PUPILS REASONING WITH INFORMATION AND MISINFORMATION

Jim Ridgway and James Nicholson
School of Education, University of Durham, United Kingdom
jim.ridgway@durham.ac.uk

Real-world problems are complex; the ability to engage with such problems should be an educational goal. We know little about the effects of exposure to information, or about the cycle of revising personal stories, beliefs and theories, in the light of information and misinformation (IAM). A second emerging educational goal is the ability to reason with IAM. We are working with mash-ups comprising interactive multivariate displays of survey data, and newspaper articles on a particular theme (e.g. smoking, alcohol abuse). We report on the 'natural' statistical skills that pupils display when engaging with the media linked to relevant complex data, and show that some 'big statistical ideas' – such as interaction, and effect size – can be acquired by young students. We argue that the statistics curriculum should be reformed to include qualitative descriptions of large scale multivariate data sets, and to include the critique of rival accounts.

EMERGING EDUCATIONAL GOALS

The world is becoming increasingly 'connected'. Events that happen and decisions that are taken in one place can have profound effects on people who are far away both spatially and temporally. There is an urgent need for actions that are coherent at the global level, and at national and regional levels. Actions and decisions need to be aligned at the global level, but need to be implemented by individuals in their day-to-day behaviours. Obvious examples of problems that need to be tackled by global cooperation include global warming, and the financial crisis. Examples of changes in the behaviour of individuals that have had a large scale impact include reductions in smoking, and increases in recycling. Changes in lifestyles are far more likely to happen if people understand the reasons for change. Appropriate political decisions are far more likely to be made if citizens understand real-world problems, and are intolerant of short-term vote-winning popularism.

Real-world problems have a number of features that make them difficult to understand. Problems are almost always multivariate. Variables interact; there are non-linear relationships; variables often have piecewise functions over different parts of their range. For example, if one considers a well understood phenomenon such as the growth of a particular plant, the problem is multivariate (temperature, light, water, carbon dioxide etc. are all relevant); and variables interact (if there is no light or water then there is no growth, and when there are sufficient quantities of everything, there is growth); there are non-linear functions (at low temperatures and at high temperatures, growth is slower than at some optimal intermediate point); and functional relationships between variables are different over different values of each variable (for temperatures below and above certain temperatures there is no plant growth; between these two values there is an inverted U-shaped function). There is a set of emerging educational goals associated with understanding key ideas associated with modelling real-world problems.

An important part of problem solving is problem comprehension. In real-world problems, the task of understanding the problem can require considerable effort. A positive aspect of an increasingly connected world is that access to information has become dramatically easier, as a result of technology. For example, The Organisation for Economic Cooperation and Development now present data contained in their *Factbook* via the internet on *Gapminder*. A negative aspect of a connected world is that less reputable groups are using the internet for 'astroturfing' (creating artificial "grass roots" movements) or simply to promote their own interests (e.g. the extensive editing of Wikipedia by the CIA and Vatican revealed by *Wikiscanner*). Citizens are exposed to information and misinformation that is relevant to their personal lives, their social lives, and their political lives. Handling information and misinformation effectively is an emerging educational goal.

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. www.stat.auckland.ac.nz/~iase/publications.php

TECHNOLOGY CAN HELP

At present, we know rather little about the effects of exposure to information, or about the cycle of revising personal stories, beliefs and theories, or actions, in the light of information and misinformation. ICT provides some exciting pedagogical opportunities that we have only just begun to explore. Here, we use an example from our work using mash-ups–presentations that comprise interactive displays together with newspaper articles targeted on the topic of the interactive displays. The SMART Centre has designed a number of software 'shells' in Macromedia Flash® that run on web browsers, that facilitate the display of multivariate data (http://www.dur.ac.uk/smart.centre/).

