Engaging learners in statistical education: some ‘whys’ and ‘hows’

Mike Camden

Statistics New Zealand, Wellington, NEW ZEALAND
mike.camden@stats.govt.nz

Abstract

Statistical educators have increasingly vital roles in equipping people for both work and life, and so we have an increasingly vital role in engaging them as learners of statistics. Luckily, the applied nature of our subject supports learner engagement, and there are many rapidly developing communication methods that we can harness to support it as well. We will view some illustrations of these from school, workplace, and official statistics. The methods may come from outside the tertiary sector, but they have implications for it.

The illustrations will include: data exploration software for 9-year-olds and beyond, media-based activities to engage 12-year-olds, videos and datasets from scientists for 17-year-olds, aims for school curricula, aims for workplace statistical education, methods for data visualisation, informal inference, resampling methods, treatment of hot issues like climate change, and the sea of information available from official statistics agencies.

I hope we’ll conclude that engaging learners in our subject is vitally important, that there are plenty of smart and developing ways of supporting this engagement, and that many of these methods need careful design, dedicated effort, and large amounts of teamwork.

Keywords: statistics, learning, media, visualisation

1. Introduction: fresh opportunities

In our century, the learning of statistics is moving from being important for human life to being vital for it. Our learners are in schools, the tertiary sector, workplaces, and in civil society. They all need skills in evidence-based working and living. One of our responsibilities is to engage these learners in statistics. MacGillivray (2009), in a submission on the Australian national school curriculum, stated the principle that mathematical and statistical skills are complementary, and are needed in all disciplines. They are needed at the other levels too.

We have made much progress in recent decades (reformations, transformations, creative approaches, and fresh resources), but it seems that disquiet continues about how our subject is perceived. Hand (2009) asks why such an exciting subject is viewed otherwise. Meng (2009) sees it as ‘both desired and feared’. However, we are in a strong situation: the nature of our subject, combined with fresh communication methods, opens up colourful new opportunities for engaging our learners.

The paper illustrates some of these opportunities, from a viewpoint that is mostly about school and the official statistics workplace. There are implications from them for the tertiary sector. The illustrations come from the writer’s experience, but there are equivalent developments in Australia.

The paper does not discuss how these opportunities succeed in engaging learners. If we are to use these opportunities successfully, we need to be clear about what engagement is, how it varies from learner to learner, how it can be measured, and how engagement impacts on the learning of sound statistical skills. These issues are outside the scope of the paper.

The conclusion is about developing, using and evaluating resources for engaging learners: these

resources can be really attractive, but they need expertise and teamwork involving statisticians, educators, and communication experts. There are many challenges in designing and building these resources.

2. Software for young learners

TinkerPlots™.is 'software for Grades 4-8', for 'dynamic data exploration' (Key Curriculum Press, 2009a). Its older sibling, Fathom™, is 'dynamic data software' for secondary students (Key Curriculum Press, 2009a). Some school curricula, like The New Zealand Curriculum (Ministry of Education, 2007), require that students enact the 'statistical enquiry cycle': an investigation starts with context and purpose, and progresses to plan, data, analysis, and conclusions. The datasets are 'multivariate'. TinkerPlots is designed so that learners engage in this cycle at the data stage. It presents them with a visualisation of the records in the dataset as a plot of randomly scattered points, and a list of the variables in the (multivariate) dataset. Learners build their own graphs by dragging one or two variables into the plot area, and clicking a third variable that shows via colour. They drag points to get the level of disaggregation that they want. The learners have to be active in deciding which distributions and relationships to look at, and in building their visualisations. We hope that they then make sound observations and interpretations.

In his recent visit to New Zealand (NZ) at the NZ Association of Mathematics Teachers' conference, TinkerPlots designer Cliff Konold commented that exploratory data analysis by children is about excitement: 'a fire to ignite', and about 'seeing the world and questioning what we see'. In 'Designing a data analysis tool for learners,' Konold (2007) describes his aims for the software.

Perhaps the solution to the attitudinal problems stated by Hand and Meng is for us to re-create the excitement with our older learners, and take a similar incendiary approach. Software and attitudes like these could also be very useful in the official statistics workplace.

Software can enable learners to explore statistical datasets, or statistical concepts (like the central limit theorem), or both. An example of software mainly for the concepts is Computer Assisted Statistics Teaching (CAST) (Stirling, 2009). TinkerPlots can or will have features for both purposes.

The R software, and in particular the R-Commander package (Fox, 2006), is a magnificent resource for data exploration. It can help ignite data analysis for learners at many levels.

