age (15-64) and elderly (65 and over) together with the youth dependency (total children divided by total working age population), old age dependency (total elderly divided by total working age population) and total dependency (sum of children and elderly divided by total working age population) ratios in related star plots. Each point in the scatterplot matrix for each of the two-way population proportions is linked to its geographic area. A univariate cartogram indicates that it is the outer areas of the city (the suburbs) that have the highest proportion of children.

Figure 4: GeoVista display of population age groups in Auckland city.



As with the flow visualisation above this tool demonstrates to the learner the multivariate and interdisciplinary nature of official statistics. It also raises the question of what will be the impact on both the teaching and the analysis of statistics when the linking of statistical data with its underlying geography is so easy that it is no longer appropriate to analyse separately from this dimension.

CONCLUSION

Even from just the few examples given above it is clear that new data visualizations not only allow us to look at currently existing data in new ways but also provide a way for us to visually demonstrate statistical concepts. The potential educational uses of dynamic graphs (that move over time), integrated graphs and maps and combinations of these is a virtually untapped area for further exploration. The increased ability to link data with its underlying geography raises issues for both the analysis and the teaching of statistics.

REFERENCES

de Róiste, M., Gahagen, M., Morrison, P., Ralphs, M. & Bucknall, P. (2009). *Geovisualisation and policy: exploring the links.* Official Statistics Research Project Report. Statistics New Zealand.

Dominguez-Dominguez, J. & Dominguez-Lopez, J.A. (2010). A visual approach to the teaching of statistics and probability. Invited paper presented to the Eighth International Conference on Teaching Statistics (ICOTS 8). Available at www.stat.auckland.ac.nz/~iase/

Forbes, S., Ralphs, M., Goodyear, R. & Pihama,N. (draft). *New ways of visualising official statistics.* Working paper series. Information and Decision Support Centre. Cairo.

Gal I. (2002). "Adult's statistical literacy. Meanings, components, responsibilities." *International Statistical Review*, 70(1), 1-25.

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

Hidalgo, C. A. (2010). '*Graphical Statistical Methods for the Representation of the Human Development Index and its Components'* United Nations Development Programme, Human Development Reports, Research Paper 2010/39. United Nations, New York.

Jackson, N. (2001). *The policy-maker's guide to population aging: key concepts and issues,* Policy Research Paper No 13. Department of Family and Community Affairs. Commonwealth of Australia.

Konold, C. & Miller, C. (2005). *Tinkerplots. Dynamic Data explorations.* Emeryville, California. Key curriculum Press.

Paparistodemou, E. & Meletiou-Mavrotheris, M. (2008). Developing young students' informal inference skills in data analysis. *Statistics Education research Journal*, **7**(2), 83-106. International Association for Statistical Education. www.stat.auckland.ac.nz/serj

Rosling, H. (2007). Gapminder. GapMinder Foundation www.gapminder.org

Statistics New Zealand (1999). *A Layperson's guide. CPI. All about the Consumers Price Index.* Statistics New Zealand. Wellington. New Zealand. www.stats.govt.nz.

Statistics New Zealand (2008) *Consumers price index resource.* Retrieved 30 October 2010. www.stats.govt.nz/surveys_and_methods/our-surveys/cpi-resource.aspx.

ten Bosch, O. & de Jonge, E. (2008). Visualising official statistics. *Statistical Journal of the IAOS: Journal of the International Association for Official Statistics.* Vol.25, No.3-4. IOS press.

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

Tucker, C. (2010). Recruitment, Training and Retention of Statisticians in the U.S. Federal Statistical Agencies. *Journal of Official Statistics*, **26** (3), 455–464

Wild, C & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry (with discussion). *International Statistical Review*, **67**(3), 223-265.

Wickham, H. (2010). Using visualisation to teaching (sic) data analysis and programming. Invited paper presented to the Eighth International Conference on Teaching Statistics (ICOTS 8). Available at www.stat.auckland.ac.nz/~iase/

Contributed paper – Kay Lipson and Glenda Francis

COMPARISONS BETWEEN MARKETING AND PSYCHOLOGY STUDENTS IN LEARNING STATISTICS

FRANCIS, Glenda and LIPSON, Kay
Faculty of Higher Education Lilydale, Swinburne University of Technology, Australia
klipson@swin.edu.au

Abstract
This study investigates how the attitudes of marketing and psychology students towards statistics differ and whether the differences are inherent from the start or develop as they progress through their degree. The attitudes towards statistics for final year marketing and psychology students were measured using a slightly reduced version of Schau's 36 item SATS scale (Schau, 2005). The third year marketing students were shown to have much less positive attitudes towards statistics than their psychology counterparts, and perceive statistics as less useful to their discipline. Marketing students see only modest value in statistics at the start of their program, and unfortunately this worsens over time, while psychology students see statistics as more valuable from the start, and increase this view over their program. Both groups of students start out with almost the same level of interest in statistics but again this reduces markedly for the marketing students while level of interest increases for the psychology students. Psychology students' perception of the level of difficulty of statistics stays relatively constant during their program. While Marketing students see statistics as less difficult at the outset than psychology students, for them statistics becomes more difficult over their course of study. It is suggested that this difference may be as a result of differences in the two course structures and that embedding statistics more fully into specific discipline areas improves student' attitudes and helps to prepare them better for the workplace.

Topic: Learning strategies

Contributed paper – Ian Gordon and Sue Finch

REALSTAT: FROM IDEAS TO DATA AND BEYOND

GORDON, Ian and FINCH, Sue
University of Melbourne, Australia
irg@unimelb.edu.au

Abstract
RealStat is a suite of multimedia case studies in statistics, designed to be accessed on the internet and used in mainstream statistics teaching. We describe the educational philosophy behind RealStat, outline its development, and present several examples. The case studies have been used at every level of undergraduate courses at the University of Melbourne, and at postgraduate level. They are designed with levels of complexity and a variety of relevant statistical techniques and applications. We aim to inspire students with rich context and material, including visual media, graphics, context background, commentary by a statistical analyst and so on. Students can explore the data and carry out their own analyses. RealStat has been used in lectures, assignments and examinations. We describe some of these uses, and the experiences of lecturers in developing curriculum material based on RealStat.

EDUCATIONAL PHILOSOPHY

"If you have only pretend data, you can only pretend to analyse it" (Watkins, Schaeffer & Cobb, 2008). In the evolution of statistical education, it is now recognised that students' learning should be grounded in real world applications to encourage the development of statistical thinking along with an understanding of the importance of data collection methods and study design. Statistical thinking relies on statistical knowledge, context knowledge and the information in data (Wild and Pfannkuch, 1999). "Context is an essential part of statistical thinking, and some of the worst teaching of statistics occurs when the teacher or textbook tries to treat context as irrelevant" (Cobb, 2007, p.338).

A common complaint about statistics majors is that they have not learned to solve real problems. This is also a difficulty for student learning statistics in service courses (Singer and Willett, 1990). The need to engage these students with genuine relevant data has been identified in many applied disciplines (e.g. Love and Hildebrand, 2002; Paxton, 2006; Singer and Willett, 1990).

The context of data-based enquiries is often rich and complex. Useful characterisations of this complexity are Wild and Pfannkuch's (1998) "worry questions", Utts' (2004) seven critical components, and the cycle of statistical enquiry that moves from problem through plan, data, analysis and to conclusions (Wild and Pfannkuch, 1999). These characterizations reflect the breadth of aspects of the context that a statistician must consider in making judgements about data and inferences from analysis. It is this richness that we believe students must learn about in order to develop their skills in the *art* of statistics (Pfannkuch and Wild, 2000) along with their capacity to carry out appropriate analyses.

RESOURCES

A variety of resources have been developed in response to the idea that statistics education should engage students with genuine applied empirical research. The first generation of texts included Freedman, Pisani and Purves (1978) *Statistics*, now in its fourth edition, and Cox & Snell's (1981) *Applied Statistics.* The next generation of introductory texts, like Moore and McCabe (1999), and those that followed typically include examples from real problems. While this is a welcome improvement, many text-book examples based on genuine data are sanitized and both the description and thinking about the context are often omitted or highly simplified.

Hand's (1994) *A Handbook of Small Data Sets* is a unique book that provides a large range of genuine data with brief contextual information. Texts such as Nolan and Speed's (2000) *StatLabs* and Peck, Haugh, and Goodman's (1998) *Statistical Case Studies* took a further step in attempting to provide richer examples of applied statistical problems.

The internet has open access to data through the web pages of statistical journals, specialist sites including the *Data and Stories Library* and the Australian resource Oz*Data and Stories Library*, and official sources and government sites. A vast array of statistical teaching and learning tools, such as Java applets, are available via the internet, but these are rarely integrated with data sources. Again, although web-based data resources do not suffer the space limitations of texts, the context detail provided is typically very limited. The same problems often arise with data sets that are integrated in statistical packages, such as Minitab and R.

Good examples from these sources get used repeatedly, and often after initial enthusiasm the development of specialist site stagnates. The potential of the internet in providing context rich material to teachers and learners of statistics has not been exploited much. Data abounds but context is given short shrift. The US series of videos *Against All Odds: Inside statistics* (Consortium for Mathematics and its Applications), released in 1989, illustrated the benefits of using rich media in characterising and understanding the complexity of data in context.

We argue that good statistical thinking about data in context requires access to an array of different kinds of information, which are essentially the components of the context: the wider context, the political framework, the funding, the design and sampling structure, the measurements and interventions, and details of what actually happened possibly in contrast to the plan. Statistical thinking encompasses asking appropriate questions and using suitable methods for inference in order to draw valid conclusions. It requires effective forms of communication, often supported by simple and focused representation. As Wild and Pfannkuch (1999) note, the end of the cycle of enquiry about one question naturally leads to consideration of what to do next. In developing RealStat, we aimed to provide a resource of statistical case studies that provided access to this complexity, giving an educationally appropriate exposure to context.

DESIGN OF REALSTAT

Our design of RealStat aimed to make the most of flexibility of the internet in providing access to a range of different types of information and materials for each statistical case study. We have included background material written in collaboration with the researcher, graphs that represent the data and model effective use of graphics, and images such as visual displays of the design and photographs of study sites or experimental setups. The resources also include QuickTime movies of the researcher and/or a statistician discussing the project; some of these movies were taken in the field where an experiment had been done. Students thus have direct exposure to the researcher's ideas and questions, rather than having them only as a neatly summarized research question or hypothesis, or (worse) needing to speculate about the aims of the research. Each case includes a commentary by a statistician. The data can be accessed by a downloadable Excel file.

The material for each case is organized in six major sections, accessed via a set of tabs that appear on each page. The different sections are the introduction, background, study design, data collection, analysis and data. In some cases, additional tabs link to glossaries. We illustrate the kind of detail found in these sections in an example in the next part of this paper.

RealStat is designed as a resource for teachers and learners, rather than as a set of materials that might correspond to a structured curriculum. In choosing case studies, we looked for examples that had features or issues that could be considered by students at many levels of statistical sophistication. We hope that students may work on a case study at an introductory level and return to it and revisit the data as their skills develop. We wanted cases that varied in their complexity, but all have the depth of the real problems where thinking and 'research' are required to carry out and interpret appropriate analysis. This also meant that we chose cases that had several (rather than one) interesting feature or result.

The material in RealStat is designed to reflect the openness of questions that can be addressed with data. We provide questions for students to consider and provide some hints and guidance, but no answers. There is no direct prescription of how to answer a question. The first questions asked are simpler and more guided, and they become increasingly open. All the questions we ask require statistical thinking rather than relying only on producing ‘output’. This also meant that we needed to make a careful choice of the visual representations and discussions of the analyses that we provided on the site; we included useful informative graphics and simple summaries of the interpretation of the primary analyses without explicitly answering the questions we asked for students to answer.

Hoti, Francis and Lancaster (2010) argued that good teaching datasets have four important characteristics. They should be topical and relevant, as well as freely available. There should be good documentation of the measured variables and context, supplemented by publications. At present, most of our case studies meet all these criteria; in some cases, publications are not yet available. In general, this will be corrected over time as the publications become available. All case studies used have arisen through projects seen in the Statistical Consulting Centre at the University, and therefore have a history that is primarily local and current, enhancing their appeal in Australian contexts.

SOME EXAMPLES

There are currently eight case studies, with another four under development. The case studies are based on:

- A legal dispute about data from a weighbridge at a garbage tip
- A randomised controlled trial of music therapy for terminally ill patients
- An observational study comparing the health of different arteries from cardiac surgery patients
- A comparison of four methods of harvesting seeds from a native lily
- A randomised controlled trial of a palatial anaesthetic for wisdom teeth removal
- A randomised controlled trial of exercise, vision improvement and home hazards removal in preventing falls in the elderly
- A field trial of three undervine treatments in growing wine grapes
- A field trial (Latin square) of mowing and fertilization in the success of growing native flowers in native grasslands

The planned additions to the case studies include a meta-analysis of cancer rates in fire-fighters, a community survey of pet owners, a longitudinal survey of trachoma in outback Australia, and a study of people with colour vision problems.

To illustrate the kind of content provided, we describe the field trial of mowing and fertilisation. Figure 1 shows the Analysis page from this case study. The Introduction page gives a very brief overview of the study including the primary question of interest, outlines the timeline, and includes a video of a statistician interviewing the researcher in front of the experimental plots. The Background page has a more detailed discussion of the motivation for the study, the final report on the study written by the researchers, a video of the researcher discussing an unanticipated finding, and photos of the site over the course of the study. The study used a four by four Latin square design, based on a combination of two levels of mowing and two levels of fertilisation. The Study Design page has discussion of the design, and a graphical representation of the design alongside a video of the experimental site. It lists the measurements taken, and includes a video of the researcher discussing the choice of the design and a graphic giving a detailed layout of the experimental plots. On the Data Collection page you find details of the protocol, a study log, and a video discussing the experimental site. Illustrative graphs and “questions to consider” appear on the Analysis page, and the data with definitions of variables in the data file are available on the Data page.

Other types of information available in other case studies include discussions of estimation of sample sizes, examples of randomisation envelopes, photos of data collection equipment, examples of consent forms, details of randomisation procedures, and illustrations explaining important aspects of the content. Details of funding and links to publications are included.

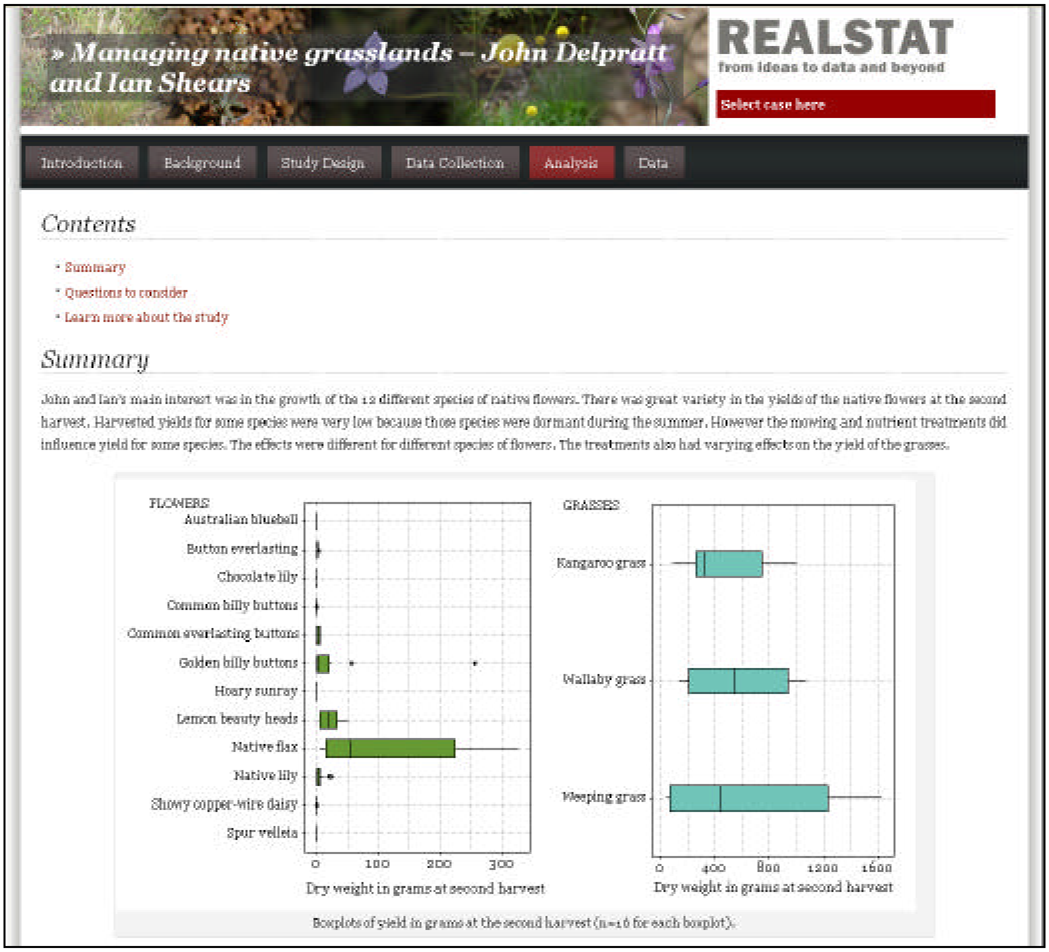


Figure 1: Part of the Analysis page from the study of native flowers in native grasslands

EDUCATIONAL USES

To date, RealStat has been used in courses run in the Department of Mathematics and Statistics at The University of Melbourne, at all levels from first year to postgraduate. Lecturers have found the material easy to integrate into lectures and the access to visual representations, photos and graphics has been particularly useful in illustrating concepts related to study design and interpretation.

Postgraduate students from a wide range of applied disciplines taking an introductory course in statistics take an examination at the end of the course, conducted in a computer laboratory. The exam has been based on a case study from RealStat; students' first experience of RealStat is during the exam where they are directed to examine relevant material. The exam questions ask about measurement variables and issues in the study design, as well requiring students to carry out and interpret analyses. The quality of students' interpretation of analyses can also be evaluated. RealStat has been used successfully in this way for several years; the volume of material available to students during the examination has not been problematic. This is handled by using simpler case studies and by directing students to relevant parts of the case study.

Undergraduate students have used RealStat in a laboratory class setting to complete work requirements based on a case study. Teachers in these classes report that students are highly engaged with the material. The primary use of RealStat to date has been in setting assignments. Take home assignments for students at varying undergraduate levels and in masters courses have been set with RealStat. The material provided in RealStat has allowed lecturers to ask questions that assess the quality of students' statistical thinking. This can be achieved by probing understanding of the context. Students can be asked to take the perspective of the researcher or of the statistical consultant and to provide context-rich explanations of statistical information. These assessment tasks can be initially confronting to some students as they are unlike any statistics assignments they have had before. RealStat provides lecturers with an ideal resource for extending students in this way, by injecting the crucial ingredient so often absent historically: context.

FUTURE PLANS

The development and addition of case studies to RealStat is primarily limited by the availability of developers' time. We plan to continue adding cases over time. At present, RealStat has only been available to students at the University of Melbourne. We hope that it will be available more widely in the near future.

The development of RealStat has been funded by a SciMet grant from the TaLMET committee and by the Statistical Consulting Centre at the University of Melbourne.

REFERENCES

Cobb, G. (2007). One possible frame for thinking about experiential learning. *International Statistical Review*, 75(3), 336-347.

Consortium for Mathematics and its Applications (1989). *Against All Odds: Inside Statistics*. Series of 26 videos. Produced by Annenberg/CPB. US: PBS.

Cox, D.R., & Snell, E.J. (1981). *Applied Statistics: Principles and Examples*. London: Chapman and Hall.

Freedman, D., Pisani, R., & Purves, R. (1978). *Statistics*. New York: W.W. Norton and Company.

Hand, D. (1994). *A Handbook of Small Datasets*. London: Chapman and Hall.

Hoti, V., Francis, B., & Lancaster, G. (2010). Resource discovery for teaching datasets. In C. Reading (Ed.), Data and context in statistics education: Towards an evidence-based society. *Proceedings of the Eighth International Conference on Teaching Statistics* (ICOTS8, July, 2010), Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute.

Love, T. E., & Hildebrand, D. K. (2002). Statistics education and the "Making statistics more effective in Schools of Business" conferences. *The American Statistician*, *56*(2), 107-112.

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

Nolan, D.A., & Speed, T. P. (2000). *Stat Labs: Mathematical Statistics through Applications*. New York: Springer.

Paxton, P. (2006). Dollars and sense: Convincing students that they can learn and want to learn statistics. *Teaching Sociology, 34*(1), 65-70.

Peck, R., Haugh, L.D., & Goodman, A. (1998). *Statistical Case Studies: a Collaboration between Academe and Industry.* ASA-SIAM Series on Statistics and Applied Probability (No. 3), Philadelphia: American Statistical Association and the Society for Industrial and Applied Mathematics.

Pfannkuch, M. & Wild, C. (2000). Statistical thinking and statistical practice: themes gleaned from professional statisticians. *Statistical Science, 15(2*), 132-152.

Singer, J. D., & Willett, J. B. (1990). Improving the teaching of applied statistics. Putting the data back into data analysis. *The American Statistician, 44*(3), 223-230.

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

Watkins, A.E., Scheaffer, R.L., & Cobb, G.W. (2008). *Statistics in Action: Understanding a World of Data* (2^{nd} edition), Key Curriculum Press, Emeryville, CA.

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).

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

Contributed paper (refereed) – Graham Barr and Leanne Scott

SPREADSHEETS AND SIMULATION – A NEW WAY FORWARD FOR TEACHING STATISTICS

BARR, Graham and SCOTT, Leanne
Department of Statistical Sciences, University of Cape Town, South Africa
graham.barr@uct.ac.za

Abstract
Fundamental statistical concepts remain elusive to many students in their introductory course on statistics. This talk will focus on the teaching of statistics within a spreadsheet environment, wherein the students are required to master the basics of Excel to perform statistical calculations. This approach has the advantages of developing the students' ability to work with data whilst also building an understanding of the algebraic relationships between elements embedded in the formulae which they use. The authors demonstrate the use of a classroom experiment aimed at exploring the distributions of a number of randomly generated pieces of information. Teaching sessions are built around a suite of Excel based simulations based on this experiment that attempt to demonstrate the concept of random variation and show how statistical tools can be used to understand the different underlying distributions. Some preliminary results are given of an investigation which assesses students' perceptions and understanding of the approach.

INTRODUCTION

Teaching mathematical and statistical principles through a spreadsheet platform offers significant advantages. The structuring of a spreadsheet develops a general algebraic way of thinking as the process requires skills in expressing numerical relationships using algebraic notation. In the field of Statistics, the advantages of spreadsheets for teaching purposes are particularly marked as spreadsheets can simultaneously present an easily navigable yet extensive vista of numeric information stored in multiple rows and columns, along with the formulaic links between them, as well as an associated rich graphical depiction of the same data. However, the richest feature that a spreadsheet offers to the teacher of Statistics is its ability to show how one can mimic the process of repeated statistical experiments. By simulating statistical sampling one can reveal a range of subtle and often misunderstood ideas which are central to basic statistical knowledge, such as those of randomness and statistical distributions. Moreover, beyond the basic structure of the spreadsheet which lends itself so well to the teaching of statistical ideas, MSExcel, the most often used spreadsheet in academia and commerce, has an extremely powerful built-in programming language, Visual Basic for Applications (VBA). This allows teachers and students to enhance and leverage basic Excel power and functionality to a new level of flexibility and sophistication with click button automation and slickness.