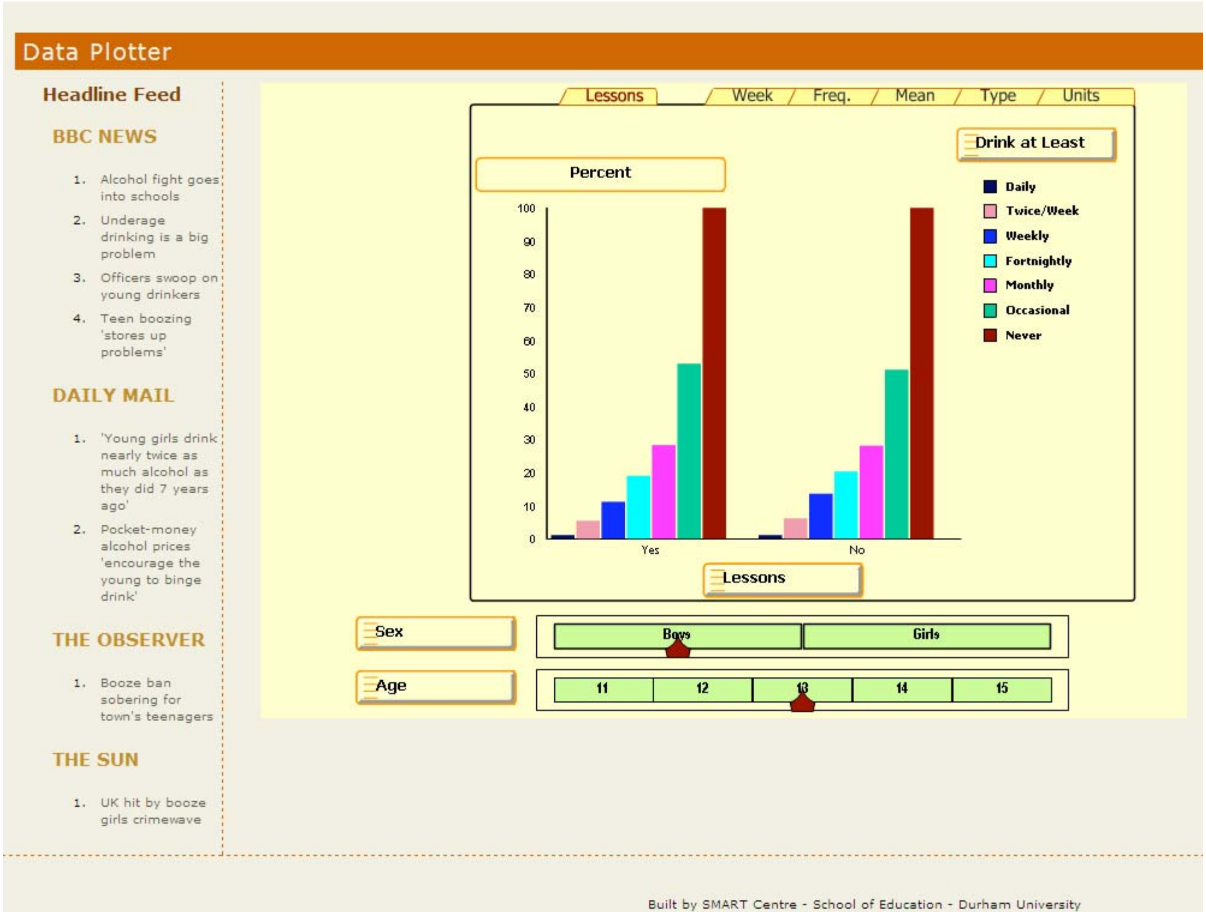

Figure 1. A Mash-up on Alcohol

Figure 1 shows a mashup where all the elements relate to alcohol consumption by young people. Data are taken from a number of national surveys. One data set presents information on the percentage of boys and girls aged 11-15 years who reported drinking alcohol in the previous week, for the period between 1994 and 2004. Another data set (published in 2008) shows the frequency of drinking alcohol by boys and girls aged 11 to 15 years**,** for students who had and had not received lessons on alcohol consumption. Other tabs in the display give access to data sets on the number of units drunk, and the type of alcohol consumed, by age and sex, over time. The interactive data displays are accompanied by newspaper articles (e.g., *Young girls drink nearly twice as much alcohol as they did 7 years ago* - Daily Mail; *UK hit by booze girls' crimewave*–The Sun) that contain statements purporting to be facts, and theories about causes of drunkenness, and crime, supposedly based on evidence. The screen shot in Figure 1 shows a comparison of the frequency of drinking between those who had experienced school lessons about alcohol in the past 12 months and those who had not. This graph is for 13 year old males, and by moving the sliders, similar data for girls, and for a range of ages, can be displayed. The positions of the age, sex and lessons variables can be interchanged to allow a fuller understanding of the relationships between them. The interface can be explored at http://www.dur.ac.uk/smart.centre1/mashup/alcohol.htm.

STUDENTS CAN DO MORE THAN WE THINK

In one study based on this display, students aged 13-15 years from 4 different schools worked in groups for about 3 hours of class time. They created ‘open responses’ such as writing a letter to the editor or a newspaper article, or creating a presentation (video or powerpoint) on the topic of alcohol use. Students were not given any formal teaching directly related to the tasks, nor any explicit instruction about how to proceed.