In a few years, we will have students and new colleagues arriving, who have been 'ignited' as data explorers from when they were aged 9. This early engagement has implications for us. The first is that these learners will have different skills, and different expectations about the tools we give them. The second is that their data exploration experience needs to have been statistically sound. We need to work with the mathematics education community to ensure that teachers are supported with sound resources and professional learning experiences.

3. Statistics in the Media: a resource for engaging primary and intermediate students

Statistics in the Media is a pair of booklets (LearningMedia, 2009a, 2009b) that seeks to engage students aged about 10 and 12 years respectively. The booklets contain 14 and 11 statistical investigations respectively that relate to cell-phone usage, TV and radio programming, and other media-related contexts that concern these students. The student-friendly design will help with engagement too. The booklets and the teachers' notes inform students and teachers about the statistical focus, key concepts and learning goals for each activity. Figure 1 shows the start of *Texting Olympics* from the booklet for 12-year- olds.

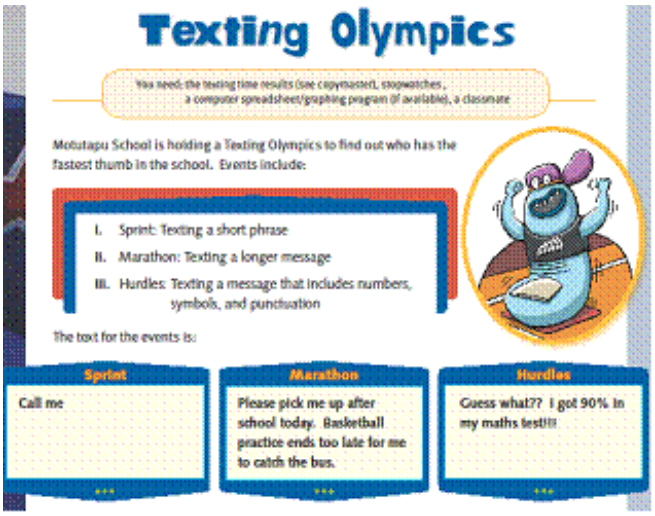

Figure 1. The authors of Statistics in the Media aim to engage statistical learners aged about 12.

The activities lead students into graphing the data, exploring its properties, and reaching conclusions. The activities have the subversive aim of engaging students by leading them to demand "If Ms Bayes's class can do this, why can't we?"

Motivation for this resource included the success of previous similarly-styled statistics booklets for these age-group, and the diverse set of statistical applications that exist in the youth media context.

A very diverse team built this resource: a secondary media teacher and an intermediate teacher (who provided the creative ideas), a group of maths educators (more ideas), the Ministry of Education (funding), the publishers (design, editing, photography, graphics),

classes of students (testing, images and data), and even some statisticians (comments about statistical and graphical soundness).

The implication for tertiary educators is that we too need resources and learning experiences that will engage students. There certainly are some (quality texts on paper or screen, and in other media), but more are increasingly possible. Their production requires creative thinking, teamwork and careful design.

4. Data Visualisation

Graphics do seem to engage most people. Today, there are more opportunities for engaging our learners with statistical graphics than ever before. *Interactive graphics for data analysis* (Theus and Urbanek, 2008) systematises the current developments, and illustrates them beautifully. There are challenges for us in why and how we use new graphical tools to engage our learners:

- they (the learners) need to become both consumers and producers of data graphics
- they may need graphics skills for data exploration, model building and checking, and for communication and presentation
- they need to apply these skills when they are in our classes, but they also need to apply them later in living and life-long learning
- they can use static but multivariate graphics (scatterplot matrices, 3D plots, lattice and trellis plots, etc), and dynamic and interactive graphics (some of which will be familiar to them from websites)
- they need sound statistical skills for dealing with all these tools.

In the statistical workplace, there is a great need for fresh skills with graphs. Behind the firewall of an official statistics agency, as well as the analysis of data, a vast amount of data cleaning and editing goes on. Graphics would greatly benefit tasks like these.