In this paper we will relate how our experience with teaching first year Statistics courses at University of Cape Town, UCT, has shown that the key concepts of randomness and distribution are most effectively taught through simulation in an MSExcel based spreadsheet environment using a 2-stage approach, first using a formula based spreadsheet and, subsequently with the enhancement of VBA. The focus of the paper is a carefully crafted teaching example which demonstrates to students how the random sampling of a set of distinct attributes associated with a set of individuals may reveal completely different distributions of each attribute.

SPREADSHEET LEARNING IN THE MATHEMATICAL AND STATISTICAL SCIENCES

As early as 1985, two years after the launch of Lotus 1-2-3, the spreadsheet had been recognized as a force in statistics education (Soper and Lee, 1985). It is now universally recognised that the two-dimensional structure of spreadsheets, along with their associated graphical components, can facilitate the comprehension of a wide range of mathematical and statistical concepts by providing a supportive platform for conceptual reasoning. Baker and Sugden (2003), for example, give a comprehensive review of the application of spreadsheets in teaching across the mathematical, physical and economic sciences. The idea of using simulation, especially within VBA, was found by Barr and Scott (2008) to be particularly useful and effective for the teaching of first year statistics to large classes and they confirm the sentiments of Jones (2005) that statistical concepts and procedures taught within the context of a spreadsheet tend to be transparent to pupils, allowing them to look inside the "black box" of statistical techniques. A comprehensive survey of the use of simulation methods for teaching statistical ideas has been done by Mills (2002). It is argued that the literature supports the notion that spreadsheets lead students into a didactically rich and effectively open-ended line of inquiry with MSExcel as the de facto spreadsheet standard.

TEACHING THE NOTION OF RANDOM VARIATION WITH EXCEL

The core part of this paper is to showcase the teaching of the two foundational statistical concepts of *randomness* and underlying *distribution* through simulation using both simple spreadsheet functions and more sophisticated VBA programs. Our experience has lead us to believe that VBA-structured spreadsheets by themselves provide a difficulty for a large cohort of students; a leap into the dark to some extent. However, when properly scaffolded by a standard spreadsheet with formulae approach, it becomes a more effective learning tool. By itself, VBA simulation programs or simulation programs written in Java on the web are neat and impressive but constitute too much of a 'black-box' for students. Leading students through a formulaically structured approach on the spreadsheet first, thereby putting the appropriate building blocks in place to support subsequent exposure to VBA programs, leverages the first-tier analysis to a second, more accessible level.

HOW WELL IS THE NOTION OF RANDOM VARIATION UNDERSTOOD?

One of the fundamental concepts of statistics is that of random variation. It is a notion that we as statistics educators frequently assume people have an intuitive understanding of. It is, however, a subtle notion that apparently random and unpredictable events have underlying patterns that can be uncovered through (inter alia) long term observation.

An open ended invitation to describe their understanding of 'randomness' and how it affects our day to day lives was extended to a group of twenty adult learners, all of whom were tertiary educators in non-quantitative disciplines themselves. A variety of notions of randomness were articulated, from which some unexpected themes emerged. All of the descriptions volunteered by the students were devoid of any notion of underlying pattern or distribution. Subsequent discussions confirmed that they believed the existence of an underlying pattern was, in fact, contradictory to the very idea of random variation. It is suggested that beginning the statistics journey with a description of the world as containing innate patterns and order which are hidden from us through the random and unpredictable way in which individual outcomes are free to vary, may open up the power and interest of the discipline in a way that the traditional approach of teaching 'theory followed by its application' fails to do. We show below, using an appropriate experiment, that the spreadsheet environment is an ideal canvas on which to sketch and unveil the ideas around random variation and underlying patterns.

SO HOW DOES ONE CONVEY THE CONCEPT OF RANDOM VARIATION? *THE CLASS EXPERIMENT – RANDOM SELECTION; DIFFERENT PATTERNS!*

We begin the class experiment with a discussion with the students about different types of numbers, reflected both by the different measurement scales we choose to assign to them, and by the process that generates them. We ask them to consider the following experiment in which each student in the class will contribute four pieces of information, viz: (1) their first name; (2) their height (in cm); (3) an integer randomly selected within the range 1 to 50; and, finally, (4) their personal results of a (to be explained) experiment involving mice! Once we have generated this data we will be collecting it from everyone and constructing four separate graphs of each of the four information types. As part of this experiment we will be constructing ways to generate data for (3) by using, and exploring, the Excel random number generator. Data for (4) involves a mouse training experiment which tests the ability of 5 (simulated) randomly selected mice to navigate a simple maze, recording the number of successful mice. Students will be able to run the Excel models to record their own data for the class experiment. The focus of this class exercise is for students to answer the key question: *What shapes do we anticipate for these graphs?* We proceed by considering the randomly generated number.

THE RANDOM NUMBERS

Suppose we are interested in mimicking the National Lottery and (repeatedly) generating a random number which lies between 1 and 50. Each draw can be likened to drawing a number from a hat with the numbers 1 to 50 in it. If we want to keep drawing a number from this hat in such a way that all numbers are equally likely, we would have to also suppose that we have a very large hat that can hold such a large (and equal) quantity of each of the numbers that it doesn't limit our thinking about the situation. What would a histogram of these numbers look like? Some discussion would probably lead us to conclude that we would expect all of the bars in the histogram to be of equal height. Now let's see what sort of patterns we get when we randomly take numbers out of the hat. The fact that the number is randomly selected means there is no way of telling exactly what number is going to pop up next. In order to see a pattern of numbers we need to observe more than one randomly drawn number. Let's see what happens when we generate 10 random numbers. Perhaps the pattern looks a little different from what we might have expected. What happens when we draw 100 numbers; or 1000? It seems that as the pool of numbers that we are drawing grows, so the pattern of the numbers in the hat gets revealed. A small pool can give quite a misleading picture of the histogram of the numbers in the hat! However, a big pool is less likely to do so. So how big a pool of numbers do we need to have access to in order to get a reliable picture of the numbers in the hat? Imagine that we hadn't known the shape of the histogram of the numbers in the hat. The numbers might, for example, have had a different (other than flat/ rectangular) pattern of distribution. Let's use some spreadsheet commands to model our thinking of the above. We can easily simulate the drawing from our hat of a number (where all the numbers in the hat are integers that lie between 1 and 50 inclusive) by typing the formula: =(RANDBETWEEN(1,50)) into our spreadsheet. This computes (and rounds to the nearest integer) a single random number in the interval (1, 50). We can then resample this number by pressing the F9 key. We can also display this visually by plotting the number in a simple histogram. We should first set up some bin intervals and use an array formula to compute the frequencies: ={FREQUENCY(data_range, bin_range)} which can then be plotted in a histogram. We replicate this procedure for a sequence of sample sizes from 1 (single random number), through to 10 (10 random numbers), then 100 and finally 10 000. By repeatedly pressing F9 (which simply recalculates the formulae and effectively re-samples) we get to replay the (random) selection of different sized pools of numbers from the hat. One feature becomes apparent. The pattern (of the numbers in the hat) becomes increasingly clearly revealed as we observe larger and larger pools of numbers from the hat, see Figure 1. Although we cannot *at any stage* predict what the next number will be, by observing randomly selected pieces of information from the hat we can begin to piece together what the pattern of numbers in the hat must look like. In real life this is likely to mean we have to observe the pattern (of randomly revealed pieces of information) over time, before we can begin to understand something about the nature (or distribution!) of the numbers in the hat.

We can then use VBA to extend these ideas. The VBA version of the uniform distribution provides a neat and slick point and click demonstration of how simulated uniform distributions can be generated. A key feature of this is the ability to compare the theoretical (or expected) frequency graph with the empirical (that is the one generated by simulation).

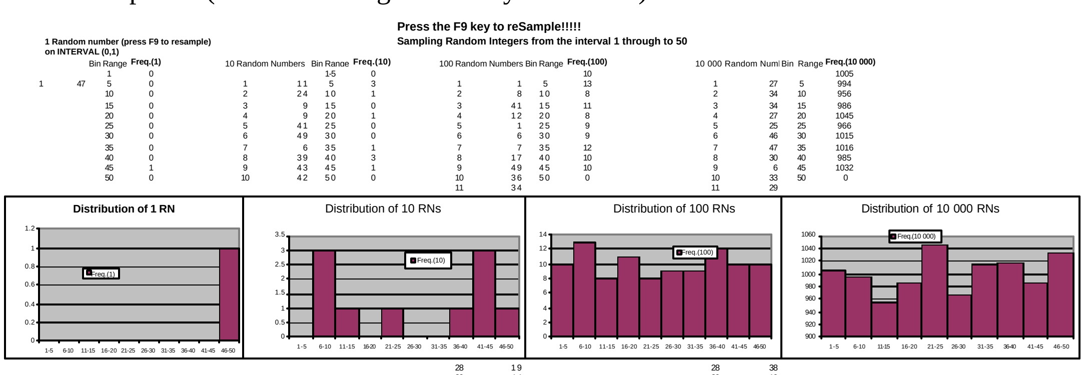


Figure 1. Simulation results from a Uniform Distribution for n = 1, 10, 100 and 10,000

THE MOUSE EXPERIMENT: GENERATING DATA FROM A BINOMIAL DISTRIBUTION

The uniform case above constitutes a starting point and necessary platform to consider more complex distributions. In particular, as it allows us to transparently allocate a binary outcome probabilistically, it leads naturally onto the Binomial Distribution. We consider a fixed number of trials, where at each trial we have assigned a fixed probability of success or failure (i.e. a uniformly generated random variable on (0,1) indicates a success if U is between 0 and p and failure otherwise). In order to generate the fourth piece of information for our class experiment we consider the following scenario: We, as statisticians, have been approached to mediate on a claim that an animal trainer has managed to train a group of ten mice to turn left at the end of a tunnel. To test this claim we select 10 mice, put them through the tunnel and record the number of successful mice, i.e. the number out of the 10 who turned left. We could repeat this many times, sampling without replacement. The mechanics of implementing this on the spreadsheet are as follows: We select a random number between 0 and 1. If the number is less than or equal to 0.5 we assume the mouse went left, if not we assume it went right. Then we do this for 10 mice, labelling the results Trial 1, through to Trial 10. This comprises one experiment; in cell C3 we put the value of p, in this case 0.5. The following formula:=IF(RAND()<C3,1,0) results in a 1 if the random number calculated is less than 1 and a 0 if it is not (that means it is greater than 1). We then copy this formula down a further nine cells to get the model results for one experiment. In this case, Trials 2, 7 and 10 were a "success" (mouse turned left). Trials 3 to 6 and 8 and 9 were failures (turned right). Excel allows us to repeat the experiment by pressing the F9 (recalculate) key. Next time we might get 6 successes for the same value of p. By repeatedly pressing F9 we see we can generate a series of results, sometimes the same, sometimes different. These simulated results form a pattern for each particular p. For example out of 10 mice we often get 4, 5 or 6 who are successful. We may repeat this experiment 50 times and store the results of many random experiments by copying the set of formulae across a number of different columns. The results are shownin Figure 2. Note that we have repeated the 10 trial experiment 50 times and revealed an interesting histogram. There are relatively high frequencies for 4, 5 and 6 successes, a number for 2, 3, 7 and 8 successes, BUT none for 0, 9 and 10 successes. The indications are that the training is partially effective (effective training might equate to a p parameter of 0.8). We don't know what the true p value is but it does seem that the observed data are consistent with a p of 0.5

p	0.5	
n	10	
		Ex1
	Trial 1	0
	Trial 2	1
	Trial 3	0
	Trial 4	0
	Trial 5	0
	Trial 6	0
	Trial 7	1
	Trial 8	0
	Trial 9	0
	Trial 10	1
	#Successes	3

# Successes	Frequency for 50 Experiments
0	0
1	1
2	3
3	4
4	11
5	12
6	9
7	7
8	3
9	0
10	0

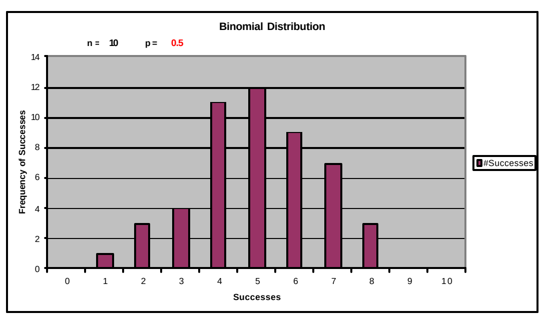


Figure 2. Simulation results from a Binary Distribution for n = 50

USING VBA TO EXTEND AND STREAMLINE THE APPROACH

Our teaching approach over the last few years has been to focus on teaching things like the binomial distribution primarily through the means of a VBA simulation. Similar simulations exist on the Web written in Java. On reflection, however, we have come to the conclusion that jumping straight to a statistical simulation is too 'black-boxy', that is, it is graphically impressive and illuminating to a significant extent but, because no (spreadsheet) building blocks have been shown to students, it is impenetrable and intimidating for the average student. The approach we recommend here is a 2-stage approach where the ideas in stage-1 are explained through Excel sheet formulae up to a certain level, as expounded upon above. Then, when the basic ideas have been properly internalised, the leap is made to showcasing the click-button-and-fancy-graphics world of VBA. For engaged students, this transition can also provide an entrée into the power and flexibility of VBA programming. Barr and Scott (2008) give an account of using a suite of simulations written in VBA (including the mouse example) as teaching tools.

PUTTING ALL THIS TOGETHER: MAKING SENSE OF THE CO-EXISTENCE OF PATTERN AND RANDOMNESS

The interesting feature of the results of our 'mouse training' model is the pattern of results it revealed. Each time we repeated the 50 experiments, each consisting of 10 trials, we observed a different set of numbers but they appeared to keep a number of common features. This pattern became clearer the more times we repeated the experiment. The pattern was different from that of the numbers we drew out of the hat. What does this tell us about randomness? What causes the patterns? Remember we said the hat was our mechanism to ensure random selection, in other words to mimic the way data might present itself to us in an unpredictable way. All the numbers were mixed up in the hat and we drew them out in a way that meant no particular numbers were favoured or prejudiced. What if we put different pieces of information (as distinct from random numbers) into the hat, shuffled them and drew samples of them out of the hat? Will the patterns related to different pieces of information all be the same? We are now in a position to conduct our class experiment and find out; we collect from each member in the class the piece of paper bearing the four pieces of information we specified earlier. We construct histograms of the samples of paper as we draw them from our hat. We might not be surprised to observe that the random numbers have a flat, rectangular distribution. We also see that the 'successful mice observed' have the same shaped distribution as the one we saw repeatedly with our electronic mice running model. The heights may show one bell shaped histogram, or may have a hint of two humps of data, with the males being taller than the females. The 'names' may well show a few modes, depending on popular names and prevalence of language groups. Our 'randomising' hat has had the effect of giving us the data in random and unpredictable order, but the distinct patterns associated with each different piece of information have been preserved, and are revealed as we have access to more and more data. In fact, our reconstructing of the data into histograms reinforces two facets of random variation. On the one hand, it is reflected in the unpredictable way we frequently encounter (information in) life (stocks on the stock exchange, increments of growth of children, number of cars on highway at a particular time, etc). But, perhaps paradoxically, randomisation also provides the best mechanism to uncover the true pattern of an unknown measurable (eg household income). Selecting data (sampling) randomly ensures we have the best chance to see as broad spectrum of the unknown pattern of data as quickly (and efficiently) as possible!

EVALUATING STUDENTS' UNDERSTANDING OF THIS NEW APPROACH

It is clearly difficult to assess the value of a new teaching approach in the absence of an unambiguous baseline (or control). We are in the process of developing clearer 'before and after' assessment tools that are of particular interest in the context of a teaching environment that is both multi-lingual and developmental as well as exhibiting one of the world's highest national measures of inequality. The current questionnaires we have used to probe the students' experiences of these approaches have focused on the student's ability to discriminate between empirical and theoretical distributions as our preliminary investigations indicate that these are likely to be new concepts for all students. In a given experimental setting (eg a Poisson process), students are asked to distinguish between distributions which are either empirical or simply reflect the underlying theoretical model. Between 70% and 95% (taken across a number of questions) of around 1000 first year statistics students who participated in the survey answered these questions adequately. Anecdotal evidence has suggested that a lower proportion of advanced statistics students (i.e. of those exposed to traditional teaching approaches) make this distinction correctly, despite these students having passed much higher levels of statistical theory.

Table 1 shows the assessed level of understanding (assessed by a first year statistics lecturer using a subjective scale from 0 to 5) based on students' descriptive responses to the indicated questions. Students were directed to use their own words ('parrot' definitions from text books were excluded from the analysis).

Table 1. Assessed level of understanding

Graded response	What is a pdf?	What is an empirical pdf?	When will the empirical dbn of data look like the (underlying) theoretical dbn?
0 (no understanding)	6%	10%	12%
1	47%	50%	18%
2	17%	19%	8%
3	15%	14%	13%
4	11%	5%	22%
5 (perfect understanding)	4%	2%	27%

Understanding of the concept of probability distributions was generally poor, as was expected to be the case at this introductory level. However, the notion of empirical data was found to be relatively well appreciated. The authors assert that this discrepancy can (tentatively) be attributed to the explicit way in which the notion of a theoretical model is linked to empirical data in the spreadsheet approach. In terms of qualitative assessment, both teachers and tutors who worked through these Excel based lectures and tutorials with the students, reported excitement at various levels of breakthrough. Our future plans for evaluating these approaches include combining both planned qualitative (focus group) type studies with qualitative diagnostics.

CONCLUSION

The spreadsheet has many powerful didactic facets. It provides a flowing, dynamic model which links cells in a transparent way. At the first level it provides a two dimensional, visible, matrix-like calculating machine where at the press of a button the whole matrix may be recalculated. This can be used to simulate samples with different underlying distributions. Randomly sampled bits of information have statistical distributions which reflect how these bits of information are generated and what attributes of life they reflect. A key insight for learners is that although each attribute is selected from the hat of our experiment *randomly* the *patterns* or *distributions* of the attributes can be quite different. This experiment thus constitutes a very powerful mechanism for learner differentiation between the concepts of randomness and the underlying pattern of the attribute itself. Our experience with the extensive use of Excel for first year statistics courses has led us to conclude that a pure VBA approach may be somewhat intimidating for learners and teachers alike so we have adopted a 2-stage approach, where Excel is used firstly in a straightforward formulaic approach, where the formulae are first explained to students in a flowing logical sequence and thereafter the jump to a VBA point-and-click automated approach is made. VBA based spreadsheets allow more flexibility and sophistication and in particular allow automated comparisons showing students how empirical distributions converge on theoretical distributions. We believe this approach is a step forward in the teaching of foundational concepts in first year Statistics, such as randomness and distribution, which are often poorly understood, even by statistics graduates.

REFERENCES

Baker, J. E. & Sugden, S. J. (2003). Spreadsheets in education. The first 25 years. *eJournal of Spreadsheets in Education*, vol *1*(1), (pp18-43). {Available at http://www.sie.bond.edu.au/, visited 25 August 2010}

Barr, G.D.I. & Scott, L.S. (2008). A new Approach to Teaching Fundamental Statistical Concepts and an Evaluation of its Application at UCT, *SA Statist. J.*, 42, 143–170.

Jones, K. (2005). Using Spreadsheets in the Teaching and Learning of Mathematics: a research bibliography, *MicroMath*, vol 21(1), (pp 30-31)

Mills, J. D. (2002). Using Computer Simulation Methods to Teach Statistics: A Review of the Literature, *Journal of Statistics Education*, vol 10 (1)

Soper, J.B. & Lee, M.P. (1985). Spreadsheets in Teaching Statistics, *The Statistician*, vol 34, (pp 317-321)

Contributed paper – David Griffiths and Doug Stirling

REVISITING THE MISUSED, MISUNDERSTOOD AND UNLOVED STEM AND LEAF PLOT

GRIFFITHS, David[1] and STIRLING, Doug[2]
[1]University of Wollongong, Australia
griffd@uow.edu.au
[2]Massey University, New Zealand

Abstract
When John Tukey unleashed some new tools and a new way of thinking about what until then were loosely categorised as descriptive statistics on the disciplinary stage nearly 40 years ago, it has been said that a new culture began.

This paper explores the effectiveness of that cultural change on how we teach and use techniques of exploratory data analysis, as well as embedding EDA into an earlier culture, when many of the EDA tools, including at least two of the supposedly new ones, were already in use.

The stem and leaf plot is one of those not so new tools. The principal foci of this paper are its history, its misuse and its underuse, as well as highlighting novel approaches to the use of stem and leaf plots.

What is its history? What are the catalysts for misuse and underuse? And what are the cures? The first answer is , unsurprisingly, that the history is not what you think. The second is easy to identify and multi-faceted, including materials (print and electronic) and resources (software) , teachers, teaching and professional practice. The third may simply be recognition that the revolution that supposedly happened in the 1970's never did. It was merely a small change in mindset as to how we teach the first week of the first course in Statistics.

May this paper help to bring on the 2010 revolution!

Contributed paper – Michael Bulmer and Daniel Kaplan

INTRODUCTORY EPIDEMIOLOGY WITH A VIRTUAL POPULATION

BULMER, Michael[1] and KAPLAN, Daniel[2]
School of Mathematics and Physics, University of Queensland, Australia
m.bulmer@uq.edu.au [1]
Macalester College, USA [2]

Abstract
We have developed an online population of virtual people that have been used as experimental subjects in student experiments for the last two years. However these virtual people also have genetic histories tracing back 240 years, records of disease and mortality, and a variety of demographic characteristics. We are now exploring this side of the virtual environment with students enrolled an introductory course on epidemiology. We will report on the types and outcomes of studies undertaken by students, their experiences in the environment and our reflections on the role of these experiences in learning.

Contributed paper – Helen MacGillivray

NATIONAL STATISTICAL CURRICULUM JOURNEYS

MACGILLIVRAY, Helen
Queensland University of Technology, Australia
h.macgillivray@qut.edu.au

Abstract
Since 1990, Statistics has featured in Australian school curricula. In the National Statement on Mathematics for Australian Schools (1991), Chance and Data was one of the five syllabi areas, with as many pages of the statement devoted to it as to Algebra. It is somewhat ironic that both these areas have been the source of much angst in the past two decades - of universal importance in each is development of their individual ways of thinking.

Reasons for, and sources of, frustrations in Chance and Data in school curricula are complex and should not be trivialized nor attributed narrowly or simplistically. They have been experienced not just by statisticians and statistics educators, but also by teachers, many of whom are eager to learn about, and teach, statistics in rich and engaging ways. It is informative and sobering to look at some history and timescales in the UK and USA, starting from the 1970's when Peter Holmes led the UK Schools Council Project on Statistical Education emphasizing that statistics should be taught, learnt and assessed in ways that reflect its thinking and practice, with real data and real experiences. This Project informed the American Statistical Association's Quantitative Literacy Project of the 1980s, and the Statistics Focus Group sponsored by the Mathematical Association of America's Curriculum Action Project in 1991 on developments for introductory tertiary courses. Their influences can be seen in the American Statistical Association's 2005 'Guidelines for assessment and instruction in statistics education' (GAISE) college report, and the 2007 GAISE pre-K-12 report (http://www.amstat.org/education/gaise/).

This paper gives a very brief overview of two decades of involvement in statistics education across school and transition levels. This has included designing and delivering many professional development workshops and enrichment programs, working with teachers and educational authorities, and, most recently, significant involvement in national curriculum challenges and writing materials for teachers. Pitfalls, progress, problems and promising avenues for ways forward are discussed. World-leading innovations and developments in statistics education have come from Australia and New Zealand, but there is much to be done in the national context.