Ninety pieces of work were produced. An analytic framework was developed to describe student work in a systematic way. The first phase of the development of the analytic framework was a discussion between two researchers based on a large sample of student work that included students of different ages and abilities, working on both letters and newspaper articles. The second phase of the development began with the writing of formative feedback on a sample of student work, in order to make visible the observations and value judgments of the researchers. On the basis of an analysis of the formative feedback, a framework was developed, designed to crystallise the key features of the formative feedback. The essential components of the analytic framework are: the use of argument; use of empirical data, and literary aspects of the responses.

Analysis of the student responses showed that students engaged well with the tasks. Responses generally showed evidence of a sense of audience, and were structured logically. Only four responses were judged to be poor in terms of style *and* sense of audience *and* structure and logical coherence. Eighty three reports made use of data. Nineteen reports described trends clearly and accurately (a further 10 reports described trends with some success). Fifty four reports used data accurately to critique or create text media. Thirteen out of 90 responses actually showed evidence that students could discuss two-way interactions.

This last result is important. The idea of interaction is central to understanding most real-world problems, but in England and Wales is judged to be too difficult to be part of the school curriculum (see Ridgway, Nicholson & McCusker, 2007a).

LESSONS FROM LESSONS

The bulk of current research on the understanding of simple statistical concepts suggests that statistical literacy is very weakly developed, and that students and adults have difficulty with simple statistical ideas (e.g., Batanero, Godino, Vallecillos, Green & Holmes, 1994; Gal, 2002; Schield, 2006). In this study, we have evidence that some young students, untaught, discuss important ideas such as interaction. Why might this be the case?

One possibility is that the literature has not examined a broad enough range of components of statistical literacy–in particular, the exploration of some of the most important conceptual ideas have been ignored in the pursuit of the exploration of the mastery of technical skill. A further possibility is that presenting short isolated technical tasks that are devoid of context leads to an underestimation of student ability. It is reasonable to suppose that an important element in uncovering student competencies is the nature of the tasks themselves–authentic, meaningful tasks that are personally relevant to students are likely to increase student engagement, and to provide a better arena for them to demonstrate what they know, understand, and can do.

In addition to the study reported above, we have conducted studies in a wide variety of classrooms (that included low attaining and disaffected students) using large scale data sets on topics such as drug use by young people, sexually transmitted diseases, and poverty. Informal evidence from observation, and from discussions with teachers, is of very high levels of student engagement with the tasks and the evidence presented. A somewhat surprising result has been the very positive affective response from low attaining students. Part of this, we believe, arises from a number of interlinked factors. The contexts are of direct and immediate interest to them and their community. Students can see that the data are authentic and have not been censored by considerations of political correctness (e.g. in the drugs data set, the group that reports itself as being ‘happiest’ uses cannabis every day). Students are asked questions about some of the real challenges of our time, not low level technical exercises set in a spurious context.

The evidence from classroom studies reported above, gives grounds for optimism about the possibility of curriculum reform, but far more evidence is needed. We do have evidence from a psychometric study on the difficulty of tasks based on multivariate data presented via computer, compared with cognitively simple paper- based tasks. Ridgway, Nicholson and McCusker (2007b) used Rasch scaling of the responses of 13 year-old students and found that some computer based multivariate tasks were no more difficult than cognitively simpler paper-based tasks.

Taken together, the evidence suggests that the difficulty of some aspects of statistical literacy have been seriously overestimated, and that it is reasonable to present students with tasks involving reasoning with data that appear to be far harder than the tasks they would meet anywhere in the current mathematics and statistics curriculum, and far closer to the challenges of dealing with information and misinformation in the world outside the classroom.

LESSONS FOR THE CURRICULUM

Ridgway, Nicholson and McCusker (2007a) analysed all the statistics examination papers for 2004 in England and Wales that students have to pass in order to gain entry to university. The analysis showed that most questions involved univariate problems; at most two variables were involved in any question, and there, the relationships between variables were always linear. Few questions required any interpretation of results – success in statistics is about the mastery of technique, not about interpreting the implications of analyses. We can conclude that this statistics curriculum is not about understanding phenomena by quantitative modelling, but rather is about developing a very narrow set of technical skills, ill-suited to understanding real-world problems. This is completely unacceptable. How might we remedy this lamentable state of affairs?

The evidence to hand shows that students across the ability range can reason with complex evidence. An obvious development for the statistics curriculum is to introduce realistic problems, and the exploration of authentic data sets. Students should be exposed to real-world problems, and big statistical ideas from an early age. There are important statistical ideas that do not depend on the mastery of statistical technique – such as effect size, and multiple causality. We believe that some big ideas are easy to grasp in complex contexts, but will appear to be difficult concepts if all the preceding experiences of statistics have been very simple. Examples include: interaction, non-linearity, and piecewise functional relationships between variables.