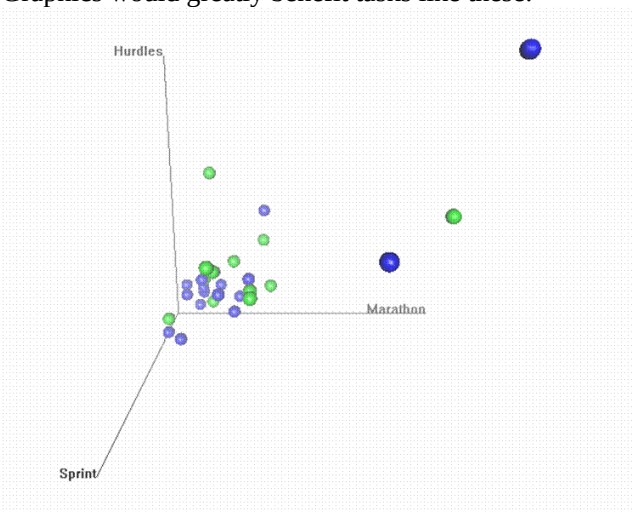

Figure 2. The graphic, from R-Commander, shows all the four variables from Text Olympics. Girls and boys show as dark and light respectively. It appears that most of the fast thumbs belong to a certain gender.

The graph (Figure 2) uses perspective and encodes 'near' as 'large'. We need to ensure that our learners perceive and use complex graphics with understanding of their strengths and weaknesses. In this case, 'large' does not mean 'more important'.

In school, there are vast areas of data visualisation to be explored. The NZ Curriculum requires 'data displays' throughout. It is obvious where these start (one-variable graphs) but it is not clear where they finish. We need to explore how far the learners can go, or whether they ever need to stop.

We need to investigate how software and graphics (with sound interpretations) can be used by learners in school, the tertiary sector, and the workplace. We also need to ensure that less visual learners have their needs met. For all learners, we need to ensure that data visualisation skills complement other investigation skills, and that the whole skill set is integrated.

5. Videos and data from researchers

Harraway (2009) has recently released a new resource that uses real data in context to engage and motivate statistical learners in the last year of school. The resource has 12 items (with more to come). In each, there is a video of about 15 minutes where a researcher describes their study, the dataset, and some student tasks. The descriptions use the curriculum's statistical enquiry cycle, in a wide range of science, health and social science contexts. For example, Stewart's item (Figure 2) on disbudding (de-horning) of cattle involves animal ethics, experimental design, data visualisation, and analysis. The resource is freely available, and can be used at other levels.

.

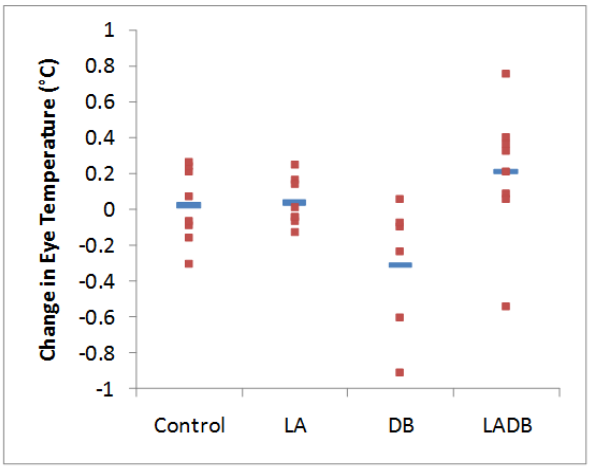

Figure 3. In Stewart's video, temperature change relates to local anaesthetic (LA), disbudding (DB), and both.

Motivation for this resource included the success of the previous set of videos (Harraway and van der Vyver, 2007), and the strong need among teachers for datasets with contexts.

Clearly, the production of this resource involved large amounts of teamwork for the statistician, the researchers, teachers, media experts, and funders.

6. Statistical learning in the workplace

Official statistics agencies are very interested in statistical learning, for several sets of people: their own methodology teams, the much larger group of statistical analysts throughout the organisations, policy and other analysts in the rest of the official statistics sectors, and the public, including the education community. A statistically literate public is important, not only as consumers of statistics, but also as people who understand and value their role as respondents.

Statistics NZ arranges learning experiences for all these sets of learners. For our own methodologists, we seek to enhance their tertiary skills with further work on topics like these:

1. statistical thinking
2. sample design for business and social collections
3. time series and seasonal adjustment
4. editing and imputation
5. data integration
6. confidentiality.

Sample design is a key concern, and leads on to courses on multistage sampling, variance, resampling and replicate weights, and estimation.

For our analysts, we offer courses on the same six topics, and a new course on survey development. For analysts here and elsewhere in official statistics, we offer a custom-designed certificate in official statistics (Statistics NZ, 2009) that provides 'the ability to use statistics to inform policy development'.

How do we engage these workplace learners? Content is related as closely as possible to the current needs of the learners. Often the courses are arranged to be 'just-in-time'. All courses contain short activities and/or longer projects that are often done in teams.