Topic: Technology and Statistics

Contributed paper (refereed) – Doug Stirling

DYNAMIC DIAGRAMS FOR TEACHING DESIGN AND ANALYSIS OF EXPERIMENTS

Doug Stirling
Massey University, New Zealand
d.stirling@massey.ac.nz

Abstract
Dynamic and interactive diagrams provide an effective way to explain concepts in introductory statistics courses through simulations and other demonstrations and illustrations. There is similar potential in the use of dynamic diagrams to help teach more advanced statistics but few have been published. A collection of over 200 dynamic diagrams for teaching design and analysis of experiments is described. These are implemented as applets and are embedded within two e-books about agricultural and industrial experiments that are freely available as part of CAST (Stirling, 2010a). Other advanced areas in which dynamic diagrams can potentially help to explain concepts are discussed.

INTRODUCTION

Many concepts in introductory statistics can be demonstrated or illustrated using dynamic and interactive diagrams. In particular, simulations can effectively demonstrate most results about sampling distributions and inference (Chance & Rossman, 2006, Erickson, 2006, and Lane & Peres, 2006). Dynamic diagrams can be generated within general-purpose authoring systems such as XLispStat (Tierney, 1990), Fathom (Finzer, 2000) and rpanel (Bowman et al, 2007), but most have been implemented as Java applets and a web search for "statistics applet" returns many links to applets that can be freely used in teaching. A few e-learning books (e-books) such as Seeing Statistics (McClelland, 2000) and CAST (Stirling, 2010a) make extensive use of applets to help teach introductory statistics.

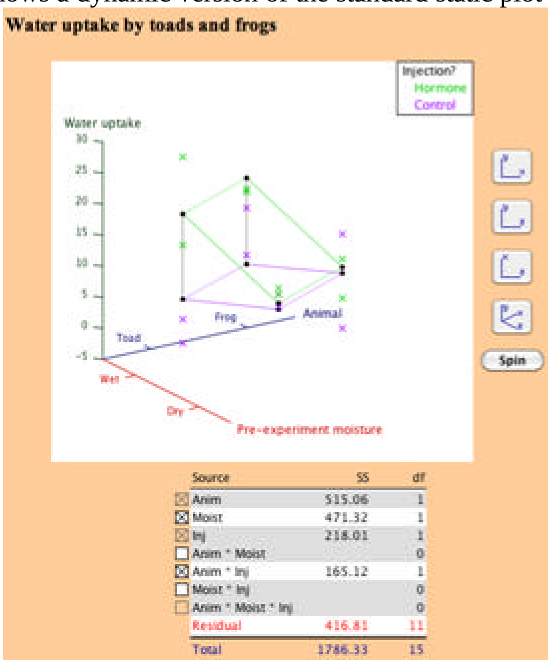
Although there is similar potential in the use of simulations and other dynamic diagrams to help teach statistics beyond an introductory course, few such applets have been published. Advanced concepts are often harder to visualise, algorithms to implement advanced statistical methods are more complex, and the target audience of students is much smaller, making it difficult to justify the design and implementation costs. However the potential of dynamic diagrams in the teaching of advanced statistics can be seen in a CAST e-book about multiple regression that contains over 250 applets (Stirling, 2006).

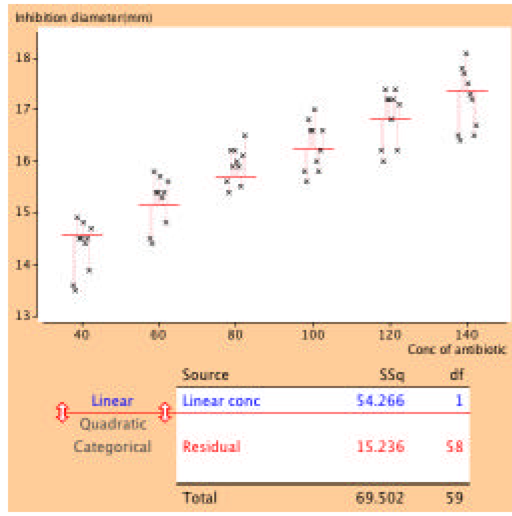
This paper introduces a collection of over 200 applets for teaching the design and analysis of experiments. These applets are embedded in two CAST e-books about agricultural and industrial experiments (Stirling, 2010b and 2010c). The paper focuses on a few concepts that are challenging for students to see conceptually through static diagrams and theoretical exposition, and describes dynamic and interactive diagrams in CAST that target these concepts. Almost by definition, dynamic diagrams are difficult to adequately describe in print, but these examples give a flavour of their potential in teaching.

GRAPHICAL DISPLAY OF MODELS

Models for complex experimental designs cannot be easily visualised but graphical displays can help to explain the properties of simple models. Two examples are described below.

The concept of interaction is central to experimental design but many students find it difficult to understand how interaction terms affect the fitted treatment means. The 3-dimensional diagram on the top left of the following page allows terms to be added and removed (using the checkboxes to the left of the sum of squares table) from models for data from a factorial experiment with three 2-level factors. Studying the changes resulting from adding/removing main effects, 2-factor interactions and the 3-factor interaction helps students to understand the flexibility that is introduced to the model from each term. Rotating the diagram to a 2-dimensional display of the response against a single factor shows a dynamic version of the standard static plot of fitted treatment means.



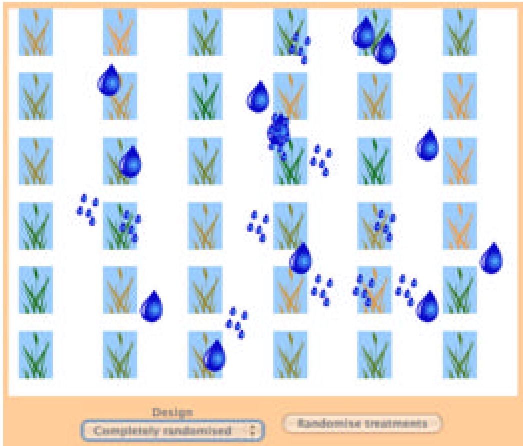


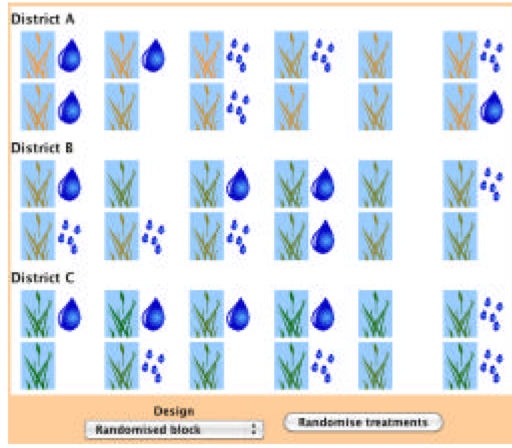
The diagram on the right above shows a jittered dot plot of data from an experiment with a single numerical factor. It illustrates the concept that generalising the model for a factor (linear to quadratic to categorical) reduces the residual sum of squares. Dragging the horizontal red line in the sum of squares table changes the model and exposes the sequential explained sums of squares in the sums of squares table as reductions to the residual sum of squares.

RANDOMISATION

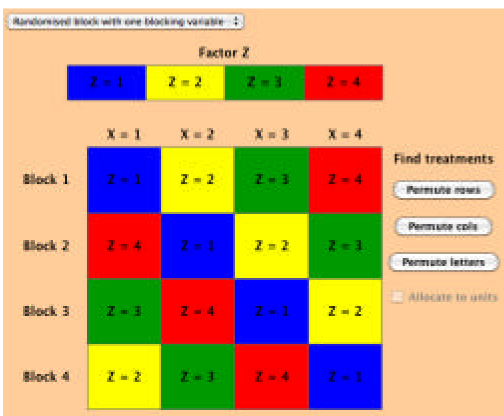
Random allocation of treatments to experimental units is an important part of experimental design. It is therefore important that students understand the different randomisations implied by different designs and can apply these randomisations in practice. Dynamic diagrams can illustrate this randomisation.

The diagram below compares randomisation for completely randomised and randomised block designs. It animates allocation of three treatments (amounts of irrigation, represented by large, small or no raindrops) to 36 plots (represented by grasses of different shades). Changing the design from "Completely randomised" (left) to "Randomised block" (right) animates the grouping of fields into three pairs of rows with icons of similar colours, after which the "Randomise treatments" button animates the randomisation of treatments within the three blocks.





Randomisation for Latin square designs is harder to explain to students (and is often not discussed in courses). The following diagram illustrates randomisation for Latin square designs with one block variable and two controlled factors. The three buttons on the right animate the permutation of rows and columns of the design and the allocation of the levels of factor Z to the four colours. After this permutation, the checkbox "Allocate to units" permutes the treatment combinations for the design to physical units within each block.



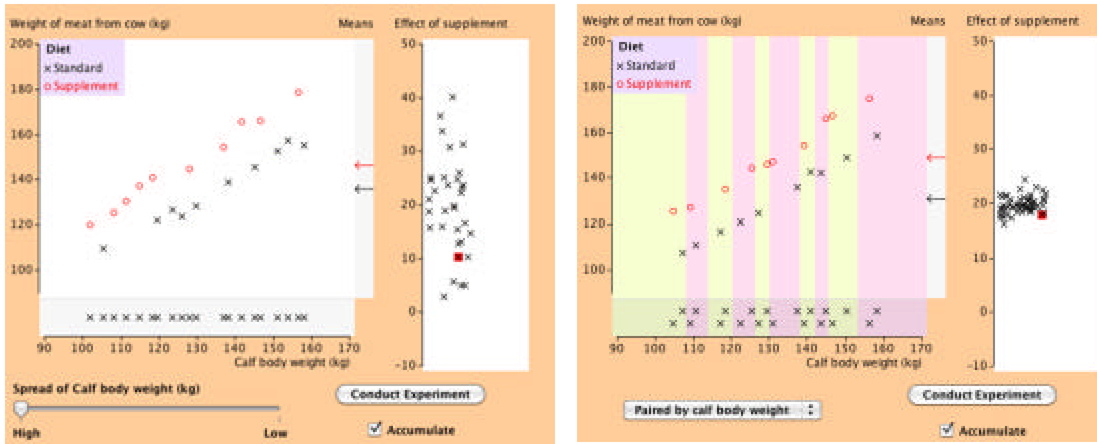
SIMULATIONS

It is important for students to understand that some experimental designs give more accurate estimates of factor effects than others using the same number of experimental units. Simulations are effective ways to demonstrate this.

The following two diagrams simulate experiments about the effect of a feed supplement on the weight of meat produced by cows. The experimental units are 20 calves and initial calf weight also affects the adult weight of meat. (The variation in calf weight and its effect have been exaggerated in the simulation.)

The diagram on the left below demonstrates the importance of reducing variability in the experimental units; it shows that the accuracy of estimating a factor effect is highest if the experimental units are similar to each other. Clicking "Conduct Experiment" randomises the treatments, measures the individual responses and plots the estimated effect of the supplement (difference between the treatment means). Running the experiment repeatedly builds up the sampling distribution of the estimated effect.

Dragging the slider to the right reduces the spread of initial calf weights but retains the other simulation parameters. Repeating the simulation clearly shows lower variation in the estimated factor effect.



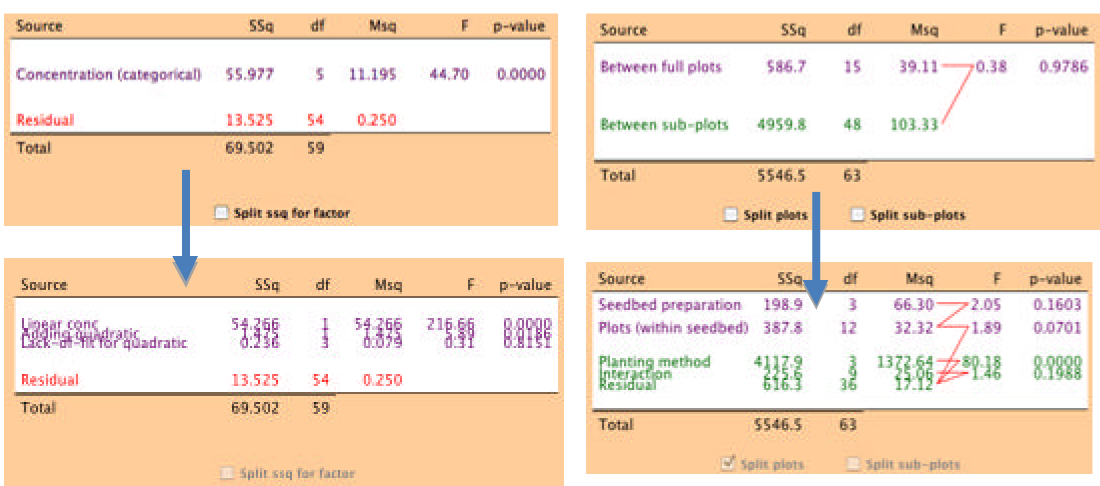
The simulation on the right is based on the same context but is used to show that a better experimental design can improve the accuracy of estimating factor effects even with identical experimental units. The diagram shows results from a simulated experiment in which the calves are blocked by initial weight into matched pairs. The estimated effect of the supplement is fairly accurate — its sampling distribution from the simulation has low spread.

If the pop-up menu at the bottom is used to specify (and simulate) a completely randomised experiment, the spread of the estimated factor effect is higher, similar to that shown in the diagram on the left. This demonstrates that more accurate estimates are found from the block design.

ANALYSIS OF VARIANCE TABLES

Analysis of variance tables underlie most inference from experimental data but are hard to explain without complex mathematics. Animated anova tables can be used to illustrate a few of their properties and make them more memorable.

In completely randomised experiments, the idea of splitting the treatment sum of squares into sums of squares for contrasts or for different factors and interactions, is important. The diagram on the left below shows an anova table for an experiment with a single numerical factor; the checkbox animates separation of the factor sum of squares and degrees of freedom into three components.



Source	SSq	df	Msq	F	p-value
Concentration (categorical)	55.977	5	11.195	44.70	0.0000
Residual	13.525	54	0.250		
Total	69.502	59			

Split ssq for factor

Source	SSq	df	Msq	F	p-value
Linear conc	54.266	1	54.266	216.66	0.0000
Adding quadratic	[illegible]	1	[illegible]	[illegible]	[illegible]
Lack-of-fit for quadratic	0.236	3	0.079	0.31	0.8151
Residual	13.525	54	0.250		
Total	69.502	59			

Split ssq for factor

Source	SSq	df	Msq	F	p-value
Between full plots	586.7	15	39.11	0.38	0.9786
Between sub-plots	4959.8	48	103.33		
Total	5546.5	63			

Split plots Split sub-plots

Source	SSq	df	Msq	F	p-value
Seedbed preparation	198.9	3	66.30	2.05	0.1603
Plots (within seedbed)	387.8	12	32.32	1.89	0.0701
Planting method	4117.9	3	1372.64	80.18	0.0000
Interaction	[illegible]	9	[illegible]	1.46	0.1988
Residual	616.3	36	17.12		
Total	5546.5	63			

Split plots Split sub-plots

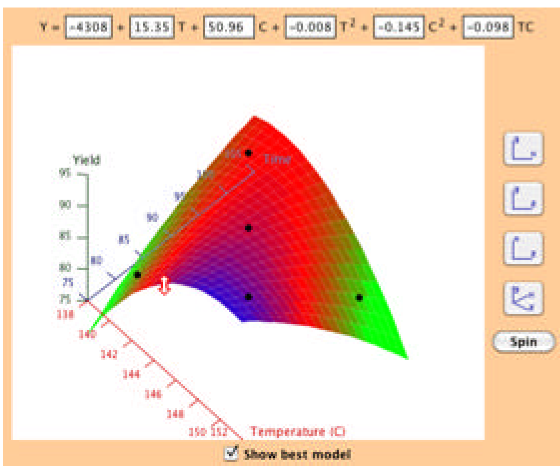
Students find analysis of variance tables for split plot designs harder to understand. The ANOVA table on the top right above initially takes no account of the factors varied at full-plot and sub-plot level. Clicking the checkbox "Split plots" animates the splitting of the full-plot factor's effect from the full-plot sum of squares, illustrating that it is really a component of the variation between full-plots. Similarly, the other checkbox animates the splitting of the sub-plot factor and interaction sums of squares from the sub-plot sum of squares, illustrating that they are determined only from variation within full plots.

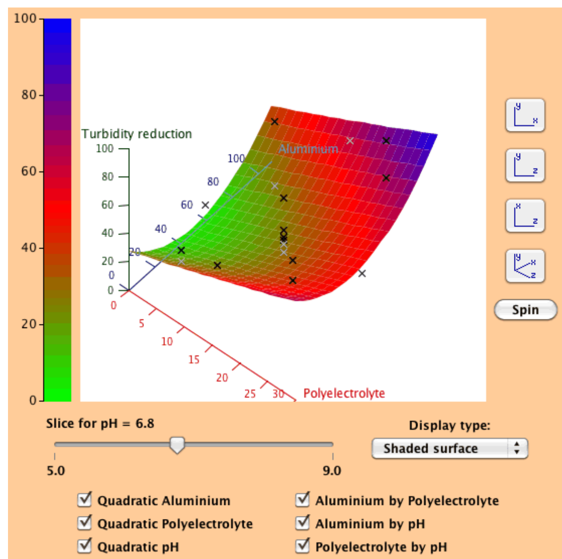
RESPONSE SURFACES AND DESIGNS FOR MIXTURES

In industrial experiments, response surface models are often used to help optimise process outputs. Three-dimensional diagrams are useful to help understand such models and designs. Although many textbooks use static diagrams of response surfaces to help explain their properties, dynamic diagrams that can be rotated are much more effective.

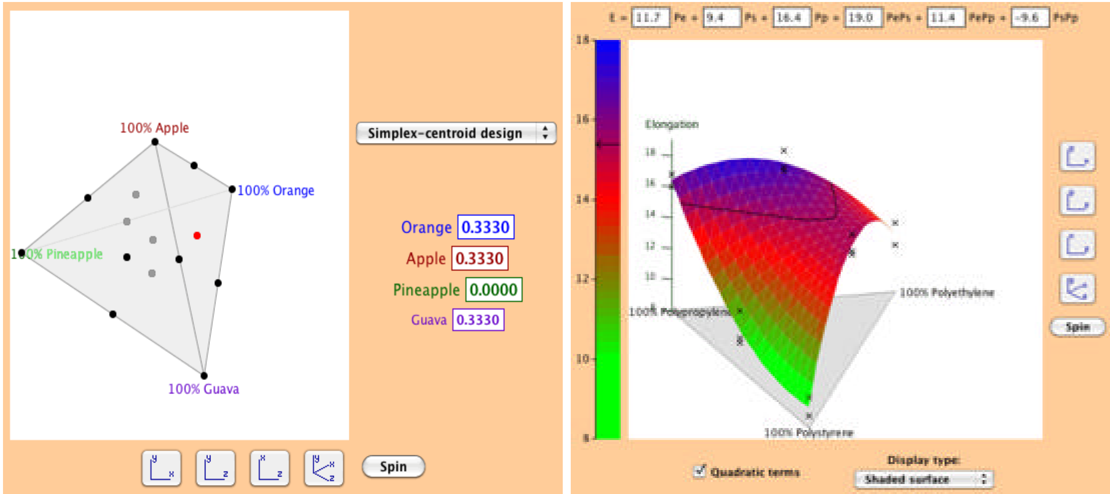
For example, the quadratic surface on the left below was fitted to data from a 2^2 factorial experiment with an added centre point. It can be used to explain to students why this design does give enough information to uniquely determine the surface; dragging the red arrow changes the shape of the surface while still passing through all five data points.

The diagram on the right shows response surface models for three factors; the six checkboxes specify the terms used in the displayed model. The surface is really four-dimensional and is represented by a slice specified by the slider for the third factor. Rotating the diagram to look down on the surface and dragging the slider helps to explain the usual way that these models are displayed in a set of 2-dimensional coloured slices (or contour plots).





Industrial experiments sometimes involve mixtures of components in which the factor levels sum to 100%. This constraint makes designs and their properties harder to explain. The diagram on the left below shows designs for mixture experiments with four factors; clicking each design point displays its mixture of levels. The diagram on the right helps to explain the similarity between response surface models for mixture data and those in the diagrams above. Dynamic rotation helps students to understand properties of the designs and models.



CONCLUSION

The saying "a picture is worth a thousand words" is particularly true for teaching statistical concepts. Dynamic diagrams are particularly effective if the diagrams require student interaction, transforming passive reading into a form of active learning (Meyers & Jones, 1993).

Dynamic and interactive diagrams, mostly in the form of Java applets, have been widely used for teaching introductory statistics but are only now being developed for more advanced statistics courses. The CAST e-books about multiple regression and experimental design provide many examples of applets that can help to explain advanced concepts.

Like applets in introductory courses, dynamic diagrams are fun for both instructor and student, but evaluation of their effectiveness in learning is difficult. Qualitative feedback from staff and students is valuable but quantitative evaluation is harder to obtain.

Other advanced statistics courses would also benefit from the use of dynamic diagrams. These include applied topics such as multivariate methods (where rotating three-dimensional diagrams would be useful) and sample surveys (where simulations could be used to compare survey designs). Dynamic diagrams also have potential within theoretical statistics courses. For example, simulations can demonstrate the independence of the mean and variance of normal samples and can compare the properties of alternative estimators; likelihood-based methods can be illustrated with dynamic graphs in which data points can be moved.

The applets described in this paper are contained in two e-books that are part of CAST (Stirling, 2010a). CAST is freely available for use under a Creative Commons Licence and can be either run directly from the CAST web site or, to improve speed and reduce the load on the CAST web server, downloaded and run from a local hard disk or web server.

REFERENCES

Bowman, A., Crawford, E., Alexander, G. & Bowman, R. W. (2007) "rpanel: Simple Interactive Controls for R functions Using the tcltk Package," *Journal of Statistical Software*, 17, 9.

Chance, B. & Rossman, A. (2006), "Using Simulation to Teach and Learn Statistics," In A. Rossman & B. Chance (eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil: International Association for Statistical Education and International Statistics Institute.

Erickson, P. (2006), "Using Simulation to Learn about Inference," In A. Rossman & B. Chance (eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil: International Association for Statistical Education and International Statistics Institute.

Finzer, W., Erickson, T., & Binker, J. (2000), *Fathom*. Key Curriculum Press.

Lane, D. M. & Peres, S. C. (2006), "Interactive Simulations in the Teaching of Statistics: Promise and Pitfalls," In A. Rossman & B. Chance (eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil: International Association for Statistical Education and International Statistics Institute.

McClelland, G. H. (1999), *Seeing Statistics*, http://www.seeingstatistics.com

Meyers, C. & Jones, T. B. (1993), *Promoting Active Learning: Strategies for the College Classroom*, San Francisco: Jossey-Bass Publishers.

Stirling, W. D. (2006), "Interactive 3-dimensional Diagrams for Teaching Multiple Regression," In A. Rossman & B. Chance (eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil: International Association for Statistical Education and International Statistics Institute.

Stirling, W. D. (2010a), *Computer Assisted Statistics Textbooks* (CAST 4.1), http://cast.massey.ac.nz

Stirling, W. D. (2010b), *CAST E-book about Agricultural and Biological Experiments*, http://cast.massey.ac.nz?book=agExper

Stirling, W. D. (2010c), *CAST E-book about Industrial Experiments*, http://cast.massey.ac.nz?book=indExper

Tierney L. (1990), *Lisp–Stat: An Object–Oriented Environment for Statistical Computing and Dynamic Graphics*. New York: John Wiley.