There is considerable scope for promulgating statistical ideas across the curriculum – or, to be more precise, in particular subject areas where they are urgently needed, such as citizenship, and personal and social education. One strong message from our work across the curriculum is that teachers who define themselves as being non-numerate can engage students in ‘qualitative’ and ‘semi-qualitative/semi quantitative’ discussions that show some good statistical insights. There is a need to offer colleagues good quality data in accessible formats, and to offer advice on valid, invalid and indeterminate conclusions.

Modesty about the applicability of particular statistical techniques is overdue. Statistics is just one sort of mathematical modelling. It has its roots in the 1920s when the absence of computers led to the creation of a set of modelling tools that made very strong assumptions about data structures (e.g., linearity) in order to make problems tractable. Mastery of these tools has become an end in itself. Now we are concerned with messier problems, and are not much limited by access to computational power. As a thought experiment, consider modelling plant growth (whose features were discussed in an earlier section). How useful would statistical knowledge (as characterised by standard statistical packages, such as SPSS®) be, compared with a qualitative exploration of data facilitated by interactive displays? For sophisticated users, there is no dilemma here–one explores before one models (and one knows of some cunning transformations to help things along). For less sophisticated adults–say people without a degree in statistics–SPSS would probably lead to quite erroneous conclusions. Students in school, we believe, could describe qualitative features of the data, but any formal statistical analysis they conducted would be wrong, supposing it were intelligible at all.

FUTURE WORK

Further research is needed to address some major issues. We need to define and describe attainment in yet more detail, and to map out the skills that are critical for dealing with information and misinformation. There are some hierarchies of knowledge that can be derived from logical principles (describing linear trends should be easier than describing interactions) that need to be supplemented by empirical studies (e.g., Watson and Callingham, 2003). Work needs to be done developing a list of appropriate heuristics that are useful when dealing with multivariate data, along with curriculum activities that support the development of these heuristics. There is a need to understand teacher and pupil misconceptions, and to develop remedial activities.

CONCLUSION

In democratic societies the global, national and local problems we face cannot be solved without citizen 'buy-in'. This requires the promulgation of ideas on how to understand and address complex problems. The challenge is made more difficult because real-world problems are set in a context of easy access to information and misinformation. Dealing effectively with information and misinformation is another emerging educational goal. In some countries, the school statistics curriculum provides a very bad preparation for engaging with real-world evidence, by focusing on the mastery of technique at the expense of understanding and interpretation. Here, we have shown that many untutored students are able to work effectively with complex multivariate data presented in a mashup, and that the current curriculum seriously underestimates the extent to which students can reason with complex evidence. We need to broaden our conceptions of the big ideas appropriate for school statistics, and should embrace semi-qualitative and semi-quantitative approaches to describing phenomena as important educational targets. We should also enlist help from an unlikely source–colleagues who declare themselves to be non-numerate, but who have a good understanding of social phenomena, and considerable skills in helping students to articulate complex ideas.

REFERENCES

Batanero, C., Godino, J. D., Vallecillos, A., Green, D., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematics, Education, Science and Technology*, *25*(4), 527-547.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*, 1-51.

Organisation for Economic Cooperation and Development (2009). *OECD Factbook 2009: Economic, Environmental and Social Statistics*. OECD: Paris.

Ridgway, J., Nicholson, J. R., & McCusker, S. (2007a). Teaching statistics–despite its applications. *Teaching Statistics, 29*(2), 44-48.

Ridgway, J., Nicholson, J., & McCusker, S. (2007b). Reasoning with Multivariate Evidence. *International Electronic Journal of Mathematics Education* 2(3), 245-269.

Schield, M. (2006). Statistical literacy survey analysis: reading graphs and tables of rates and percentages**.** In A. Rossman & B. Chance (Eds.), Proceedings of *the Seventh International Conference on Teaching Statistics, Salvador, Brazil.* CD-ROM. International Statistical Institute and International Association for Statistical Education. Online: http://www.stat.auckland.ac.nz/~iase/publications.php.

Watson, J. M., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3-46.

Alcohol mash-up http://www.dur.ac.uk/smart.centre1/mashup/alcohol.htm.
Gapminder http://www.gapminder.org/
SMART Centre http://www.dur.ac.uk/smart.centre/
Wikipedia http://www.wikipedia.org/
Wikiscanner http://wikiscanner.virgil.gr/