An ultimate statistical learning situation is the workplace one where a person has a new and hard problem, and no existing solutions. We support this learning by networks of expertise, mentoring, in-house research projects, contact with academics, and contact with other official statistics agencies. Engagement here often involves data visualisation.

An example of workplace learning and engagement is our recent study of information disturbance in turning an original dataset into a confidentialised unit record file (CURF). An established method is to make tables from both files, subtract, and find the average deviation in the differences. If instead we graph the percentage difference with other variables (Figure 4), we discover useful relationships. We discover more about the whereabouts, sources, and effects of the information disturbance.

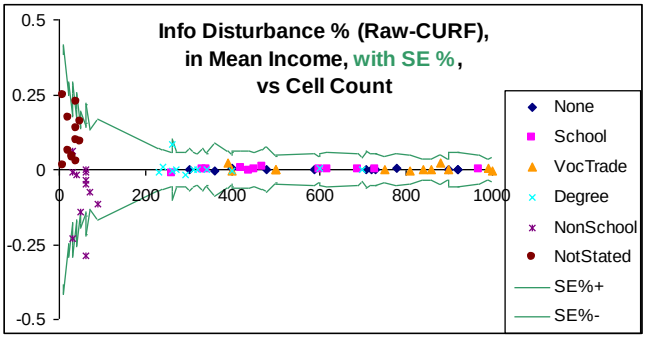

Figure 4. The noise added by confidentialisation is related to cell count, category of qualification, and sampling error. The largest noise comes from two categories with low counts. Noise is similar to or much smaller than sampling error.

Statistics NZ engages learners from businesses and community groups with interactive seminars that are customised to the audience, and centred on how they can use our data. Statistics NZ engages similarly with schoolteachers: they, with their students, can become very intensive users of our data products.

7. Resampling, intuition, and informal inference

At first, resampling methods may seem abstruse. A closer look shows that they are ideal for engaging some learners in statistical inference. Resampling methods can be related to intuitive and informal concepts about populations and samples, they can use very active and visual learning experiences, and they are very Australian in origin. The NZ Curriculum states 'make inferences … using methods such as resampling and randomisation' at its final level, but the concept can be used earlier. As statistical educators, we have exciting opportunities to try out these methods with learners.

The use of resampling complements current research into informal inference, such as that in the special issue of the Statistics Education Research Journal on informal inferential reasoning (Pratt and Ainley, 2008), and the work at the University of Auckland (Pfannkuch, 2007, and CensusAtSchool, 2009).

An example of using resampling methods to engage learners is the learning of confidence intervals via the bootstrap. The concept is smart commonsense: if we have a good sample, the rest of the population will look a lot like it. The learning process can be active: students resample first by putting themselves in groups, then by using data cards, then by using software. All the results need to be made visual. The confidence interval itself becomes visual.

The use of resampling needs to complement the learning of the more traditional forms of inference.

Perhaps resampling is one of several statistical innovations, where we can combine a commonsense basis with intuition and visualisation.

8. Hot issues

Nicholson, Ridgeway and McCusker (2006) met with teachers of mathematics and citizenship, and noted : 'A striking feature … was how strongly they felt about the need for materials which were genuinely relevant to the main issues facing the pupils … . These included teenage pregnancy and sexually transmitted infections'. Harraway's videos include one on 'effect of circumcision on sexually transmitted diseases', and one on this controversial local issue: 'measurement of public opinion on support for funding of the new Otago Stadium' (2009). Nolan and Temple Lang believe that 'statistics education is in dire need of exposing students to real scientific and social problems' (2007).

A particularly hot current issue is climate change. In seeking to engage learners, do we avoid these issues because they are 'too hot to handle', or use them because they are 'too hot not to handle'?

Climate change has aspects that relate to the learning of statistics: masses of new time-series and other data, modelling, data visualisation, risk analysis, and evidence-based decision-making. If statistics is a key part of the academic and scientific process for creating and applying new knowledge, perhaps we have a role in making sure that statistical learners understand this process. In using hot issues like these, we need to be prepared for how people respond to potentially unpleasant prospects. Perhaps our role is to equip them to move into evidence-based problem-solving.

A recent discussion on the ANZSTAT email network concludes that we need to have a grip on the underlying science, and that we need to adhere to scientific processes and balanced risk assessment.

9. Excitement from Official Statistics

Excitement and official statistics do not need to be mutually exclusive. Official statistics agencies need to communicate with an audience that mostly grew up before the revolutions in statistical education. The agencies have new tools that they can use to communicate that are interactive, dynamic, graphical and cartographic. Examples can be found on the Australian Bureau of Statistics website (2008), like the dynamic population pyramids.