Contributed paper – Dineika Chandrananda and Chris Wild

A FREE STAND-ALONE SYSTEM FOR STATISTICAL DATA ANALYSIS WRITTEN IN R

CHANDRANANDA, Dineika and WILD, Chris
Department of Statistics, University of Auckland, New Zealand
c.wild@auckland.ac.nz

Abstract
The authors are building a system for data analysis designed initially for use in New Zealand high schools but now with capabilities that will also make it useful in early-level university courses. Our desire has been to provide a tool that will actively encourage the exploration of multivariate data sets and enable emphasis to be kept almost entirely on seeing what data is saying rather than learning how to drive software. We want student-time primarily to be spent thinking about the questions they want to ask of the data and what the things they see mean rather than on the complexities and grind of getting hold of them. We also want to keep the data students are working with always "in their faces" to minimise abstraction. We have done these things using a drag and drop metaphor in which the software is driven by dragging the names of variables from the top of the data spreadsheet and dropping them in a small number of appropriate places. What is delivered instantly, and with an almost nonexistent learning curve, is graphics (plots). Numerical summaries and inferential information can also be obtained but the user has explicitly to ask for them. Plots involving up to three variables require almost no system knowledge. Gradually, as students get more experienced and the questions they want to ask become ambitious, they can learn more about system enabling them to dig in deeper. Relationships involving up to six variables can be explored using only very basic plot types. A useful set of tools for re-expressing variables is provided, but these are seen as existing for more experienced users. While the system is built in R, the user should never have to interact with R. We discuss pedagogical imperatives and describe system capabilities, design choices and the reasoning that led to them.

Contributed paper (refereed) – James Baglin and Cliff Da Costa

AN EXPERIMENTAL STUDY COMPARING STRATEGIES OF LEARNING HOW TO USE STATISTICAL SOFTWARE PACKAGES IN INTRODUCTORY STATISTICS COURSES

BAGLIN, James and DA COSTA, Cliff
School of Mathematics and Geospatial Sciences, RMIT University, Melbourne
james.baglin@rmit.edu.au

Abstract
This paper reports the findings of a preliminary study assessing different strategies of training students in the use of statistical packages. The study compared active/exploratory training, which prompts students to complete tasks with minimal instruction, to guided training, which takes the student step-by-step through completing a task. A total of thirteen participants were randomly allocated to one of the two training strategies and completed a one hour training session. During the session participants rated the perceived difficulty of the training, their statistical package anxiety, and their statistical package self-efficacy. One week following the session, participants also completed an online quiz to measure analogical transfer. The limitations of the study are discussed within the context of the results. Overall, the results of this study suggest no evidence of any differences between the two conditions after one week.

INTRODUCTION

In 2005, the American Statistics Association made the use of statistical packages, such as *SPSS/PASW*, *Minitab*, and *R*, a key recommendation in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report (American Statistical Association, 2005). Statistical packages present instructors and students of introductory statistics courses with a number of benefits. These include automating difficult statistical formulae, offering instructors unique tools for demonstrating statistical concepts and giving students experience with packages that they may one day utilise in their careers. However, training students to use statistical packages effectively and efficiently is another matter.

Statistics educators are conversant with some of the learning issues students encounter in using these packages. Anecdotally, most students demonstrate poor recall of statistical package operational knowledge between week-to-week statistics computer tutorials and the poor recall is even more pronounced between semesters. As the student's proficiency with these packages is rarely formally assessed, little is known about efficient and effective ways of promoting this proficiency. Most strategies for training students in the use of statistical packages can be broadly group into two categories: Guided training versus Active/Exploratory training.

Guided training assumes students are passive participants in the learning process (Keith, Richter, & Naumann, 2010). Guided training, which may be delivered orally, in writing, or by a demonstrator, uses step-by-step comprehensive and explicit instructions and screen shots to successfully guide the learner correctly through a procedure. The student then practices these steps until they are ready to move on to more advanced procedures. There are a number of Guided training textbooks available for popular statistical packages such as SPSS (see Allen & Bennet, 2008; Francis, 2007).

Conversely, *Active/Exploratory training* views students as taking an active role in the training process and the learning achieved is under the direct control of the student (Bell & Kozlowski, 2008). Active/exploratory training gives the student minimal information to perform a task, encouraging the student to explore the statistical package themselves. Research shows that these two training strategies have different outcomes depending on the type of outcome that is measured (Keith & Frese, 2008).

When the student is training in the use of a statistical package, the student can be assessed on two types of tasks. *Analogical transfer* tasks are based on tasks that are similar to what the training covered (Keith, et al., 2010). On the other hand, *adaptive transfer* tasks are distinct from training tasks and require the student to generalise their knowledge to novel situations (Keith, et al., 2010). A recent meta-analysis of 24 studies investigating the effect of an active/exploratory training method, known as

Error Management Training, versus procedural (guided) training for common computer software (e.g. word processors, email and spreadsheets) found that active/exploratory training was slightly better compared to guided training for analogical tasks but was moderately better compared to guided training for adaptive outcomes (Keith & Frese, 2008). These findings suggest that "less guidance is more" in terms of training students in the use of regular computer software.

Very little research has focused on evaluating the merit of active/exploratory training as a strategy for learning statistical software packages. One study by Dormann and Frese (1994) investigated active/exploratory error management training methods for teaching psychology students to use the statistical package *SPSS*. While the results were in favour of the active/exploratory group, the outcome was measured immediately after the training session. This does not allow for inferences to be drawn regarding learning transfer to a following week's tutorial.

Training students in the use of statistical packages also presents instructors with other important contextual issues. Guided training might be beneficial for weaker students who need extra assistance, while active/exploratory training might be more suitable for engaging stronger students. However, it appears a recent study by Keith et al., (2010) refutes this claim when training students on the use of word processing programs. Keith et al. found that regardless of motivation and cognitive ability, students in the active/exploratory program out-performed students on guided training. However, it is quite apparent that learning to use a statistical package is quite different to learning to use a word processing program. Active/exploratory training might put students who are already in a challenging course into a situation that is all too much for them. This would lead to lower statistical package self-efficacy and higher statistical package anxiety.

From a practical perspective, if a published guided training booklet is not available, developing guided instructions is often a very time consuming process. Many statistical packages undergo regular updating, requiring a constant revision process to be put in place. Active/exploratory training, on the other hand, is far easier to implement. However, it might increase the time students need to complete a scheduled tutorial and entail extra work effort on the tutors supervising the sessions. It is apparent many questions regarding the selection of an appropriate strategy for learning statistical packages remain unanswered.

Therefore, the aim of this preliminary study was to investigate the practical, contextual, and analogical outcomes of guided versus active/exploratory strategies in training students to use a statistical package.

METHOD

Participants: The sample consisted of 15 participants who had previously completed a first year introductory statistics course. Two students were dropped from the study leaving a final sample of 13 (one student failed to complete the quiz and another was noncompliant during the session).The students came from Business (N = 4) and Applied Science programs (N = 9). There were 6 males and 7 females. All students were full-time 1st year students and two were international students. The sample mean age was 25.9 years (SD = 8.3). The mean time taken to complete the tutorial was 64 minutes (SD = 18.8). The mean follow-up time for the quiz was 8.08 days (SD = 2.1). All participants had prior knowledge of statistical software packages (E.g. *Excel* and *Minitab*), but none had experience with *SPSS*, the package used in this study. Table 1 shows the breakdown of students in the two groups used in the study.

Measures: Participants were given a tutorial booklet which contained the training strategy instructions, tutorial activities and outcome measures. This tutorial booklet gathered demographic information as well as measures of statistical package self-efficacy, statistical package anxiety, and perceived level of difficulty.

A measure of the change in statistical package Self-efficacy was adapted from three items of Finney and Schraw's (2003) Current statistics self-efficacy (CSSE) and Self-efficacy to Learn Statistics (SELS) scales. These scales have evidence of good psychometric properties (Finney & Schraw, 2003). The 3 items from the SELS and CSSE we modified to relate specifically to conducting statistical analysis using a statistical package. The participants were asked to rate their self-efficacy on a 10-point scale ranging from 1 (no confidence at all) to 10 (complete confidence) before and after the tutorial. The self-efficacy change score was calculated by subtracting the self-efficacy rating taken before the session from the self-efficacy rating given upon session completion. Scores could range from -27 to 27. High scores are indicative of self-efficacy improvement. A score of 0 indicates no change in self-efficacy.

Table 1.
Group Characteristics

	Guided	Active/Exploratory
N	6	7
Business	2	2
Applied Science	4	5
Males	4	2
International	1	1
Age: Mean (*SD*)	26.3 (10.8)	25.4 (6.4)
Tute Time Mins: Mean (*SD*)	63.3 (15.4)	65.0 (22.6)
Follow-up Days: Mean (*SD*)	8.5 (2.7)	7.7 (1.5)

Statistical Package anxiety was measured using two items adapted from Cruise, Cash and Bolton (1985) Statistics Anxiety Rating Scale (STARS). Once again, these two items were modified to relate specifically to statistical packages. The STARS has well established psychometric properties (Baloglu, 2002). Participants responded to these items on a 10-point scale ranging from 1 (no anxiety) to 10 (very strong anxiety) with scores ranging from 2 to 20. The final item in the tutorial booklet required participants to rate the perceived difficulty of the tutorial on a 10-point scale ranging from 1 (very easy) to 10 (very difficult).

Analogical transfer was measured using a 10-question online quiz to measure the participant's recall of the previous week's tutorial. The quiz had 14 possible marks. The quiz focused on assessing the participant's recall of operational knowledge of the statistical package. For those who are familiar with *SPSS*, this quiz covered differentiating variable view and data view, labelling variables, matching different descriptive outputs with the correct commands, using split file and select cases, running *t*-tests, finding *p*-values in *SPSS* output, and setting up graphs. Questions were a combination of multiple-choice and text/numerical responses. All questions contained an "I do not know" option to minimise guessing.

The tutor present at the sessions was also required to record time taken to complete the tutorial and the number of times the participant sought the tutor's assistance. This was to give an indication of the practical issues relating to the implementation of the training strategies.

Procedure: On obtaining ethics approval, participants were recruited following the completion of a one semester introductory statistics course. Involvement in the study was strictly voluntary and was completed outside regular university attendance. A raffle for a major prize was used as an incentive for students to participate. All participants were randomly allocated to a training strategy prior to attending. All participants were blinded to the exact nature of the study and the differences between training strategies.

The statistical package *SPSS* for Windows Version 17 was used in all sessions. *SPSS* was chosen as this package was not taught in the participant's previous introductory statistics course. The two tutorial conditions were designed to take approximately one hour to complete. However participants were given all the time they needed. Both conditions covered the following topics in *SPSS*: Entering data, editing datasets, descriptive statistics, data file manipulation (select cases and split file), comparing means via *t*-tests (dependent and independent) and basic graphical displays.

In the Active/Exploratory condition, participants were first given a question to answer and given a few prompts to get started (e.g. Use the Analyse - Compare Means command). The idea was to give participants minimal information and a few pointers to get started. The participant would then attempt the exercise and in the process actively explore the statistical package. Students were encouraged to seek assistance only when they were really stuck.

In contrast, the Guided training strategy had explicit explanations, screenshots and step-by-step instructions on how to do a specific analysis which the participant would work through. The participant would then be given another activity to practice the analysis that was previously explicitly explained. This condition deliberately avoided uncertainty and aimed to explain every aspect of the statistical package in the implementation of an instruction.

RESULTS

The results of this study were restricted to a descriptive analysis due to the small sample size. Descriptive statistics comparing the two conditions are show in Table 2. The median was the preferred measure of central tendency due to the susceptibility of the mean to outliers in small samples. Dot plots were also included to gain an insight into the variability of the outcomes between conditions and to also present the raw data which would otherwise be encompassed in the descriptive summaries (see Figure 1).

Table 2.
Descriptive Statistics Between Conditions on Outcome Variables

Outcome	Guided					Active/Exploratory				
	Mdn	*M*	*SD*	*Min*	*Max*	*Mdn*	*M*	*SD*	*Min*	*Max*
Tute Time (Mins)	67.5	63.3	15.4	40	80	55	65	22.6	45	105
No. Questions	3	3	2.5	0	7	3	3.3	2.6	1	8
Perceived Difficulty	2.5	3.3	2.0	2	7	5	5.3	2.1	3	8
Anxiety	7	7.7	3.8	4	14	11	9.7	3.0	6	13
Self-efficacy Change	5	7.5	8.1	-1	19	4	5.9	4.7	2	16
Quiz Score	7.5	8.2	3.3	4	12	8	7.9	1.7	5	10

Note. Mdn = Median.

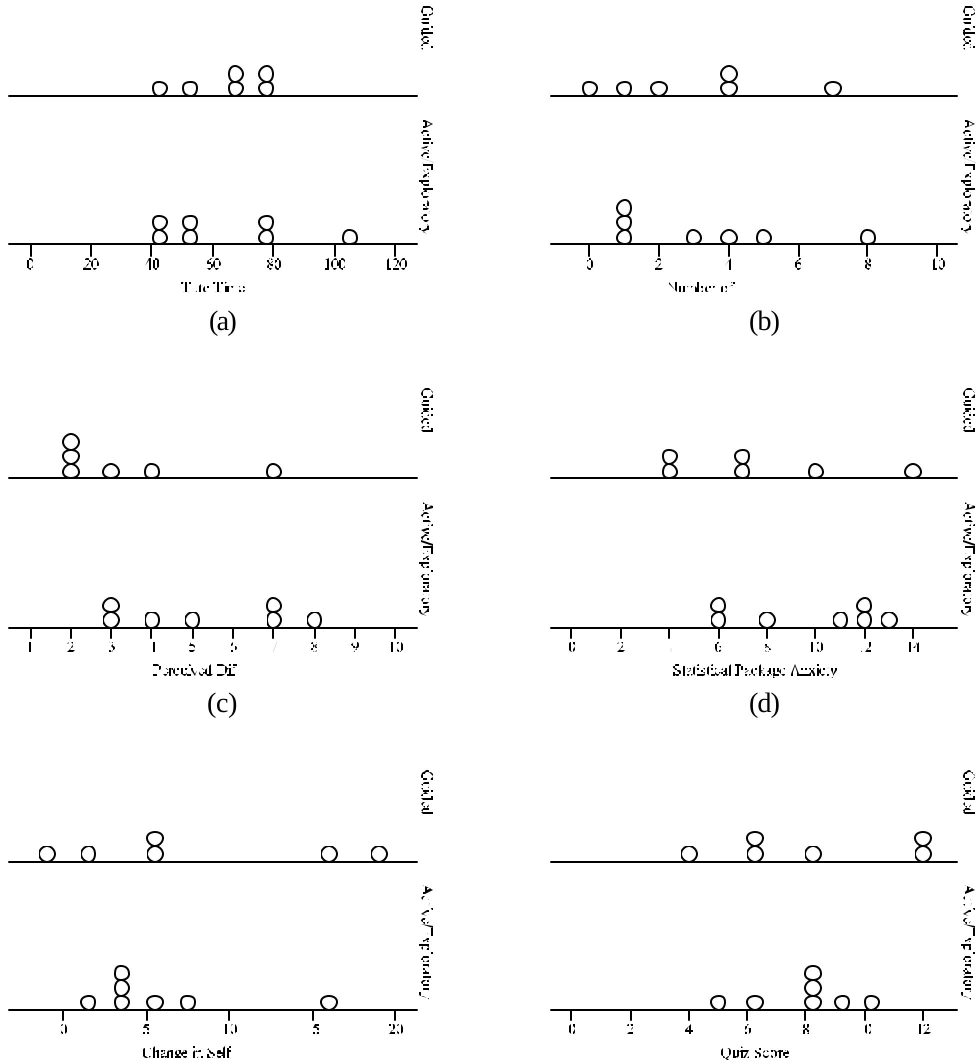
For tutorial session times, the median was lower in the active/exploratory group (*Mdn* = 67.5 mins) compared to the guided group (*Mdn* = 55 mins). Inspection of the dot plot reveals less variability in the guided condition. With the exception of an outlier in the active/exploratory group, both training strategies appeared to finishing within a similar time frame. This runs counter to intuition which would dictate that the more difficult active/exploratory condition to take longer.

The median number of questions asked was the same between conditions (*Mdn* = 3). The active/exploratory group would be expected to ask many questions given the nature of the training strategy, but the similar number of question in the guided group needs explaining. According to the researcher, the guided group would ask the tutor for validation in what they were doing. Participants in the guided group would also constantly run into problems because they had not followed the instructions properly. They would then persist in asking the tutor where they had gone wrong.

As one would expect, the median perceived difficulty of the session was higher in the active/exploratory condition (*Mdn* = 5) compared to the guided condition (*Mdn* = 2.5). This provides evidence of the validity of the difference between the nature of the two training strategies. This probably led to a higher median statistical package anxiety rating in the active/exploratory condition (*Mdn* = 11) versus the guided condition (*Mdn* = 7).

In terms of statistical package self-efficacy change, both groups were comparable with a median of 5 and 4 for the guided and active/exploratory groups respectively. However, examination of the dot plot showed a large degree of variability in the guided condition, whereas the change score for the active/exploratory group, with the exception of one value, was clustered close to the median. One interpretation of this result is that the guided condition created very mixed perceptions of self-efficacy change, whereas the perceived change in the active/exploratory condition was more uniform.

The performance on the online quiz assessing analogical transfer one-week after the tutorial session between training strategies showed that the active/exploratory condition scored marginally higher (*Mdn* = 8) than the guided condition (*Mdn* = 7.5). Once again, the active/exploratory condition was associated with a lower degree of variability as demonstrated in the dot plots. Overall, there appears to be little discernable difference on analogical scores taken from the quiz between the two training strategies.



(a) (b) (c) (d) (e) (f)

Figure 1: Dot plots showing the distribution of each outcome variable between conditions.

DISCUSSION

The aim of this preliminary study was to compare active/exploratory and guided training strategies for the training of students in statistical packages. The outcomes measured included both contextual and performance indicators. As expected, the active/exploratory training strategy was associated with higher perceived difficulty and higher statistical package anxiety. This was expected as the idea of the active/exploratory training strategy was to engage students by giving them minimal information. Also, as anxiety was only measured after the session, the discrepancy might be explained by individual differences even though random allocation was used. This type of question could only be addressed with the use of a larger sample where the probability of pre-existing group difference in randomly allocated designs reduces as the sample size increases. The trade-off of higher anxiety is partially compensated by no perceivable difference between training strategies on statistical package self-efficacy change. Both training strategies seemed to have a similar positive effect, but this effect seemed vastly more variable in the guided training strategy.

There were also a few unexpected results. There were no noticeable differences between the training strategies on the number of questions asked during the tutorial sessions. From a practical perspective, this challenges that notion that using an active/exploratory strategy would increase the demand on tutors supervising sessions. The results for session time also suggest that an active/exploratory strategy would not necessarily increase the required time taken to complete a set tutorial.

The major outcome of interest in this study was the difference in performance between the two training strategies on an analogical transfer quiz. The active/exploratory group did score higher, but the large degree of variability in the guided group made any definitive conclusions difficult. Overall, the results from the quiz suggest no clear difference between the two training strategies. These results were in agreement with what would be excepted in a low powered study (Keith & Frese, 2008).

The major limitation of this preliminary study was the sample size. Every possible measure was taken to maximise participation, but convincing students to attend a statistical tutorial outside of regular course activities was difficult indeed. Future research should incorporate the evaluation of the training methods within the framework of a regular course. Future research should also include measures of adaptive transfer. Adaptive transfer is a more desirable outcome to achieve given that students will need to apply what they learned in their statistical courses to other academic and professional tasks.

In conclusion, this preliminary study suggests that, at the very least, training students in statistical packages using active/exploratory training strategies is comparable to guided training in terms of analogical performance for week-to-week transfer. More importantly, it is hoped that this preliminary study will stimulate large sample comprehensive studies leading to more definite conclusions in order to ensure that students are getting the best statistical package training available.

REFERENCES

Allen, P., & Bennet, K. (2008). *SPSS for the health and behavioural sciences*. South Melbourne, VIC: Thomson.

American Statistical Association. (2005). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) project*: Author.

Baloglu, M. (2002). Psychometric properties of the statistics anxiety rating scale. *Psychological Reports, 90*, 315-325.

Bell, B. S., & Kozlowski, S. W. J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology, 93*, 296-316.

Cruise, R. J., Cash, R. W., & Bolton, D. L. (1985). *Development and validation of an instrument to measure statistical anxiety*, Proceedings of the American Statistical Association, Las Vagas, Nevada.

Dormann, T., & Frese, M. (1994). Error training: Replication and the function of exploratory behavior. *International Journal of Human-Computer Interaction, 6*, 365-372.

Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology, 28*, 161-186.

Francis, G. (2007). *Introduction to SPSS for Windows: v15.0 and 14.0 with notes for Studentware*. Frenchs Forest, NSW: Pearson Education Australia.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology, 93*, 59-69.

Keith, N., Richter, T., & Naumann, L. (2010). Active/exploratory training promotes transfer even in learners with low motivation and cognitive ability. *Applied Psychology: An International Review, 59*, 97-123.

Contributed paper (refereed) – Martin Gellerstedt and Lars Svensson

WWW MEANS WIN WIN WIN IN EDUCATION – SOME EXPERIENCES FROM ONLINE COURSES IN APPLIED STATISTICS

GELLERSTEDT, Martin and SVENSSON, Lars
Dept. of Economics and IT,
University West Sweden,
Sweden
martin.gellerstedt@hv.se

Abstract
This paper reports on the experiences from online courses in applied statistics. The courses were designed with the ambition of making studies in statistics, fun, interesting, useful, not that difficult and directly supported the possibility to combine studies and work. When designing the courses we considered three dimensions: "pedagogies","community" and "structure". Experiences after giving a first-year course three times show that the online course succeeds in attracting new students since 90% of the participants would not be able to follow an on-campus course and 62% worked full time. The pedagogies were highly appreciated because focusing on the interpretation of results and using computer analyses really changed the prejudices about statistics. Structure and prompt feedback was experienced as important factors. It was possible to combine online studies with employment, and the student completion rate was (84%, 55% and 61%), with a potential for further improvements.

INTRODUCTION

The last two decades have been an expansive era for online courses, and the numerous accounts of success stories are encouraging for online educators. Still, with new groups of students being attracted to online courses, new challenges are likely to emerge. These challenges could serve as an asset for improving the quality of online education (Trentin 2002), but there are still major obstacles to overcome, and we can still note that there is a gap between promises and reality of online education (Wilner & Lee 2002). The literature on Educational technologies highlights some of these concerns in online settings, for instance social presence (Gunawardena 1995), sense of community (Svensson 2002), student frustration with technical problems, ambiguous instructions, and a lack of prompt feedback (Hara & Kling 1999). A problem of significant importance for educators is student dropout rates, which may be partly explained by factors such as sense of community and student interaction (Rovai 2002, Rovai 2003). At present, online education seems to be moving from pilot testing and small-scale initiatives into a more mature phase where large-scale programs are rigorously evaluated (Svensson & Gellerstedt 2009), and there are a number of suggested guidelines and principles, which summarize lessons learned that are possible to use in design and evaluation of an online course (e.g. Graham et al 2001, Buzzetto-More and Kaye Pinhey 2006, Herrington et al. 2001, Wood and George 2003).

The University West Sweden was a pioneer institution, designing the first online B.Sc. program in Scandinavia, and has now been offering online courses in a wide range of educational disciplines for 15 years. Enthusiastic colleagues have taken part in practical work and research over the years. Also a learning management system (DisCo) has been developed in-house and is now standard at the University for both campus and distance courses (Svensson et al. 2010). The courses in Statistics have changed and adapted new technologies and the online era has demanded new teaching strategies and course material. Furthermore, the pedagogical strategy has gradually increased focus on time on "mathematical masochism". We believe that this approach enhances skills useful for working practical applications, interpretation of results and using computers for analyses rather than spending life and that it increases the conceptual understanding of statistics. We also believe that the attitudes towards statistics are changed to the positive (see Wore & Chastain 1989 for a similar conclusion). In 2008 we decided to consolidate our experiences from the last two decades and try to design modern attractive courses available on campus and online as well. The online courses should be possible to combine with work. In this paper, we will describe the design strategy and present experiences after giving a first-year online course in descriptive statistics three times.

DESIGNING THE SPSS ACADEMY

According to our experiences there seems to be a rather common prejudice among students that statistics is boring, difficult and not that useful. Since we believe in the opposite we wanted to find a way of making studies of applied statistics fun, interesting, useful and not that difficult. And, as mentioned above, we wanted to offer the possibility to study online courses while employed. Based on our own and collegial experiences the following three dimensions guided the design process: (i) *Pedagogies* -learning activities supporting the pedagogical aim. (ii) *Community* - activities for creating continuity and a sense of community. (iii) Structure: instructions and technical solutions.

The most important pedagogies were:

- *Authentic and engaging cases*, preferably based on news of current affairs. This should stimulate curiosity and a desire to find out the truth by exploring data.
- *Assignments that can be solved on your own.* The assignments should be progressive over the course and successively increase self-directed learning. Assignments on campus are solved in groups, while the assignments online are solved individually.
- *Continuous computer assisted teaching*. Learning processes driven by assignments and the use of lectures, books and whitepapers as support.
- *Automatic SPSS skills*. Detailed SPSS instructions that are reduced once the students' skills increase.
- *Understanding & interpretation*. Focus on understanding statistical concepts and interpretation of results. A minimum of mathematical formulas.
- *A Collaborative project* including data collection, coding, analyzing and reporting.
- *Engaging and informal.* Using a relaxed and informal teaching style – "show that statistics is fun and interesting". Teachers must have experiences from solving real life problems.
- *Professional productions.* Pre-recorded lectures designed especially for online purposes and not just copies of classroom lectures.
- *Learning and examination.* The examinations should be a part of the learning process rather than a test of achieved knowledge.

The most important strategies for Community were:

- *Vital course home page and continuous activity*. High degree of activities, e.g. discussions and even distribution of material, news, push-mails, etc.
- *Participation*: Frequent visits from teachers, with comments, suggestions, answers to questions, links to news related to statistics, etc., preferably on a daily basis. Students are encouraged to cooperate and to continue ongoing discussions. A collaborative survey is one part of the examination.
- *Prompt feedback*. Visit and leave comments on discussion boards daily. Respond to e-mail as quickly as possible. Results on assignments/examinations during the course are presented within one working day. Specific times for synchronous discussions via e-mail are also offered.

The most important strategies for structure were:

- *Modules.* Each course is divided into 3-5 modules. Each module is supported by study guidelines, assignments and an examination. Students are provided prompt feedback after the examination of each module. Modules are used for keeping an even pace in the course and for giving feedback and a possibility to diagnose knowledge at examination "check points".
- *Study guidelines:* Guides that include the learning objectives, key concepts, what to do, links to pre-recorded lectures, data sets, computer assignments, and templates for reports.
- *Deadlines.* We have clear deadlines and no exceptions are accepted. If a student misses a deadline, a new examination assignment is offered after the course ends.
- *Accurate and reliable.* Instructions must be detailed, error free, up to date, and possible to follow on your own.

During the development of the SPSS Academy we contacted the Swedish office for SPSS Inc. and started cooperation. This cooperation means that we could exchange experiences, data, etc. Furthermore, students who complete a course receive a certificate signed by SPSS and the University. The SPSS academy consists of four 5 weeklong courses, which together constitute a minor in applied statistics. All the course names start with "Applied statistics" followed by the following subtitles: "To collect and summarize data", "To draw conclusions from data", "Predictive models" and "Independent analytical project". We offer all courses both on campus and online as well. All courses are based on the strategies discussed above, i.e. learning by solving assignments using SPSS. The first course "To collect and summarize data" contains descriptive statistics, basic survey methodology, simple linear regression and a short introduction to inference. The course examination consists of two computer assignments and a collaborative survey project. In the survey projects the students: choose a theme - teachers construct a draft questionnaire of really bad quality, students suggest improvements - teachers produce a refined questionnaire, students collect 10 questionnaires each and enter data in SPSS – teachers pool data, and finally students analyze and write report.

EVALUATION OF THE FIRST ONLINE COURSE IN THE SPSS ACADEMY

The evaluation of the course: "To collect and summarize data", given 3 times (Fall 2009 to Spring 2010) is based on data from several sources presented in Table 1. In this paper we primarily focus on the results from the Course Conference and the Study Context Survey. However, both the administration data and the results from the student satisfaction survey are indirectly included since these data are used in the Course Conference.

Table 1.
Evaluation sources

Data Source	**Description**
Administration Data	Data concerning student recruitment, dropouts, proportion of student completion, and student grades.
Student Satisfaction	At the end of each course, students fill in a standardized questionnaire with questions about content, learning activities, examination, etc.
Course Conference Notes	Each course is also evaluated and discussed during a seminar (4 h) where the course content, learning activities, examples of examinations, and results from student evaluation are presented and discussed. Course teachers, subject matter experts, e-learning experts and researchers participate in the seminar.
Study Context Survey.	During the Spring semester 2010 we sent out a questionnaire with questions about: age, city of residence, working situation (not working, part-time or full time) and study situation (other courses?). The response rate was 71% (33 out of 46 students).

EXPERIENCES FROM THE EVALUATION OF THE SPSS ACADEMY – MAIN FINDINGS:

In this section we briefly summarize the most central aspects that were discussed and analyzed during the course conference. Focus is set on positive change and challenges that can be related to the implemented design principles regarding *pedagogies*, *community* and *structure.* Furthermore, we present data related to the *study context survey*.

We confirm that the *pedagogies* really change attitudes and prejudices. Student evaluations frequently contain comments indicating that studying statistics was not as expected, e.g. *"not as difficult as expected and it was even fun!"*, *"Surprisingly interesting and important"*, *"A true relief for me who is afraid of formulas"*. There is however some frustration over the fact that the text-book contains several formulas, and is not aligned with the course pedagogy. There is a need for a more suitable book. We have also noticed that our informal and relaxed profile is adopted also by students and we believe that an informal, yet respectful, attitude makes the discussion more vital and open for everyone. Pre-recorded lectures and the quality of the clips are appreciated. Finally, we believe that authentic assignments engage the students, especially the final collaborative assignment – a questionnaire survey, really engaged the students. However, the first time the course was given, we noticed that the skills in writing a report varied a lot. Several students had to spend time on finding out how to write a report instead of focusing on producing adequate statistics. In subsequent courses we gave students a report skeleton, with titles and hints for suitable content.

Regarding *Community and structure*, the most frequent comment in student evaluations regards the structure and feedback, which confirms the importance of these factors. We also believe that it is essential to keep the course home page vital. When the course was given online for the first time involved teachers were enthusiastic, were frequently present in discussions, and encouraged discussions, all of which stimulated vitality. On the second occasion the enthusiasm was somewhat decreased, and on the third occasion it was somewhat strengthened again. We believe that this partly explains the variation in student completion (84%, 55%, 61% respectively).

The study context survey shows that more than 9 out of 10 participants would not have been able to follow the course on campus (based on geographical distance (at least a two hour's drive from Campus) or working full-time). Eighty-two percent are working at least part time, 62% are working full time. Thus, the online course attracts students that would not be able to follow a campus course. The proportion of immediate dropouts (participants who never actively participate in the course) is around 20%. At the university in general, this figure is estimated to be around 20-25% for online courses. One possible explanation is that these dropouts are students who have underestimated the available time and motivation. The proportion of completion (calculated as the proportion that passes the course among active students, i.e. immediate dropouts are excluded) was 84% (53/63), 55% (16/29) and 61% (28/46) respectively.

Among the responders of the questionnaire, 68% passed the course. Student completion by working status is shown in Table 2.

Table 2.
Student completion by working status

	Fail	**Pass**	Total
Non working	1 (16,7%)	5 (83,3%)	6
Part time	2 (28,6%)	5 (71,4%)	7
Full time	8 (38,1%)	13 (61,9%)	21
Total	11 (32,4%)	23 (67,6%)	34

If a student who passed the course is categorized after grade (passed or passed with honors) we found that among non working and part time working students the proportion: (pass with honors/pass with or without honors) is one out of five (20%), while the same proportion among full working students was nearly one out of two (46%, 6/13). Another interesting finding is that the proportion that also studies another course in parallel was 100% (6/6) among the non workers, 71% (5/7) among part time workers and 43% (9/21) among the full working participants.

DISCUSSION

Offering online courses in applied statistics helps us to reach new students. As teachers, we can also note more reality based questions and discussions, which is stimulating both for us and for other students. In fact, some questions have been nearly of a "consultancy nature". There was a trend in the data indicating that the proportion of completion decreases gradually among students who work part- or fulltime. We know that among participants who are working full time, the importance of higher education credits is not that important as it is for a student who is putting together a degree. We have received comments like *"It was interesting and I have learned a lot, but I don't need the credits so I don't want to spend time on examination tasks"*. A majority of students among "non workers" and "part time workers" actually studied other courses in parallel, indicating that they primarily are "students" and not "workers", which makes the credits more valuable. During discussions with dropouts, increased working load or other exogenous factors, e.g. changing job, moving or sickness, are given as reasons for the drop-out. In short, getting the credits is not a primary concern for all full-working participants and if available time for some reason decreases, delivering examination tasks is not prioritized. Hence, we expect completion rates to be lower among full working participants than compared to non-working or part-time workers.

We believe that the online courses are suitable for combining with work, even full time. The best examples of examination reports are actually produced by students working full-time. We find the completion rates satisfactory, but we can see potential improvements. Firstly, we experience that our engagement for keeping the course home page vital and for keeping up discussions, sending push-mails, etc is important. We believe that it affects the sense of community and we can see that the completion rate goes down when vitality goes down. Furthermore we would like to find ways that stimulate working people to finalize examination even if credits do not matter. We have tried to construct examination forms as a natural part of the learning process, but maybe this could be refined further, i.e. to get the learning effect to outweigh the effort needed. Another possibility would be to involve problems useful for the workplace, i.e. construct examinations that can be performed at work, or using participants' real work problems in the course. We experienced that in several cases examination reports from participants with working experiences were of truly good quality, the proportion "pass with honors/pass with or without honors" was more than twice as high among full working students than among others. Thus, engaging experienced working participants could be an asset. Our pedagogical strategies seem to be working, and student evaluations are in general very positive. Using a computer for analyses, focusing on interpretation, and finding interesting patterns in data stimulate curiosity and engage students, no matter if it is on campus or online. According to student evaluations this approach seems to be a true relief for students with math-anxiety and in many cases the attitudes, or rather prejudices, about statistics are changed radically. Furthermore, we confirm that the dimensions community and structure are of great importance for students. We have especially noted the high frequency of comments regarding structure and feedback.

In summary, focusing on the interpretation of the results and using computer assignments, is appreciated among students and could even lead to changed attitudes about statistics. Structure and prompt feedback is highly appreciated. We believe that our course design makes it possible to combine work and studies and that the completion rates are acceptable. However, we believe that the completion rates could be increased further, especially among full-working students. It would be interesting with further research using more experimental designs for studying in order to explain factors for completion.

REFERENCES

Buzzetto-Moren N., & Pinhey, K. (2006). Guidelines and Standards for the Development of Fully Online Learning Objects. Interdisciplinary Journal of Knowledge and Learning Objects vol 2, p95-104

Graham, C., Cagiltay, K., Byung-Ro, L., Craner, J. ,& Duffy, T. (2001). Seven principles of effective teaching:A practical lens for evaluating online courses. The Technology Source, March/April 2001. Available at http://technologysource.org/article/seven_principles_of_effective_teaching/

Gunawardena, C. N. (1995). Social Presence Theory and Implications for Interaction and Collaborative Learning in Computer Conferences. International Journal of Telecommunications, 1 (2/3), pp.147-166

Hara, N., & Kling, R. (1999). Students' Frustration with a web based distance education course. First Monday, Peer-reviewed Journal on the Internet, http://firstmonday.org

Herrington, A., Herrington, J., Oliver, R., Stoney, S., & Willis, J. (2001). Quality guidelines for online courses: The development of an instrument to audit online units. In (G. Kennedy, M. Keppell, C. McNaught & T. Petrovic (Eds.) Meeting at the crossroads: Proceedings of ASCILITE 2001, (pp 263-270). Melbourne: The University of Melbourne.

Rovai, A. P. (2002). Building a sense of Community at a Distance. International Review of Research In Open and Distance Learning, ISSN: 1492-3831

Rovai, A. P. (2003). In search of higher persistence rates in distance education online programs, The Internet and Higher Education, Vol 6, Issue 1, pp 1-16

Svensson, L. (2002). Communities of Distance Education, Doctoral thesis, Göteborg Unversity Series in Informatics, No 25, Göteborg, Sweden

Svensson, L., Gellerstedt, M. (2009). A Reality Check for Work-Integrated e-Learning, in Proocedings of e-Learn Asia Conference, Seoul, South Korea

Svensson, L., Nilsson, S. (2010). Learning Models in Online Education: On the transformation from tutors to producers. In Proceedings of Global Learn Conference, Penang Island, Malaysia

Trentin, G. (2002). From Distance Education to Virtual Communities of Practice: The Wide Range of Possibilities for Using the Internet in Continuous Education and Training, *International Journal on E-Learning*, Jan-Mar 2002

Wilner, A., & Lee, J. (2002). The promise and the reality of distance education. Update. Washington, D.C.: National Education Association, Office of Higher Education, 8, 3.

Wood D., & George R. (2003). Quality standards in online teaching and learning: a tool for authors and developers. In G.Crisp, D.Thiele, I.Scholten, S.Barker and J.Baron (Eds), Interact, Integrate, Impact: Proceedings of the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education. Adelaide, 7-10 December 2003.

Wore, M., & Chastain J. (1989). Computer-Assisted Statistical Analysis: A Teaching Innovation? Teaching of Psychology, vol 16, No. 4, p222-227.

Contributed paper – Carey Biggs

GENSTAT FOR TEACHING

BIGGS, Carey
VSN International, Hemel Hemstead, UK
Carey.Biggs@vsni.co.uk

Abstract
VSNi believe that statistics and data analysis is central to many disciplines trying to bring meaning and deeper understanding to a student's research and work. A solid grounding in statistical practice is vital to many of today's researchers, as a part of this, students should be taught with tools they will encounter in later life, and tools that add value to their learning.

GenStat for Teaching is a set of new software tools, based on GenStat 13th edition. Currently a Schools edition is being developed for high school students of mathematics and statistics, and a university edition for undergraduates. Both systems retain the core statistical routines and strengths of GenStat, but have been specifically tailored to suit the needs of different level of students. Our developers have worked with high school teachers and university lecturers to provide a simpler menu based system highlighting the more usual statistics techniques that students come across; the more advanced university version unlocks slightly more detailed and complex statistics in line with the types of analysis techniques studied by undergraduates. At VSNi we are keen to support students and teachers alike, which is why GenStat for Teaching is free, at both schools and university level.

VSNi is a prime supplier of data analysis software for the biological and life sciences markets worldwide. We were formed in 2000 as a spin off from Rothamsted Research (RRES) and the Numerical Algorithms Group (NAG). We are backed by UK government through RRES being the largest land-based research institute in the UK and arguably the original home of statistics. As a result we are uniquely placed to provide statistical software to the world's educators and students. Our ethos is collaboration and partnership, ingrained within our psychology from our government links.

Topic: Introductory and transition

Contributed paper – Maureen Townley-Jones

DIAGNOSTIC TESTING FOR STATISTICS COURSES

TOWNLEY-JONES, Maureen
School of Mathematical and Physical Sciences, University of Newcastle, Australia
Maureen.Townley-Jones@newcastle.edu.au

Abstract
It is generally recognised that a part of any cohort of students entering tertiary education are underprepared in mathematical skills which are essential for understanding statistics at a tertiary level. As well as being underprepared, students are entering tertiary education with a misconception that mathematical and/or statistical skills are not required or misplaced perceptions of statistics due to learning experience in earlier school years prior to tertiary study. As tertiary educators, we recognise the importance of students having these skills to successfully navigate their way through university study. But students do not understand why basic mathematical and statistical skills are important or related to their discipline of study.

The consequences of allowing students to enrol in tertiary education with varying mathematical and statistical skills have been the basis for many Australian universities to implement various strategies that will allow students and staff to identify the weakness and strengths in students quantitative skills and direct students to other remedial actions, thereby to improve student retention and give students skills that are important graduate attributes.

This paper will discuss the journey that was initiated by the recognition of a need to do diagnostic testing for statistics and mathematics courses. The challenges of meeting different university student cohorts are discussed as well as strategies and the results of showing performance in first year subjects requiring mathematical skills compared to diagnostic test results.

Contributed paper – Isnander Slamet

THE IMPROVEMENT OF LEARNING PROCESS OF MATHEMATICAL STATISTICS I SUBJECT THROUGH STUDENT TEAMS ACHIEVEMENT DIVISIONS (STAD) METHOD USING ENGLISH AS MEDIUM LANGUAGE

SLAMET, Isnandar
Department of Mathematics, Sebelas Maret University,
Surakarta, Indonesia
isnandar06@yahoo.com

Abstract
Learning and teaching process of Mathematical Statistics I subject in the Department of Mathematics, Sebelas Maret University, Surakarta, Indonesia used to follow teacher-centred paradigm. In Indonesia English is learnt as a foreign language and compulsory subject in all levels of education. However, prior studies report that tertiary students faced an English language difficulty as a significant barrier to understanding the literatures. The study investigates the achievement of student-centred learning process of this subject during the academic year 2007/2008 through STAD method using English as passive medium language. The result shows percentage of students passing this course is 100%. The study confirms the importance of student-centred learning process and the problem of using English in learning process.

BACKGROUND OF SUBJECT

Mathematical Statistics I is a compulsory subject for students from all areas of concentration, namely statistics, applied mathematics, and computers enrol in Department of Mathematics, Sebelas Maret University. This subject is 3 credits worth. To take this subject, students must meet the prerequisite i.e. have to pass a Probability subject. In addition, this subject is a prerequisite for the subject Mathematical Statistics II.

The Mathematical Statistics I is given as an attempt to prepare students for having the ability to explain phenomena related to the uncertainty condition through a mathematical approach to the form of probability distribution models. In addition, this subject supports the core competencies of graduates majoring in mathematics, which will be able to think logically and systematically i.e. having proper knowledge and understanding of statistics and be able to continue their studies into higher education i.e. having intellectual skills.

The average percentage of achievement for this subject in the 3 academic years 2004/2005, 2005/2006 and 2006/2007 is shown in Table 1.

Table 1. The Average Percentage of the Grade for 3 Academic Years: 2004/2005-2006/2007

Grade	The Average Percentage
A	16,30
B	23,00
C	48,30
D	9,00
E	3,30

Considering the above circumstances, we will develop a method of student-centred learning process using English as a passive medium language. This improvement plan is expected to raise student's achievement and English proficiency.

COURSE COMPETENCY

Standard competency of this subject is that students are able to construct and apply some inferential statistical concepts. The basic competencies are:

i. derive the limit of a distribution
ii. find a statistic and derive its sampling distribution
iii. determine a point estimation
iv. determine sufficient and complete statistics
v. construct an interval estimation
vi. construct the concept of hypothesis testing

ASPECT/COMPONENT DEVELOPED

Instructional method proposed here is active learning using English as passive medium language. The method chosen is Student Teams Achievement Divisions (STAD). According to Sibermann (1996) and Sri Anitah (2006), STAD method is one of the learning methods that can activate students. This learning method is a learning theory of constructivism based on cognitive learning theory.

In STAD, students are placed in learning groups of four or five people who were a mixture according to performance level, gender and ethnicity. The lecturer presents materials, then students work within their groups to ensure that all group members have mastered the material. Next all students are given a quiz about the material and at the time of taking the quiz, students are not allowed to help each other. The score of each student is compared with an average of their previous score, and points are awarded based on how well students meet or exceed previous performance. The scores of each member of this group are added up to get the score of the group, and groups that reach a certain criteria will be given a certificate or award.

In addition in developing this teaching method, the learning process is run using English as a passive medium language. Students are encouraged to use English in their academic work. Their ability of reading English literatures as well as their content knowledge will be developed. Likewise, their capability on oral presentations, speaking and writing will increase. Media used in this process learning are computers, LCD projector, OHP and whiteboard.

RATIONALE

In previous years, the method used in teaching this subject was the delivering method. In this method, lecturers explain materials, give examples and problems. Students take notes and solve the given problem. In other words, the learning process follows the paradigm of teacher centred.

According to Gulo (2002) and Buzan (2005), there are weaknesses of the delivering method by the teacher. These are 1). The learning process tends to be teacher centred, 2). Learners tend to be listeners and note takers, 3). The delivering process depends on the speed of speech and dialect language used by teachers or lecturers.

This can cause unfavorable impact for students. Students will rely on notes from lecturers, are less able to learn independently and are less able to develop course material. The less participation in classrooms and lacking of the critical skills of the students were reported by Chalmers and Volet (1997) and Biggs (2001) as implications of this learning process. Dardjowidjojo (2001) stated that conformity is still the most dominant rule in education systems in Indonesia rather than individuality which imply students become passive in the class.

It is our concern on developing learning strategies that encourage students to be active in the teaching and learning process. The method used in this process is STAD.

Although in Indonesia, English is taught as a foreign language for all levels of education, it is seldom used in classroom. The teacher tends to use *Bahasa Indonesia* in teaching English (Mustafa, 2001) in Novera (2004). Introducing English as a passive medium language will help the students increase their ability in academic work.

The technical implementation of this process is as follows:

i. At the beginning, motivation related to the subject as well as the method used in the learning process is given. At the same time, a contract of the learning process is conducted.
ii. All materials are presented in order of basic competence. The materials presented via slides or transparencies are written in English. The lecturer delivers the material using English as the medium language. When students face difficulties in understanding the contents because of the oral presentation, then the lecturer uses Indonesian and English.
iii. Students are divided into several teams or heterogeneous group consisting of 4 or 5 students. This group will help its member to remember and understand the material given so that they will be able to do all tasks later.
iv. After one or two presentation and team work, a quiz is given in English and students should working individually. The result is to determine each individual's understanding of the material.
v. Individual scores and group scores are computed. Group award is given based on the achievement of the group.
vi. Assignments, a mid semester exam and a semester final exam are given in English.

TARGET

Students can be more active in the process of learning and teaching in a group. They have many opportunities to express opinions and to argue, so that they will understand the material better. Thus the student-centred learning will be formed and can enhance students' learning achievement.

i. With the learning process using English as the passive medium language, it will encourage students to read and study the modules and the text book. Therefore, it can improve students' ability in English language
ii. Model documents of learning Mathematical Statistics I using STAD method can be produced in the form of softcopy i.e. on CD and hardcopy including syllabi, learning design, module theory. Module theory is not prepared in English since the reference book is already in English.
iii. The STAD method is applied in the course of Mathematical Statistics I.