Our learners can use and analyse these products, more intensively than their intended audience can. Resources include reports to critique for statistical literacy, and series datasets to support time series analysis. The need for confidentiality limits the use of case data, but some synthetic unit record datasets (SURFs) may be publically available.

10. Conclusions: 'whys' and 'hows"

The 'whys' for engaging learners in statistics are strong and clear: science and social science are essential for our development and decision-making, and statistics is a vital part of that.

The 'hows' exist because our discipline is based on applications and commonsense. Many of our recent new methodologies and tools are therefore not only accessible to a range of learners, but fit in well with their intuitions and current skills.

Variation is central for us as statisticians, and variation exists in the learning preferences of our clients. Some of the opportunities above will appeal to some learners, and some will appeal to others. Any resource needs to be designed so that it meets a range of learning preferences. Any resource should ideally be examined for the impacts it has, positive and possibly negative for some, on engagement and on learning. We can make use of the variation among learners to achieve quality improvement for the resources.

Exciting new resources now exist, and there are opportunities for many more. Development requires research into learning processes, expert design, and teamwork. There are many challenges for us all, in envisioning the resources and leading the teams.

References

Australian Bureau of Statistics. (2008). *Animated population pyramids*. Retrieved, 10 November, 2009, from http://www.abs.gov.au/websitedbs/d3310114.nsf/home/population%20pyramid%20preview

Census at School NZ: (2009). *Informal inference*. Retrieved, 10 November, from www.censusatschool.org.nz/2009/informal-inference

Fox, J. (2006).*Getting started with the R Commander*. Retrieved, 10 November, from http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/

Hand, D. (2009). Modern statistics: the myth and the magic. *Journal of the Royal Statistical Society*, 172(2), 287-306.

Harraway, J. (2009). *Statistics video presentations*. Retrieved, 10 November, 2009, from www.maths.otago.ac.nz/video/statistics

Harraway, J. & van der Vyver. Editors. (2007). Statistics in research: a visual teaching resource. [DVD and CD-ROM]. Dunedin, New Zealand: University of Otago.

Key Curriculum Press. (2009a) *Fathom ™: Dynamic Data Software*. Retrieved, 10 November, from http://www.keypress.com/x5656.xml

Key Curriculum Press, (2009b*). TinkerPlots ®: Dynamic Data Exploration*. Retrieved, 10 November, from. http://www.keypress.com/x5715.xml

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

MacGillivray, H. (2009) *General principles for statistics within national school curriculum*. Submission from National Forum on Building Networks in Statistical Education. Queensland, Australia.

Meng, X-L. (2009). Desired and feared – what do we do now and over the next 50 years?. *The American Statistician*, 63 (2) 202-210.

Ministry of Education. (2007). *The New Zealand Curriculum*. Wellington: Learning Media. Retrieved, 10 November, from http://nzcurriculum.tki.org.nz/Curriculum-documents

Ministry of Education. (2009a). *Statistics in the media*: Levels 3+-4. Wellington: Learning Media.

Ministry of Education. (2009b). *Statistics in the media*: Levels 4-4+. Wellington: Learning Media.

Nicholson, J., Ridgway, J. & McCusker, S. (2006). Reasoning with evidence – developing curriculum. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. The Netherlands: International Statistical Institute.

Nolan, D. & Temple Lang, D. (2007). Dynamic, interactive documents for teaching statistical practice. *International Statistical Review*, 75(3), 295–321.

Pfannkuch, M. (2007). Year 11 students' informal inferential reasoning: A case study about the interpretation of box plots. *International Electronic Journal of Mathematics Education*, 2(3), 149-167. Retrieved, 12 November, from http://www.iejme.com/

Pratt, D. & Ainley, J. (2008). Introducing the special issue on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 3-4. Retrieved, 12 November, from http://www.stat.auckland.ac.nz/serj

Statistics NZ. (2009). Certificate of official statistics. Statistics New Zealand. Retrieved, 10 November, from http://www.statisphere.govt.nz/certificate-of-official-statistics.aspx

Stirling, W. D. (2009). *CAST: Public e-books: Textbooks for learning Statistics*. Retrieved, 1 November, from http://cast.massey.ac.nz/core/index.html?collection=Public

Theus, M. & Urbanek, S. (.2009). *Interactive graphics for data analysis: principles and examples*. London: Chapman & Hall.