SUCCESS INDICATORS OF IMPLEMENTATION

i. All students can complete the tasks properly.
ii. All students can work on mid and final of semester exams.
iii. Students are more active and able to work in team.
iv. The student achievement increases.
v. Students have the ability to read and understand the resource (modules and English text book) and communicate in English. This was reflected in the students' ability to understand and translate quizzes, assignments, mid semester and final semester exam as well as presentations which was given in English.
vi. Model documents of learning Mathematical Statistics I using STAD method can be produced in the form of softcopy i.e. on CD and hardcopy including syllabi, learning design, and module theory.

EVALUATION TOOLS

i. Assignments. There are 3 assignments that must be done by students in groups. The task should be presented and written in English. Presentations can be given in Indonesian or English.
ii. Quizzes. There are 7 quizzes in the form of essays given after every 1 or 2 times of lecturer’s presentation and 1or 2 working group.
iii. Mid semester exam. This is independent work. The problems are given in essay form and taken from the material covered from week 1 to week 8.
iv. Final semester exam. The problems are taken from the material covered from week 9 to week 16.

INDICATORS OF EVALUATION

In accordance to the basic competencies, indicators can be explained as follows:

Basic competency 1: derive limit distribution

Indicators:

- state the sequence of a random variable
- mention the central limit theorem
- derive the binomial distribution and say what it approaches
- derive the asymptotic distribution
- mention the properties of its stochastic convergence

Basic competency 2: find statistics and derive the sampling distribution

Indicators:

- state the definition of the statistic
- derive the sampling distribution
- derive the distribution of t, F and beta
- derive the approximation for large-sized samples

Basic competence 3: determine the point estimation

Indicators:

- find a point estimation
- use criteria to evaluate the estimator
- mention the large sample properties
- calculate the Bayes estimator and minimax

Basic competency 4: determining sufficient and complete statistic

Indicators:

- explain the sufficient statistic and its properties
- explain the complete statistic and its properties

Basic competency 5: construct interval estimation using general and PIVOT method

Indicators:

- determine the confidence interval using PIVOT
- determine the confidence interval for the problem of pairs of independent random samples
- determine the confidence interval for the problem of pairs of dependent random samples
- determine the confidence interval using an approximation method
- determine the confidence interval using a general method

Basic competency 6: explain the concept of hypothesis testing

Indicators:

- carry out the procedure of hypothesis testing
- construct the binomial test
- construct the Poisson test

TECHNIQUE OF ANALYSIS

The score of assignments, quizzes, mid semester exam and final semester exam are in the scale 100. Student final score is calculated using the formula

Score = 20% (Assignments) + 25% (Quizzes) +27.5% (Mid semester exam)
+22.75% (Final semester exam)

From the final score obtained, then it is converted into the grade by referring to the Rector's Decree No. 177/PT 40.H/I/92 October 7, 1992.

DISCUSSION

From the classroom observation and questionnaire given, the result of the study can be reported. At the beginning, a contract related to learning process was made. Motivation with regard to the aim and benefit of studying this subject was given.

Students are divided into several teams / heterogeneous group consisting of 4-5 people to assist all members in remembering and understanding the material that has been given, so that later they will be able to do the quizzes, mid-terms and final exams and presentations well.

All materials were written in English in order to improve the students basic competence in using English. Most the time the lecturer has to use Bahasa Indonesia to help the students understand the contents.

At the time of preparation of the group tasks, there were some groups who face difficulty in understanding the textbook. To overcome this, the students were encouraged to consult their lecturer before making or doing group assignments or presentations.

All students reported that team work had helped them to understand the content better and had motivated them to be active in academic activities.

For oral presentation session, only a few students were not reluctant to ask or answer questions or to express opinions. Not all students participating in presentation sessions had been able to follow the discussion actively.

Most students agreed that English is a significant barrier to learning. These difficulties include grammatical mistakes, oral presentation problem, speaking nervously, problems with written material, and having a sufficient level of English for comprehension of reading materials.

Based on evaluation on assignments, quizzes, presentations, mid semester exam and final semester exam, the score average and percentage of grade's achievement are presented in Table 2 and Table 3, respectively.

Table 2. Score Average

No.	Evaluation	Score average
1.	Assignments	91,25
2.	Quizzes	71,56
3.	Presentation	73,47
4.	Mid Semester Exam	82,31
5.	Final Semester Exam	79,28

From Table 2, it can be noted that the average value for all evaluations is more than 60. There is no student failing in this subject as shown in Table 3. The percentage of 55,5 of students achieve grade B indicates the positive implication of the method proposed.

Tabel 3. The Percentage of Grade

No.	Grade	Percentage
1.	A	25
2.	B	55,5
3.	C	19,5
4.	D	0
5.	E	0

From this teaching method, the syllabus, assessment system, learning process design, and group papers were produced. These products can be used as a reference by lecturers in teaching and learning process for other subjects.

CONCLUSIONS

Based on discussion, it can be concluded:

1. The STAD method as a student-centred learning process can be applied to the Mathematical Statistics I subject successfully.
2. The study confirms problems associated with using English in teaching and learning process in tertiary level in Indonesia.

REFERENCES

Biggs, J. (2001). Teaching international students. In J. Biggs (Ed.), *Teaching for quality learning at university. What the student does*. Buckingham: Open University Press.

Buzan, T. (2005). *The Ultimate Book of Mind Maps,* diterjemahkan oleh Susi Purwoko, PT Gramedia Pustaka Utama, Jakarta.

Chalmers, D. and Volet, S. (1997). Misconception about students from South-East Asia studying in Australia. *Higher Education Research Development*, 16(1), 1-11.

Dardjowidjojo, S. (2001). Cultural constraints in the implementation of learner autonomy: The case of Indonesia. *Journal of Southeast Asian Education*, 2(20), 309-322.

Gulo, W. (2002). *Strategi Belajar Mengajar,* Gramedia Widiasarana Indonesia, Jakarta.

Novera, I. A. (2004). Indonesia Postgraduate Students studying in Australia : An Examination of their Academic, Social and Cultural Experiences. *International Education Journal*, 45(4), 475-487.

Silberman, M.L. (1996). *Active Learning: 101 Strategies to Teach Any Subject,* Temple University, Allyn & Bacon, Singapore.

Sri Anitah. (2006). *Metode Pembelajaran Berorientasi Active Learning,* disampaikan pada Pelatihan Pembelajaran Berorientasi *Active Learning,* Jurusan Matematika FMIPA UNS.

Contributed paper (refereed) – Maureen Morris

'She Ca'nt do sums a bit' or can she?: Tracking student learning (Carroll, 2002)

MORRIS, Maureen
University of Western Sydney, Australia
m.morris@uws.edu.au

Abstract
This case study formed part of an action research project reviewing development of an aligned teaching/learning framework effectively targeting statistical thinking. It tracked changes in teaching and learning across five sessions in a foundation statistics subject at the University of Wollongong. Evidence was derived from the multiple perspectives identified in student and teacher surveys, and assessment data. Intended learning was overtly defined to align the curriculum; marks and evaluation surveys detected improvements in student results and support for the pedagogy; and examination of the assessment tasks themselves indicated increases in cognitive demand. Pedagogical focus on the development of the higher order skills required for statistical thinking also promoted the framing of questions and marking criteria that gauged demonstration of achievement.

BACKGROUND

Context

The aim of this study was to develop an aligned teaching/learning framework that engaged students actively in learning to *think statistically* in order to promote the development of a *deeper approach* to learning in statistics. A consequent challenge was to effectively evaluate the impact of the framework on the desired learning. Chance's notion of *statistical thinking* was adopted for the study: "… the statistical thinker is able to move beyond what is taught in the course, to spontaneously question and investigate the issues and data involved in a specific context." (Chance, 2002, p.4). The learning required for such thinking requires employment of higher order knowledge structures and cognition.

Students in the New South Wales high school system have a limited exposure to statistics. The subject challenges students to solve problems differently to the traditional approaches they have used in other areas of mathematics and they are unprepared for the multiple alternative solutions that may arise in resolving a statistical problem. In addition there has been no attempt to require the use of technology to support the teaching and learning of statistics. (Statistical Society of Australia Inc., 2005) The result of such teaching practice has been to deliver students who have been introduced to very basic procedures accompanied by undeveloped concepts, and who have been provided with little or no modelling of statistical thinking (although changes to address these issues are being discussed). Consequently, students frequently arrive at university with negative attitudes towards the study of a subject that has become compulsory for many degree programs and which underpins espoused university graduate attributes (University of Wollongong, 2008).

The Subject under scrutiny

This study formed a focal part of the doctoral thesis of the author (Morris, 2008) and close reference has been made to that work. It tracked the teaching and learning of undergraduate statistics at the University of Wollongong in New South Wales, Australia, in a compulsory, session-long foundation course in statistics (STAT131) for Science and Computing students across five consecutive sessions. While the university is highly regarded and the students drawn to it both nationally and internationally enter with substantial academic credentials, no specified mathematical pre-requisite was defined for entry to STAT131.

STAT131 focused on "… understanding concepts and evidenced-based decision making, and topics (that) covered exploratory data analysis, probability models, regression and hypothesis testing"

(Morris, Porter and Griffiths, 2007, p. 1). Between 150 and 200 students were enrolled in the subject each session during the study. Prior to the study, students were engaged actively in the study of statistics under the guidance of committed educators who targeted statistical thinking and fair assessment of that thinking. The pedagogy was grounded in current and relevant teaching practice, but students still found the subject difficult and the assessment challenging.

THE STUDY

Gauging student learning

delMas (2002) and other researchers (Chance, 2002; Rumsey, 2002; Garfield, 2002 have described three domains as fundamental to the desired cognitive outcomes for an introductory course in statistics: basic literacy, reasoning, and thinking. The teaching intent for this subject was to engage students in all three domains, but to particularly develop their abilities to think statistically, that is to " … challenge students to apply their understanding to real world problems … (to) generalize knowledge obtained from classroom examples to new and somewhat novel situations" (delMas, 2002, pp.5-6). This intent drove the selection of teaching strategies and task design.

Traditionally teachers have accepted results of assessment as their primary evidence of student learning. This is of course appropriate if assessment detects outcome achievement and scores student achievement against those outcomes. In this foundation course in statistics, it was necessary not just to check student application of concepts and procedures, but also to detect their capacity to think statistically, transferring the basic knowledge and skills to solve new and relevant problems. This required an aligned teaching/learning/assessment framework.

A taxonomy was used to cross-classify the desired levels of knowledge and skills (Anderson & Krathwohl, 2001). Behaviourally framed learning objectives were used to align all aspects of the teaching/learning framework and alert students to the *teaching intent*. In particular, alignment was used to alleviate the cognitive mismatch between teaching and learning that arises from implementation of teaching strategies that do not address/model the expected thinking for students. All learning tasks, instructional activities, and assessment contained explicitly defined learning outcomes, and marking criteria and feedback for assessment related directly to them. This alignment of *teaching intent* and *assessment practice* facilitated reliable commentary on student achievement as reflected in assessment results. (delMas, 2002; Biggs, 1999)

Student personal awareness of learning has also been regarded as indicative of deeper learning (Watkins, C. Carnell, E. Lodge, C., Wagner, P., and Whalley, C., 2002). Students were asked to rank their perceived level of learning for the topic areas addressed in the course. The cohort proportions of these ranked topic learning perceptions were correlated with the cohort topic means to investigate any relationship between general perceptions and general performance.

Pedagogical Design

delMas argued that "… defining the student behavior that exemplifies a learning objective provides the impetus for instructional design." (delMas, 2002, p.3) Since statistical thinking was targeted in this subject, the literature was reviewed for strategies that would support student development of such thinking. The instructional design incorporated:

- opportunities for active engagement in collaborative and experiential tasks that required students to collect and analyse data;
- authentic tasks that sought solutions to relevant problems of interest to students;
- scaffolded tasks that provided guided exploration, analysis and conclusions, and culminated in evidence based recommendation/decision making;
- assessment tasks that modelled on classroom learning experiences;
- collaborative partnerships for assessment tasks;
- Detailed marking criteria that provided clear definition of targeted learning and formative feedback for students (Biggs, 1999);
- Analysis supported by technology (SPSS) that freed students to engage in more complex discipline thinking rather than focus on facile calculation (Chance, 2002);
- Student laboratory manuals that documented their subject learning. Students were able to take these manuals into their final examinations. Inherent in the manual design were key

elements of a learning portfolio. They promoted ownership of and responsibility for learning and a critical commentary on personal understanding that furthered development of the meta-cognitive skills essential for *deeper learning* (Biggs, 1998).

Assessment of learning

As many students regard assessment as defining curriculum it can provide a useful driver for student learning (Biggs, 1999). In STAT131 learning outcomes were defined for all assessment tasks, with the final exam canvassing outcomes from all of those defined for the subject. Student assessment included two or three assignment tasks that addressed discipline knowledge and skills and students' abilities to apply them in unfamiliar contexts, and a final examination. In addition, student laboratory manuals were marked for 'completeness' and as students were permitted to take them into the final examination, most students achieved close to maximum marks for them.

Assessment tasks were modelled on classroom experiences. Team approaches in assignments enabled collaboration in problem solving. Partners worked on parallel questions using different data and complementary questions requiring a partner's results for final decision making. *Marking guides* accompanying all assignments, not only aligned the tasks with the defined assessment outcomes, but also provided scaffolding for responses that connected with classroom tasks. More detailed *marking criteria* (expansions of the marking guides) enabled specific and timely feedback to both student and teacher.

Evaluation of the teaching/learning framework

Evidence of an effective framework was sought in student assessment results, in their perception of their own learning and in the detail of the tasks themselves. In order to evaluate successful learning it was necessary to:

1. survey the perceived value of the aligned teaching/learning/assessment framework;
2. deconstruct the assessment tasks using the revised taxonomy and match them with the marking criteria in order to ensure that the desired learning was appropriately recognised;
3. track proportional representation of the higher order knowledge and cognitive skills targeted in assessment across the implementation cycles.
4. track assessment results to observe changes across the five implementations;
5. survey student confidence in their own learning for the topic areas taught; and
6. examine general student perceptions of their achievement (ranked cohort perceptions of topic learning from student surveys) to see if they were associated with general cohort performance in the final exams (topic means).

Alignment of all facets of the course was evaluated with reference to: student surveys; attendance and submission rates; improvements in student mark and grade distributions and survey of participating teachers.

Success of the teaching/learning framework then, was anticipated to derive from three key areas: facilitation of student learning (through the subject presentation); constructive alignment of the pedagogy (through overt specification of the outcomes); and the pedagogical address of the specifically targeted skills (especially statistical thinking).

Methodology

Action research underpinned the methodological approach, allowing documentation of the pattern of change across a *spiral* of five successive sessions of observation, implementation, evaluation and refinement of the teaching/learning framework for STAT131 (O'Leary, 2005). The first session in 2003, was purely an observation phase. The following four sessions involved implementation followed by review and re-implementation of the renewed pedagogy. The primary task across all implementations was to determine *what worked* in the classroom. This required a pragmatic approach to review of action within the complex social dynamics of a classroom learning environment. Recognition was given to the social construction of knowledge and the collaborative roles of both teacher and student. In consequence, evidence was sought that acknowledged the multiple perspectives of the learning taking place and a mixed methods approach was engaged. The participant researcher and final arbiter of the framework's success was also the teacher. She was appropriately placed to observe, document and reflect upon the action in the classroom.

Developing and evaluating the framework for teaching statistical thinking required focus on two key issues: alignment of the teaching intent and practice with assessment; and grading of that assessment to substantiate outcome achievement. Reference to the literature and reflective examination of teaching practice enabled:

- identification and selection of a relevant pedagogical design;
- identification of recognisable hallmarks of student learning;
- alignment of teaching intent and examination practice in assessment; and
- systematic evaluation of the effectiveness of the framework in terms of evidence of student achievement.

Sources of evidence identified included:

- survey of students (attitudes to the subject presentation and assessment, and perceptions of their learning) and participating teachers (perceptions of effectiveness of subject presentation and assessment);
- peer discussion and review by educators;
- the teacher/researcher's annotated journal;
- student assessment marks and grades across the five sessions of implementation; and
- critical analysis of the knowledge and skills demanded by assessment tasks and the corresponding marking criteria with reference to a taxonomy of learning (Anderson and Krathwohl, 2001). This enabled comparison across sessions of the proportional representation of marks allocated to high and low order cognitive demand questions.

RESULTS

Although there has been evidence of enhanced learning throughout the study, the reasons underpinning the improvement have been complex and interwoven. Key data have been summarised in Table 1.

Table 1: Summary of the data supporting the aims of the project
(Note: 2003 was an observation phase and prior to framework implementation)

1. Perceived value of alignment	• Positive student survey responses on the value of key facet of the teaching/learning framework to their learning (>80% of students regarding them as important to their learning (all sessions from framework implementation); • Positive attendance (not compulsory) and submission rates evidenced perception of the worth of the subject presentation (mean lab completion mark >8.7/10 for all sessions from framework implementation); • Improved completion rates attested to student understanding of what was expected of them and their value of assessment tasks to their learning (in excess of 80% for all sessions from framework implementation; • Supportive comments and responses from participating teachers and markers. Note: although teachers and markers detected the alignment, students only commented that the subject was 'well structured' and assessment was 'fair'.

2. Tasks deconstructed to match cognitive demand of questions with graded responses	• 3 repetitions of deconstruction of random tasks yielded reliable classification using the taxonomy
3. Changes in proportional representation of higher order skills in assignments	• Proportions of marks allocated to higher order knowledge or skill rose from 45% in 2003 to 75-80% post implementation of the framework. Simultaneously student marks increased in the final exams i.e. increased cognitive demand with concurrent increase in student achievement.
4. Changes in final examination results across five sessions	• Mean exam mark rose from 39.1% in 2003 to between 51-64% for all sessions post implementation of the framework.
5. Student perceptions of topic learning	• A modest proportion of students (40-57%) declared confidence in hypothesis Testing and Confidence intervals but most students (70-94%) were confident in data exploration, and exploring relationships between quantitative variables)
6. Degree of association between general student perception of competence and their general performance	• In 2004 the cohort mean topic marks were correlated with the cohort proportions reporting confidence in their topic learning yielding $r=0.9$ ($p<0.02$)

DISCUSSION

Alignment of all aspects of the subject was secured through the defined learning outcomes (Biggs, 1999). The enhanced focus in teaching and the connections between teaching, learning and assessment enabled more effective promotion of higher order knowledge and skills. Although students commented on the 'structure' inherent within the subject, they failed to connect this with the explicitly defined outcomes. Student survey responses also highlighted their perception of the 'fairness' of the assessment and their belief that completion of classroom tasks prepared students for undertaking that assessment. Teachers indicated that the dichotomous (demonstrated/not demonstrated) marking criteria expedited marking considerably.

The means of student marks increased significantly following the first implementation. Of even greater relevance was the increased cognitive demand in the assessment task questions, with the proportion of marks allocated to higher order knowledge and skills increasing from 45% in the observation phase to 75-80% in the implementation phases. It should also be noted that although the increased cognitive demand was concurrent with significant rises in the mean marks, there were also accompanying increases in the standard deviations.

The intention was to correlate each student's rankings of confidence in topic learning with their proportional achievement in the topic assessment tasks. However the sample of examination scripts (restricted by signed permission for such analysis) was small and demonstrated a strong bias, with a much higher assessment mean than the cohort mean. As a result only cohort topic means and cohort proportions reporting topic confidence were available. The correlation between these measures does however offer a limited support of the contention that students exhibit a capacity to assess their own learning in a constructively aligned curriculum ($r=0.9$ and $p<0.02$) (Morris et al, 2004).

CONCLUSIONS

Teachers and markers in this study acknowledged the benefits of alignment of all aspects of the teaching/learning/assessment framework in promoting *statistical thinking*. Although student comments indicated that they had detected 'structure' in the subjects, none of them recognised the defined learning outcomes as the agents of the alignment. However, the impact of aligning all aspects of the framework was observable in the assessment tasks and in student performance on those tasks. Following implementation of the aligned framework, student results significantly improved.

Moreover these improvements were demonstrated on tasks that exhibited a substantial increase in cognitive demand in comparison with those in pre-implementation sessions.

The results of this study support the implementation of an aligned pedagogy that focuses on development of evidence-based reasoning and judgment in the statistics classroom.

REFERENCES

Anderson, l. W. & Krathwohl, D. R., eds. (2001). *A Taxonomy For Learning, Teaching, And Assessing: A Revision Of Bloom's Taxonomy Of Educational Objectives*. New York. U.S.A.: Addison Wesley Longman Inc.

Biggs, J. (1999). What the Student Does: teaching for enhanced learning, *Higher Education Research & Development* 18(1): 57-75.

Biggs, J. (1998). Assessment and Classroom Learning: a role for summative assessment?, *Assessment in Education* 5(1), 103-110 [Accessed 9/10/07 from Proquest]

Carroll, L. (2002). *Alice's Adventures in Wonderland and Through the Looking-Glass: and What Alice Found There*, The Modern Library: New York.

Chance, B. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment, *Journal of Statistics Education*, 10(3). [Accessed 3/7/2008 from http://amstat.org.ezproxy.uow.edu:2048/publications/jse/v10n3/chance.html]

delMas, R. (2002). Statistical Literacy;Reasoning and Learning: A Commentary, *Journal of Statistics Education*, 10(3). [Accessed 7/7/2008 from http://amstat.org.ezproxy.uow.edu:2048/publications/jse/v10n3/delMas.html]

Garfield, J. (2002). The Challenge of Developing Statistical Reasoning, *Journal of Statistics Education*,10(3). [Accessed 3/7/2008 from http://amstat.org.ezproxy.uow.edu:2048/publications/jse/v10n3/garfield.html]

Morris, M., (2004). *Getting Down and Dirty in Statistics*, Proceedings of the 2004 Workshop on Research Methods: Statistics and Finance. University of Wollongong School of Mathematics and Applied Statistics: pp. 173-186.

Morris, M., Porter, A. and Griffiths, D. (2007). Using the end to justify the means: enabling student learning of statistics, *International Journal of Learning*, 13(10), 157-168 [Available at: http://www.Learning_Journal.com].

Morris, M. (2008). *Evaluating University Teaching and Learning in an Outcome-Based Model: Replanting Bloom*, PhD thesis, School of Mathematics and Applied Statistics, University of Wollongong.

O'Leary, Z. (2005). *The Essential Guide to Doing Research*, Sage Publications: London.

Rumsey, D. (2002. Statistical Literacy as a Goal for Introductory Statistics Courses, *Journal of Statistics Education*, 10(3). [Accessed 3/7/2008 from http://amstat.org.ezproxy.uow.edu:2048/publications/jse/v10n3/rumsey.html]

Statistical Society of Australia Inc. (SSAI) (2005). *The State of Statistics in Australian Universities* [Accessed 24/1/08 from http://www.statsoc.org.au/review0405/ReviewofStatsFinalReport.pdf]

University of Wollongong, n.d., *Attributes of a UOW Graduate*. [accessed 14/05/08 from http://www.uow.edu.au/about/teaching/graduate_attributes.html]

Watkins, C., Carnell, E., Lodge, C., Wagner, P., and Whalley, C. (2002). Effective Learning, *National School Improvement Network Research Matters*, No. 13. Institute of Education: London. [Accessed 4/7/07 from http://www.ioe.ac.uk/iseic/research.pdf]

Contributed paper – Norhayati Baharun

A LEARNING DESIGN TO SUPPORT STUDENT LEARNING OF STATISTICS WITHIN AN ONLINE LEaRNING ENVIRONMENT

BAHARUN, Norhayati[1] and PORTER, Anne[2]
[1]University of Technology MARA (UiTM) Malaysia
[2]School of Mathematics and Applied Statistics
University of Wollongong, Australia
nbb470@uow.edu.au

Abstract
This paper presents the results of a study examining the use of a learning design map within an online learning environment in supporting the student learning of statistics at the University of Wollongong. This study compares two cohorts of undergraduate students who enrolled in an Introductory Statistics subject which were eighty-nine students in Autumn 2009 session (before the implementation of learning design map within subject) and hundred eighty-four students in Autumn 2010 session. The aim of the learning design map is to provide guidance to students in learning this subject on a weekly basis through variety of learning resources made available via a page of interactive links on e Learning site. The findings show the learning design map has the potential to support the student learning of statistics as to improve their learning outcomes in the subject. The paper concludes with a discussion on impact of learning designs on student learning of statistics and followed by suggestions for further research.

INTRODUCTION

Over the past twenty years, online learning or e-Learning has played an important role in the higher education sector around the world as a medium of teaching and learning using technology. Online learning allows students to manage time and space providing them with the opportunity for learning with no constraints of time zones, location, and distance (Suanpang, Petocz, & Kalceff, 2004). Suanpang et al. (2004) describes online learning as using a computer with Internet connection and other forms of information technologies to create learning experiences among students in a given course or subject. This paper will use the definition of online learning as "[t]he use of the Internet to access learning materials; to interact with the content, instructor, and other [student]s; and to obtain support during the learning process, in order to acquire knowledge, to construct personal meaning, and to grow from the learning experience" (Ally, 2008; p. 17).

According to Agostinho (2006), designing effective online learning experiences, or effective educational strategies remains a significant challenge for university educators. There are a variety of learning designs that is ways of designing student learning experiences using learning materials or resources provided within online learning environment. Designing online learning experiences may involve planning schedules, preparing or designing course or subject outlines and materials, determining assessment tasks, anticipating students' needs, implementing new learning strategies or modifying a previous course or subject by updating materials. Baharun and Porter (2010) stated it is essential for educators to understand the impact of learning designs within online learning environment so as to know how students experience the teaching and learning environment; to determine their needs and current level of expertise, and to assign appropriate learning materials for them to select from to achieve their desired learning outcomes. In other words, the implementation of appropriate learning designs within online learning systems can serve as learning supports to improve or enhance the student learning experience.

The unsatisfactory experience and student performance in statistics subjects worldwide (Onwuegbuzie, 2003; Pang & Tang, 2004) highlights the need for learning designs to improve outcomes particularly in the online learning environment. With the Internet as a subject delivery platform, the authors sought to conduct a case study which looked at the impact of learning designs within the online learning or e-Learning environment of an Introductory Statistics subject at University of Wollongong.

In this paper, we have explored the students' use and perceptions of the learning design map provided within the subject's e-Learning site. There are three questions to address: To what extent do students use the learning design map? What do the students feel they gain from using this resource? Does the learning design map support their learning in this subject?

METHOD

Subject design

In Autumn 2010 session, this subject was designed based on the learning design representation (refer http://www.learningdesigns.uow.edu.au) which focuses on three major elements of learning activities such as tasks, resources, and supports. These elements were put together and displayed as a flowchart also known as a "visual representation" or a learning design of teaching and learning activities on a weekly basis. The aim of the learning design map was to provide guidance to students in learning this subject through variety of learning resources that were made available via a page of interactive links on e-Learning site (using Blackboard system). Through this map, the students could access to the resources by "click-on" the links provided within the map (see figure 1). The primary resources included lecture notes, Edu-stream (audio recorded lectures); the tasks i.e. laboratory work, laboratory tests, worked solutions, data sets; other specific learning resources such as video resources to support most topics, SPSS notes; and ongoing support materials i.e. learning strategies, student forum, past exams and laboratory tests, student support advisers, learning modules. In Autumn 2009 session, these resources were put as a list using folders system in e-Learning site and labeled, for example; lectures, assessment, laboratory and data sets. Note, this subject used the same learning resources in both sessions, except the tests and assignment but these were designed with the same level of difficulty. The assessment system was different in that in 2009 session, students sat three in-class tests whereas in 2010 session students sat a mix of four in-class and take-home tests, with a pass mark of 70%. Students who did not pass the test the first time were required to sit a second version of the test again at the 70% level. Students availed themselves of the opportunity were then provided with feedback to aid them correct their work. This assessment system was useful allowing the lecturer to identify at risk students with a variety of learning issues and to provide the possibility of their gaining additional assistance.

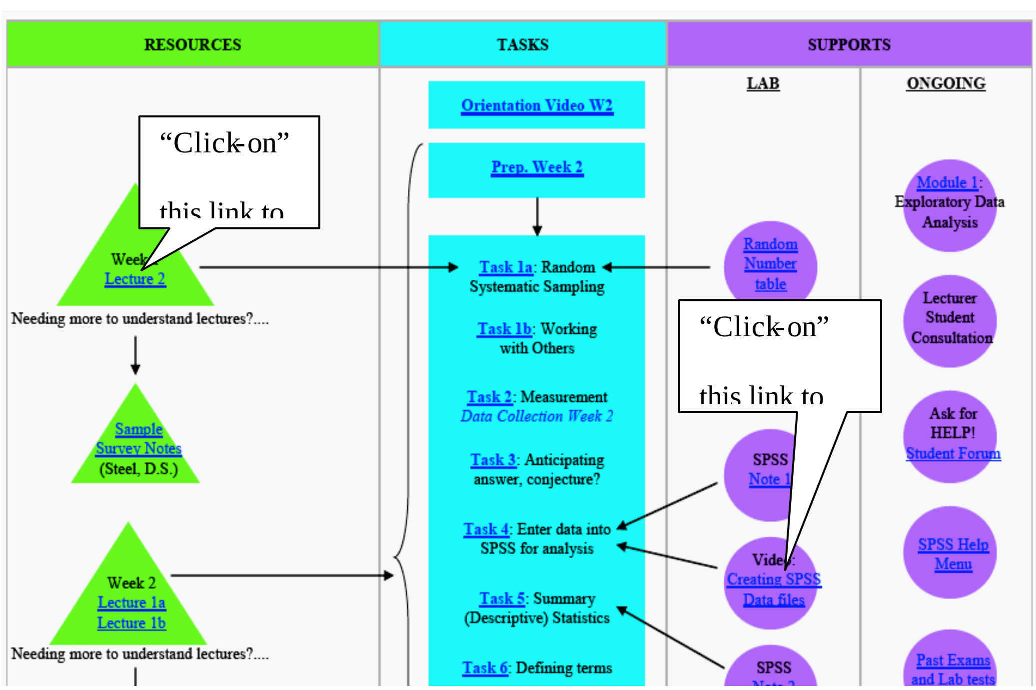


Figure 1. Learning design map for a weekly work in the subject

Procedures

Four sources of data were used in this study. First, a survey questionnaire was used to collect the background information on the students such as gender, their nationality either international or domestic students, also included items on their expectations in the subject (anticipated grades). Other questions were on their use of the learning resources particularly learning design map, their confidence in the subject on several topic areas, and their suggestions on areas to be improved in the subject. The students were asked to complete this survey via eLearning site at the end of session before their final examinations. Second, the students' learning outcomes were measured based on their assessment scores and final grades. Third, on top of that, the students were assessed on their basic statistical literacy and reasoning using a CAOS test (*Comprehensive Assessment of Outcomes in Statistics* from https://app.gen.umn.edu/artist/) at the beginning (pre-test) and the end of session (post-test) in Autumn 2010 session. This test is designed to measure students' basic literacy and it generally represents as accepted measure of statistical literacy developed by statistics education researchers (delMas, Garfield, Ooms, & Chance, 2006). Fourth, tutors were asked to comment on the value of the subject learning designs through a survey questionnaire at the end of session.

Participants

Two different cohorts of students were included in this study: 89 students who enrolled in Autumn 2009 session (before the implementation of learning design map within subject), and 191 students in Autumn 2010 session. The number of students enrolled in Autumn 2010 session almost doubled to Autumn 2009 due to a timetabling clash where the students might complete another statistics subject in Spring 2010 session as an alternative to the Autumn 2010. There was a large percentage of computing students enrolled in this subject for both cohorts. Of the Autumn 2009 students, 38 students (43%) took part in the survey while of the Autumn 2010 students, 109 students (57%) took part in survey. The information displayed in table 1 refers to those students who completed the questionnaire at the end of session. The distribution of gender highlights significant differences for the two cohorts that there were more female students enrolled in Autumn 2009 compared to Autumn 2010 session.

Table 1. Comparison of Autumn 2009 and Autumn 2010 student backgrounds

	Autumn 2009	Autumn 2010
% female*	45%	17%
% domestic students	71%	81%
% expecting credit or above	71%	59%

**significant at $p < .001$*

RESULTS

Use of the resources

In the survey questionnaire, the students were asked to rate each of the learning resources in terms of their usefulness in helping them learn and understand the subject. From previous experience (Porter, 2007) we expect the ratings of primary resources (lectures, assessment, laboratory tasks and worked solutions) to be high, above 80%, noting that the value of one resource changes with the improvement of another. Support resources such as textbooks and other learning supports tend to rate lower as they are not necessary learning aids for all students.

The students in Autumn 2010 session were also asked on their use of the learning design map that provided in eLearning site. These students were asked if they found the learning design map useful and what, if anything they thought they had gained from using it. Further, students were asked to provide their comments and recommendations on ways to improve this resource.

For the Autumn 2010 students, the majority (70%) of them found the learning design map useful. Comments indicated that some students (83%) appreciated the use of this resource because of the linking between tasks and other learning resources in the subject, the provision of references (77%), improved ability to organize their work and update learning materials (71%), for revision

(82%), and as a study checklist (62%). For example, '[I] used the flowchart to help with the tests and retests", "I used the flowchart to check that I was doing all the work and up to date with all my work", and "As a learning compass". However, there was room for improvement in the development of this resource. For example, "Integrate the map into e-Learning as a web page rather than a PDF", "Sometimes it is more convenient to have access to all lecture notes in the one folder", and "Possibly include an overall subject map pointing to resources about each topic (i.e. a map to the maps)".

In Autumn 2009 session, there were various learning resources provided and listed as folders on e-Learning site, whereas in 2010 session, the learning design map was used in the subject to provide links to these resources. The results show insignificant differences on the value of the learning resources such as lecture notes, worked solutions, video resources, laboratory tasks, and laboratory tests between sessions (see figure 2). It is almost desirable that the students in both sessions had similar resources on e-Learning site except the learning design map, which was introduced in Autumn 2010 session.

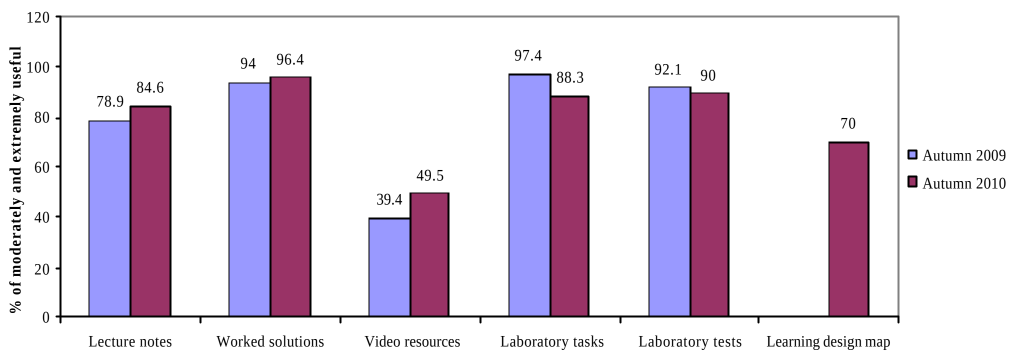


Figure 2. Value of learning resources in helping students learn and understand

Some comments made by the tutors regarding the use of learning design map on the subject e-Learning site, for example, "As a tutor, the weekly map identifying tasks and associated resources was extremely useful. It was an incredibly useful organizer in the lab class, allowing download of the particular task either the teacher or students wanted to be worked together", and "It is better than normal teaching via e-Learning site because the teaching materials are more organized".

Perceived comfort with topics

At the end of session, students were asked to indicate how confident they were in relation to each of the major topics. The results show insignificant differences in the percentage of students who indicated they "can do" on major topics between sessions (see table 2).

Table 2. Can do and moderately confident with topics

	Autumn 2009		Autumn 2010	
	Can do %	Total % (Moderately confident and Can do)	Can do %	Total % (Moderately confident and Can do)
Exploratory (measures of centre, shape, spread and outliers)	42.1	97.4	50.5	86.7
Correlation and regression	26.3	76.3	32.0	81.5
Binomial and Poisson distribution	18.4	68.4	32.0	74.7
Confidence intervals	34.2	73.7	29.5	74.3
Using SPSS	39.5	81.6	34.9	72.6
Model fitting	21.1	68.5	30.8	71.2
Hypotheses tests	31.6	79.0	24.0	68.2
Normal and exponential distribution	13.2	68.5	16.5	58.2

CAOS test scores

In Autumn 2010 session, the students were assessed on their basic statistical literacy and reasoning using the CAOS test scores at the beginning (pre-test) and the end of session (post-test). It was found that the differences in time (pre-test minus post-test) taken by students to complete the tests were positively correlated ($r = 0.248$, $p < .01$) with the differences in test scores (post-test minus pre-test). This indicates the students took less time to complete the test for the second time (post-test) and that they achieved slightly better marks. Further, the students showed significant learning gains throughout session from an average percent correct of 49.3% on the pre-test to an average percent correct of 52.5% on the post-test ($t_{171} = 3.821$, $p < .001$). While statistically significant, this was only a small average increase of 3 percentage points, with a 95% confidence interval of the difference between 1.51 and 4.75. In this subject, the students were assigned marks worth 2% of their final marks for completing both tests, in other words, they were awarded participation marks of 1% for each test irrespective of the marks they achieved in both tests. This might be a reason of low percentage of changed in CAOS test scores, as some students were not doing these tests seriously. Note, since this test was taken out of class (students completed it in their own time) the above results only based on the scores of students who completed the test between 10 to 60 minutes. The reason was to eliminate the students who did not engage sufficiently with the test items or who spent an excessive amount of time on the test, possibly those who searching for answers (delMas et al., 2006).

ASSESSMENT AND FINAL GRADES

The grades distribution of the two cohorts revealed insignificant differences in the percentage of grades achieved by students in Autumn 2009 (average mark = 52.97, s. deviation = 27.72) and in Autumn 2010 session (average mark = 59.30, s. deviation = 24.12) (see figure 3). The grades represent marks within categories: High Distinction (HD), 85-100%; Distinction (D), 75-84%; Credit (C), 65-74%; Pass (P), 50-64%; Pass Conceded (PC), 45-49%; and Fail (F), less than 45%. With respect to failures, the rate would have been much higher had the assessment system not allowed the identification of students at risk and the subsequent work with them to get them through.

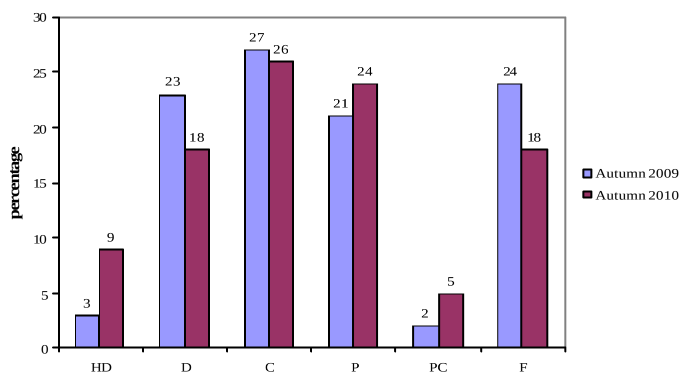


Figure 3. Grades distribution from Autumn 2009 to 2010

An analysis of assessment results indicated the average marks achieved by students in Autumn 2010 session was significantly higher than the students in Autumn 2009 session for the second test ($t_{237} = 2.136$, $p < .05$, see table 3) . However, there was no significant difference in the average marks between the two cohorts of students for the other two tests.

Table 3. Comparison of assessment marks

		N	Average	Standard.deviation
Test 1	Autumn 2009	90	5.12	2.76
(Exploratory data)	Autumn 2010	175	5.49	1.92
Test II	Autumn 2009	73	5.83	1.79
(Correlation & Regression)	Autumn 2010	192	6.52	3.39
Test III	Autumn 2009	66	7.28	1.67
(Probability & Models)	Autumn 2010	169	7.20	2.67

CONCLUSION

To name a few, Francis (2010), and Lovett, Meyer, and Thille (2010) highlighted some issues and outcomes from their studies investigating the use of online learning in statistics education area at the Eighth International Conference on Teaching Statistics held in Ljubljana, Slovenia mid of this year. In line with these studies, this paper emphasized on the importance of learning designs within online learning environment particularly in statistics subject as to deliver the learning materials or resources and support the student learning effectively.

Based on the perceptions of students and tutors of the subject, it revealed that the impact of learning design map used in the eLearning site potentially support the students to achieve their desired learning outcomes in the subject. In this paper, we were able to highlight the results on the experience of the subject design within online learning environment from two cohorts of students, Autumn 2009 and Autumn 2010 sessions. Although the resources in the subject were similar, the manner in which these resources integrated in the subject's eLearning site differed resulting a changed in student learning experience. The Autumn 2010 students commented they gained benefits from the use of learning design map, however this study suggests there is a need to design the subject within online learning environment be more interactive, easy access, multiple browsers compatibility, choices of designs variety (folders system, learning design map, webpage, concept maps, etc), and better layout. Provided complexity is not increased, it is likely that students would benefit from a choice in approaches, listing resources via type of resource as well as the learning design map that links tasks with resources and supports.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the important contributions and support provided for the production of this paper from the University of Wollongong. The first author also would like to express special appreciation on the financial support from the University of Technology MARA Malaysia, which allows her to pursue a PhD degree at the University of Wollongong.

REFERENCES

Agostinho, S. (2006). The use of visual learning design representation to document and communicate teaching ideas. In L. Markauskaite, Goodyear, P., & Riemann, P. (Eds), *Who's Learning? Whose Technology? Proceedings of the 23rd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE)*, (pp. 3-7). Sydney, Australia: Sydney University Press.

Ally, M. (2008). Foundation of Educational Theory for Online Learning. In A. Terry (Ed.), *The theory and practice of online learning (2nd ed.)* (pp. 15-44). Can: AU Press, Athabasca Univ.

Baharun, N., & Porter, A. (2010). The impact of video-based resources in teaching statistics: A comparative study of undergraduates to postgraduates. *Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: Int. Stat. Inst.

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2006). *Assessing students' conceptual understanding after a first course in statistics*. Paper presented at the Annual Meetings of The American Educational Research Asscociation, April 9, 2006. San Francisco, CA.

Francis, G. (2010). Online learning materials: Are they put to different uses by online and on campus students? *Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.

Lovett, M., Meyer, O., & Thille, C. (2010). In search of the "perfect" blend between an instructor and an online course for teaching introductory statistics. *Proceedings of the eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: Int. Stat. Inst.Onwuegbuzie, A. J. (2003). Modeling statistics achievement among graduate students. *Educational and Psychological Measurement*, 63, 1020-1038.

Pan, W. & Tang, M. (2004). Examining the effectiveness of innovative instructional methods on reducing statistics anxiety for graduate students in the social sciences. *Journal of Instructional Psychology*, 31(2), 149-159.

Porter, A. (2007). *Transitions*, Faculty of Informatics, University of Wollongong.

Suanpang, P., Petocz, P., & Kalceff, W. (2004). Student Attitudes to Learning Business Statistics: Comparison of Online and Traditional Methods. *Educational Technology & Society*, **7**(3), 9-20.

Contributed paper – Nazim Khan

ENGAGING FIRST YEAR STUDENTS

KHAN, Nazim
School of Mathematics and Statistics, University of Western Australia, Australia
nazim@maths.uwa.edu.au

Abstract
Most first year service units are difficult to teach as students are not interested and cannot see the relevance to their chosen major field of study. Additionally, students in such units are from a variety of backgrounds and mathematical ability, and are enrolled in very diverse courses. I will discuss one such unit, compulsory for all first years who come into Science with the lowest level mathematics qualification. Over the last few years this unit has had a negative perception amongst students and client faculties. I tell the story of how I approached this unit over two semesters. The student response, attitude and performance improved remarkably in these two semesters.

Topic: Evaluating Learning

Contributed paper – Brian Jersky

A CASE STUDY OF KNOWLEDGE OF KEY STATISTICAL CONCEPTS BEFORE AND AFTER AN INTRODUCTORY STATISTICS CLASS

JERSKY, Brian
Department of Statistics, Faculty of Science,
Macquarie University, Australia
Brian.jersky@mq.edu.au

Abstract

In this paper, a case study is presented that investigates some of the key concepts that students taking an introductory statistics class bring with them as they begin the class, as well as those they take with them as they leave the class. Because of the small size of the class that was investigated (n = 36), any conclusions are only indicative. The idea behind this study is that as our first-year students are increasingly exposed in high school to statistical concepts, techniques and facts, they should bring this information into our first-year classes, and hence it would be useful to investigate what, if anything, should change in our courses or curricula as a result. This study showed that students arrive in a first-year class with some knowledge, and that they leave at the end of the class with just a little more than they started with.

INTRODUCTION

All over the world, secondary and even primary school students are being introduced to statistics, usually as one or more strands in their mathematics curriculum. In the USA, as well as Australia, for example, fairly sophisticated statistical and probability facts, techniques and concepts are expected to be acquired by high school leavers. Naturally, this has led to an increase in interest among high school teachers in how to teach statistics effectively. Many statistics educators, both here and abroad, have been involved in designing such curricula, and in pre- and post-teacher qualification training in this area.

Regardless of the difficulties and challenges that are experienced by secondary or primary mathematics teachers in the teaching of statistics, we must all assume that students have acquired at least some statistical knowledge by the time they enter our first-year statistics classes. Since those were traditionally designed on the assumption that students knew very little or nothing about statistics, it is clear that this will have, or should have, quite an impact on our curricula.

For example, a common way to begin a first-year introductory statistics class is to discuss what data are, how to present them visually, and ways to summarise them numerically. Almost all students coming into my introductory classes in the past few years have already seen these ideas, often many times. So it would seem clear that our presentations of these concepts should be abbreviated, and at a higher level.

Although I had always been involved in teaching the introductory statistics classes at several colleges and universities in and around San Francisco, I took on an administrative role in the last 5 years that prevented me from teaching this class. So, when I taught it again in (northern) autumn 2009, I resolved to try to formally assess what the students knew coming into the class, preferably using some established instrument, and then to assess them after they had completed the class, so I could see what statistical knowledge, if any, my class was adding.

This paper discusses the instrument I chose to use to measure the students' knowledge, and briefly presents the results of the assessment. It closes with some discussion of my conclusions, and gives some suggestions for future research in this area. Since the particular class that this work was done with was only one small class (n = 36) from one particular college, no very general conclusions can or should be drawn from the results. Nevertheless, the students in this class seemed to be, at least informally, much the same as most other introductory-level students I have taught in California over the past 20 years, so there is some hope that the results might be indicative of more general issues.

THE ASSESSMENT INSTRUMENT

Bloom's well-known taxonomy of learning presents 6 levels of learning, hierarchically arranged as follows (Bloom, 1956):

1) Knowledge: arrange, define, duplicate, label, list, memorize, name, order, recognize, relate, recall, repeat, reproduce, state.
2) Comprehension: classify, describe, discuss, explain, express, identify, indicate, locate, recognize, report, restate, review, select, translate.
3) Application: apply, choose, demonstrate, dramatize, employ, illustrate, interpret, operate, practice, schedule, sketch, solve, use, write.
4) Analysis: analyse, appraise, calculate, categorize, compare, contrast, criticize, differentiate, discriminate, distinguish, examine, experiment, question, test.
5) Synthesis: arrange, assemble, collect, compose, construct, create, design, develop, formulate, manage, organize, plan, prepare, propose, set up, write.
6) Evaluation: appraise, argue, assess, attach, choose compare, defend estimate, judge, predict, rate, score, select, support, value, evaluate.

It would be helpful to have an assessment instrument that could measure learning in introductory statistics with this degree of precision. However, to my knowledge, no such instrument exists. An approximation to it does, however. It is the ARTIST website, a nexus of assessment tools and research into learning in introductory statistics classes. (ARTIST is an acronym for Assessment Resource Tools for Improving Statistical Thinking). It was developed by Garfield, delMas and Chance, 2006.

Among other resources found there is a rather large database of questions, covering most of the areas traditionally taught in introductory statistics courses, from which instructors can create test instruments to assess various levels as well as areas of understanding.

Garfield et al. argue that their three levels of differentiation (Statistical Literacy, Statistical Thinking, and Statistical Reasoning) are sufficiently closely related to Bloom's taxonomy, without becoming too finely divided (as they claim might be a problem if all 6 levels were to be assessed using a similar instrument). Questions can be selected from all areas of the curriculum, at the 3 levels of learning.

It is thus possible to create a test that can give the instructor quite a good snapshot of a student's learning at any one time.

Although there are both multiple-choice type questions and written response items available in the ARTIST database, I chose to use only the multiple-choice type questions, for ease of administration and of marking. The test was a 31-question test that I thought covered reasonably the kinds of statistical knowledge that a well-trained high school student would have learned, as well as a few questions that I thought would be beyond their capability as they entered the introductory class.

The test was administered twice to the class, once in the very first week of the 15-week semester, and once in the penultimate week of the semester. Students were informed that participation was voluntary, but that the instructor would greatly appreciate their help with the study. A further incentive was to offer a minimal amount of extra-credit to those who took the test. No time limit was set, and the students took the test home with them, returning it 2 days later in each case.

Somewhat to my surprise, all 36 students turned in the test in the first week. Towards the end of the semester, after 3 students had dropped the class, 27 took the test in the penultimate week. A natural pairing thus existed for analysis.

RESULTS

Gain score

The first way of looking at the data was to test the gain scores to see if there was a significant improvement in the number of correct answers given by the students. The mean increase in scores, from pre-test to post-test, was 4.3, with a p-value of 0.11 for the one-sided test of significance, so we cannot conclude that there was a significant increase at a reasonable level of significance.

This was a somewhat disappointing result, although it is of course possible to develop *post-hoc* rationalizations as to why this might have occurred. These might include such factors as the low-stakes nature of the test; inadequacy of the questions; poor teaching; lack of concordance between the multiple-choice questions and the content of my course, etc., but since we have no way of knowing which, if any, of these is operating, it is a somewhat pointless speculation.

More interesting was the fact that all but one of the 27 students who took the test in the penultimate week passed the class, and reviews of the course and of my teaching were very positive.

More detailed analysis

Generally, in multiple-choice type tests, the distractors (choices) are chosen so that some are “more” correct than others. I realized that for my data, it would perhaps be more helpful to focus on any changes in distractor choice over time, to obtain a more finely calibrated idea of what changes were occurring among the students from the beginning to the end of the class.

For example, it might be the case that students come into the first-year class with already established, firmly held beliefs about particular areas (some of which are incorrect), and additional instruction does not dislodge this belief. Another possibility is that students generally understand the concept better after the course than before it. Alternatively, it might be the case that students *move towards* the correct solution, abandoning initial incorrect beliefs, but do not quite reach the correct solution. Examples of these 3 cases follow. Within the box is the question (the stem), the distractors, and the distribution of the choices made by the students, both before and after the class.

A set of data is put in numerical order, and a statistic is calculated that divides the data set into two equal parts. Which of the following statistics was computed?

a. mean
b. interquartile range
c. standard deviation
d. median

Distribution of distractors chosen by students is below:

	Before Course	After Course
a	14%	26%
b	11%	22%
c	8%	0%
d	**67%**	**52%**

This question was chosen to address the area of “measures of centre” at the level of statistical literacy, the lowest of Garfield *et al*’s levels. This is an example of a firmly-held belief that appears not to change much in spite of instruction. Just for reference, students in California schools are introduced to the mean and the median in 5th grade (Year 5), and are taught the material over and over through 12th grade. As can be seen, most students understand that the median is the correct answer (at least to begin with), but after the class was over, somewhat fewer of them were able to conclude this. For me, it was surprising to see the relatively large number of students who believed that the mean is the statistic that was measured here.

I would be fascinated to know what prompts these students to decide what they do. Some work has been done in this area (delMas, *et al*., 2007), but I believe it would be interesting future research to investigate, perhaps using focus groups, what prompts students to choose the options they do. Work is currently in progress in this regard, and will be reported on in future.

Naturally, it may be the case that the question itself is poorly worded. That would however also come out in focus group discussions with students. I suppose there is some comfort here in the 3 students who no longer believe that this is the standard deviation, but in general, it appears that students are not as comfortable with the concept of different measures of centre as one might expect.

Students in a statistics class designed a survey about spending habits and gave it to a *random sample of 300 students, of whom 282 responded. Please read and evaluate the* following statement.

"There are over 4000 students at the college. Therefore, the results of the survey may not be valid for drawing conclusions about how all students at the college spend money."

a. Agree, 282 is too small a percentage of 4000 (7%) to allow us to draw conclusions about the population.
b. Agree, you should have a sample that is at least 50% of the population in order to make inferences.
c. Disagree, 282 is a large enough number to use for these purposes if the sample of students is random.
d. Disagree, if the sample is random, the size of the sample does not matter.

Distribution of distractors chosen by students is below:

	Before Course	After Course
a	36%	26%
b	8%	0%
c	**42%**	**63%**
d	8%	11%
-	6%	0%

This question was chosen to illustrate the area of "data production" at the level of statistical reasoning, the highest of Garfield *et al*'s levels. This is an example of a decline in the most egregious incorrect choices, with a concomitant increase in the correct choice, and is presumably evidence that students have learned something over and above their initial baseline level of knowledge. Note however the slight increase in the percentage choosing option d. Although this is clearly incorrect, it is I think easy to understand why a student with only a glimmer of statistical reasoning ability might choose this option – the word "random" makes everything work fine, for this student, perhaps. In any case, this is the distribution that one would have hoped to see more of throughout.

When conducting a formal hypothesis test, there are different errors that may be made, depending on your decision. One decision is to reject the null hypothesis. If you falsely reject the null hypothesis, what type of error has been made?

a. a Type I error
b. a Type II error
c. a Critical error
d. no error was made

This question was chosen to illustrate the area of "Test of Significance" at the level of statistical literacy. Before the class, 25% chose the correct option, and the distribution was roughly uniformly distributed across the remaining options. (Of course, a thorough reading of the question would reveal to a careful reader that option d could not be correct, but no doubt these questions were not very thoroughly read). The pre-class distribution was as expected, as this material would have been quite new to all the students, and random guessing could have produced this result.

After the class, 44% got this correct – I was relieved. Just over a third chose option c, and the rest chose b. None chose option d. Is there some connection for these students between critical values and type I errors? If so, option c might be a "better" choice in some sense than the other incorrect options. A vague connection for these students between critical levels of a test and Type I errors might be evidence that there is some rudimentary knowledge that could be built on in future courses. Again, further research will provide some answers.

Of course, this confirms that even "basic" literacy type questions are not well incorporated into students' knowledge, as is well known in the literature (Chance, et al, 2004, for example) .

DISCUSSION

Although the overall results for the students did not show significant improvement in scores, it seems clear to me that most of the students showed some kind of movement towards the correct answer. As already mentioned, of the 27 students who took the test in the penultimate week of the class, all but one passed the class overall. I believe that this bolsters my suggestion that the students have moved towards understanding, even if multiple choice questions do not demonstrate this very clearly.

Further analysis of the 31 questions on the test showed that they could quite easily be divided into the 3 types outlined in the results section above, namely: 1. Students remained at a previously established level of belief; 2. Students now generally understand the concept; 3. Students move towards understanding of the concept.

Of the 31 questions, 6 (19%) showed students remaining at previously established beliefs. The areas tested in these questions were mostly probability-based questions, confirming the widely-established idea that probability is a very difficult concept. Of the remaining 25 questions, 11 (35%) showed students now generally understanding a given area within the curriculum. The areas tested in these questions ranged over all the material covered. The remaining 14 questions (45%) showed students moving towards "more correct" understanding, and it is these that would perhaps most benefit from a second course in statistics that would cement their knowledge and make it more firmly set.

One of the crucial ideas that emerges from this kind of analysis is the fact that one course at university level in statistics is not sufficient to cement ideas for many students. This confirms what researchers are finding worldwide (Wild *et al*., 2011): "Many of the problems with students learning statistics stem from too many concepts having to be operationalized almost simultaneously. Mastering and interlinking many concepts is too much for most students to cope with. We cannot throw out large numbers of messages in a short timeframe, no matter how critical they are to good statistical practice, and hope to achieve anything but confusion. We need complexity reduction strategies. One of these is to cluster concepts into smaller, more manageable sets that share fairly well-defined spheres of influence."

There appear to be two possible solutions to this issue. The first would be to reduce the numbers of concepts we teach in first-year statistics classes. The second would be to repeat the concepts in a second class in statistics at university level. Either of these options has benefits and challenges, and I will not go into those in great detail here.

Suffice it to say that reducing the number of concepts we teach in first-year classes will not be popular with the departments we service, unless we educate our colleagues a great deal more successfully than we have to date about what students can usefully acquire in a one-semester course.

Of course, I assume that my statistical colleagues would not be averse to a second course in statistics, but this may be a minority opinion within the university.

A third possibility is to somewhat reduce what we attempt to cover in the first course, and repeat concepts to try to improve the retention of key concepts by students.

In general, it is clear that we must be aware of increased levels of statistical understanding of incoming first-year students, and that this must have an impact on what we teach and certainly how and at what level we teach our classes. On the other hand, we must also be aware of where high schools are not teaching effectively, and must enhance the teaching of secondary school teachers to assist them. Ultimately, this will of course benefit university-level instruction.

CONCLUSION

A short multiple choice test, administered before and after a first year introductory statistics class, showed no significant improvement in scores for students.

However, closer analysis of the results seems to show that students generally move towards correct understanding, even though not quite reaching full understanding. However, some incorrect ideas, especially in probability, appear difficult to change.

It would be of interest to repeat the study on a larger group of students, and to follow up the results with small focus groups of students, perhaps divided by achievement level in the first year statistics class, to understand what leads students to choose the options they did.

This will be the subject of a future paper.

REFERENCES

Bloom B. S. (1956). Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay Co Inc.

ARTIST website: https://app.gen.umn.edu/artist/index.html (Garfield, J. et al., 2006)

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, *6*(2), 28-58.

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about Sampling Distributions. In D. Ben-Zvi and J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking*. Dordrecht. Kluwer

Wild, C., Pfannkuch M., & Regan M. (2010). "Towards more accessible conceptions of statistical inference", *J. R. Statist. Soc.* A (2011) 174, *Part* 2, *pp.* 1–23 (http://www.rss.org.uk/pdf/Wild_Oct._2010.pdf)

Contributed paper – Mitra Jazayeri

UNDERSTANDING OF SAMPLING VARIABILITY: CONFRONTING STUDENTS' MISCONCEPTION IN RELATION TO SAMPLING VARIABILITY

JAZAYERI Mitra, LAI Jerry, FIDLER Fiona and CUMMING Geoff
Department of Mathematics and Statistics, School of Psychological Science
La Trobe University
m.jazayeri@latrobe.edu.au

Abstract
Psychology students typically have difficulty understanding the relationship between the sample size and variability. For example, they often do not understand that variability of sample means decreases as the size of the samples drawn increases and instead believe variability increases with sample size. Understanding this is crucial to their understanding of standard error. We designed a teaching intervention which directly confronted this misconception. The intervention was developed by applying cognitive conflict as the basis for conceptual change. We tested this intervention on a class of n=185, 'Statistics for Psychology' students. The result of the Pearson Chi-Square test of independence indicates moderate evidence regarding the association between the level of the understanding of the students about the variability of the sample means for different sample sizes and the method of teaching.

INTRODUCTION

Students studying psychology at a tertiary level typically have difficulty understanding sampling variability, and in particular, how sampling variability is influenced by changes in the size of samples drawn. Students' misconceptions about basic principals obviously influence their ability to understand more advanced material. Posner, Strike, Hewson and Gertzog (1982) explain in more detail how students' misconceptions can act as a barrier towards learning new concepts. A lot of attention has been paid to identifying student misconceptions, especially the misconceptions of psychology students. However, there has been surprisingly little research on intervention strategies of teaching methods for addressing this, with the exception of Kalinowski (2008). Kalinowski tested two interventions developed to mitigate against the inverse probability fallacy and found that even short instruction in Bayes theorem and the logic of NHST (Modus Tollens arguments) decreased levels of misconception. The current study uses Kalinowski as a model.

As highlighted by Özdemir and Douglas (2007), conceptual change theory has been implemented in different fields of science to help students overcome misconceptions. In practice, this involves asking the students a question which directly involves their misconception. In other words, as stated by Duit (1999) the students are first led to commit the fallacy or mistake that is to be eradicated. To create a state of cognitive conflict, students are then challenged with a series of arguments and examples exposing the misconceptions. The affect of this is to instigate feelings of dissatisfaction with their current reasoning and understanding. This process may involve challenges from peers as well as from the instructor. Finally, the correct concept is introduced. As a result, when the correct concept is introduced, students should find the new concepts more reasonable and easier to understand.

The research reported in this paper applied conceptual change theory to a class of psychology students studying statistics, focussing on understanding the relationship between sample size and variability.

METHOD

Participants were 185 first year undergraduate students enrolled in the Statistics for Psychology unit at La Trobe University, Australia. The statistical background of the students varied, however all students were familiar with the concept of sample size and variability.

A simple questionnaire consisting of questions about feelings and attitudes towards statistics as well as content questions about sampling variability was administered to students at three time points. 1. As a baseline measure in week 2 of semester. 2. As a direct post-test measure following the intervention in week 9 of semester and 3. As a follow-up in week 13 of semester. The sampling variability content items are shown in Figure 1.

The intervention itself was a brief instruction carried out in the weekly compulsory computer lab classes. Each lab class has approximately 20 students. Half of the lab classes received the intervention; the other half received 'standard' instruction. Allocation to 'standard' or 'intervention' instruction was random. Topics covered in all lab classes were: sample size, population, sample means, standard deviation and standard error concepts. All lab classes were taken by the lecturer for this course, namely the first author of this paper.

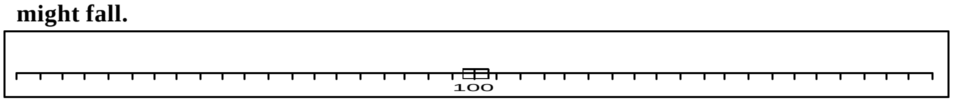
The distinguishing feature of the intervention group was direct confrontation. At the beginning of the session, the lecturer asked a question about the comparison of the variability of the sample means when there were a high number of samples with very small sample sizes and when there were many samples with very large sample sizes.

First students were asked what they thought would happen to the variability of the sample means if the sample size was increased. Many students had the misconception that the variability would increase. To illustrate this, we worked through an example, in which there was a small sample size followed by looking at larger sample sizes. The students were then directed to how the size of the sample affected the variability of the sample means. As part of the intervention procedure, the lecturer highlighted to the intervention groups, when the students were giving the wrong answer. The lecturer explained to the students the correct answers, in an attempt to alter their previous misconceptions, and confront the students' confusions.

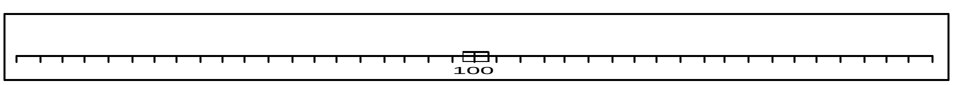
A follow up study was given to the students on week 13 of the semester in order to investigate how much information was retained by students.

Let's say that the average IQ score of the population of science students is exactly 100. There are three identical axes below and the population mean is marked on each one. For any sample, the sample mean is unlikely to be *exactly* the population mean. **This question asks you estimate some typical sample means from samples of different sizes**.

a) Imagine you take 10 samples, each of size **n=5**. Where might 10 typical sample means for this sample size fall? **Please make 10 little marks on the horizontal axis, where you think the means might fall.**



b) Imagine you take 10 samples, each of size **n=20**. Where might 10 typical sample means for this sample size fall?



c) Imagine you take 10 samples, each of size **n=80**. Where might 10 typical sample means for this sample size fall?

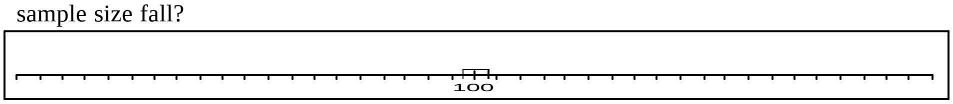


Figure 1: Questionnaire design

RESULTS

Descriptive statistics about the participants

Out of 185 students enrolled in first year psychology unit, 140 participated in the survey (110 female, 28 male, 2 missing values). The majority of students (79.7%) were females. One hundred ten students provided valid information about their Enter score. The Enter scores ranged from 53.55 to 97.50 with the average Enter score of 80.63 and the standard deviation of 9.07. Nine percent of the participants had no mathematical background. Forty one percent of the students had studied "Fundamentals of Mathematics" and 36% of the students had carried out "Mathematical methods" and 3.6% of the students studied "Specialist Mathematics" as one of their year twelve subjects. Ten percent of the students had other mathematical backgrounds. Table 1 shows students' feelings and attitudes towards 'statistics' subject at the beginning of the semester.

Questions:	All positive	Neutral	All Negative
1. How would you describe your general **feeling** towards statistics?	21.5%	29.3%	49.3%
2. How **confident** do you feel about this subject?	30.7%	28.6%	40.7%
3. How **hard** would you say statistics is for you?	18.6%	30.7%	50.7%

Table 1. Students' feelings and attitudes towards 'statistics' subject [at the beginning of the semester]

Analysis of the baseline data:

Only 59.3% (n of 140) of students provided responses for all the three parts of the question regarding the sampling distribution of the means. We assume that the remainder had too little knowledge to complete all the items and that in general students found the questions to be difficult. The remaining 40.7% answered some parts of the questionnaire; there were no completely unattempted surveys.

Figure 2 represents the comparison of the performance of standard and intervention groups, at the post-test (immediately following instruction). The 'Incorrect Response' category on the x-axis represents the category of the students which claim variability of the sample means increases with the size of the samples. The 'Correct Response' category on the x-axis in the other hand represents the category of the students who correctly identify variability as decreasing when increasing the size of samples.

The result of the Pearson Chi-Square test of independence indicates significant evidence regarding the association between the level of the understanding of the students about the variability of the sample means for different sample sizes and the method of teaching.(chi-square=3.766, df=1, *p-value*=0.048)

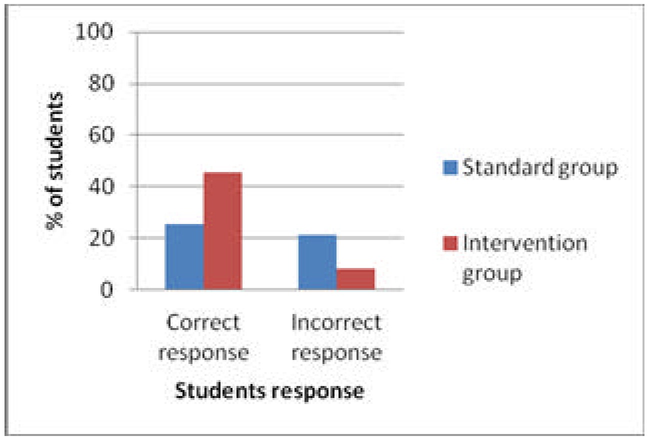


Figure 2: Percentages showing correct and incorrect responses in the standard versus intervention groups.

Figure 3 represents two line graphs overlayed on the same axes. The lines display the SD from the sample means of the students' responses for three different sample sizes (n=5, n=20 and n=100) for the Standard and Intervention groups. As can be seen, in the Standard group, the standard deviation of the sample means of the students' responses increases when the sample size increases. On the other hand, in the Intervention group the standard deviation of the sample means of the students' responses decreases as n increases.

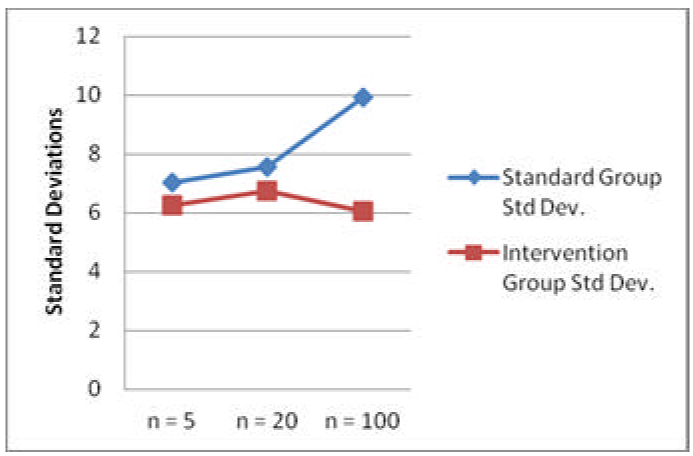


Figure 3: Comparison of standard deviations of the standard and intervention groups

CONCLUSION

The research reported in this paper concerns the effects of applying conceptual change theory to a class of psychology students studying statistics, with a focus on understanding the relationship between sample size and variability. The findings of this research indicated that of the two teaching methodologies evaluated, the application of the conceptual change theory results in a 29.3% positive effect on the students learning. This is a considerable effect, given that the intervention developed was only 10 minutes long and very simple to administer. 89% of the students in the intervention group retained the information regarding the sampling variability whereas only 45% of the standard group retained this information in the follow up study at the end of the semester. Future work could include the investigation of different perspectives of conceptual change theory to ascertain an optimal approach for teaching statistical concepts to tertiary students.

REFERENCES

Duit, R. (1999). Conceptual change approaches in science education. In W. Schnotz, S. Vosniadou, & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 263–282). Amsterdam, NL: Pergamon.

Kalinowski, P. (2008). Identifying Misconceptions about confidence intervals. *ICOTS 8 contributed paper*.

Özdemir, G., & Douglas, C. (2007). An overview of conceptual change theories, *Eurasia Journal of Mathematics, Science & Technology Education,* 3(4), 351-36.

Posner, G.J., Strike, K.a., Hewson, P.W., Gertzog, W. A. (1982). Accommodation of a scientific conception: toward a theory of conceptual change. *Science Education* 66(2): 211-227.

Sinatra, G., Pintrich, P., and Erlbaum, L. (2003). Intentional conceptual change, Mahwah, NJ: Erlbaum.