MAPPING NEW STATISTICAL ILITERACIES AND LITERACIES

RIDGWAY, Jim, NICHOLSON, James and McCUSKER, Sean
University of Durham, DH1 1TA, UK

ABSTRACT

Statistical literacy (SL) has long been a concern in mathematical education. Information is becoming increasingly important in policy making for governments and in business environments, highlighting the need for improving levels of statistical literacy – including the need to deal with multivariate data. New technologies are providing new modes of communication which change our understanding of what constitutes literacy. We describe initiatives to introduce new progress indicator measures into governments' thinking. We argue that the rise in People Net (PN) websites, and the familiarity of young people with this technology could provide a vehicle for improving statistical literacy. However, analysis of PN sites suggests that a great deal of work has yet to be done. Current literature provides evidence that SL is hard to acquire. We present evidence that understanding MV data need NOT be difficult, and we outline some key targets for future research.

INTRODUCTION

Statistical literacy (SL) has long been a concern in mathematical education, and a number of authors have set out to describe the nature of SL, and to tease apart relevant constructs such as 'statistical knowledge', 'statistical reasoning', and the like (e.g. Callingham and Watson, 2005; Gal, 2004; Garfield and Ben-Zvi, 2007; Watson and Callingham, 2003). Authors do not agree on definitions, but there is a consensus on the need for citizens to be able to reason effectively with evidence presented as numbers. The ability to reason with numerical evidence appears to be rather difficult (e.g. Batanero et al, 1994); many displays of data in the media are inappropriate (e.g. Tufte, 1997; Wainer, 2000), despite the availability of some excellent advice (e.g. Miller, 2005).

In this paper, we outline the potential educational benefits of 'people net' software in helping the development of SL – and contrast the potential benefits with current actualities. We then describe some classroom trials where students across the ability range – including students with special educational needs – engage successfully with multivariate data.

MULTIVARIATE DATA, GLOBAL POLITICS, AND INFORMED DEMOCRACIES

There is an emerging international consensus driven *inter alia* by OECD and the EU on the need to measure a country's progress on a range of indicators such as social cohesion, 'wealth' as judged by a composite of measures such as natural and renewable resources as well as GDP, and the happiness of citizens (e.g. the 2007 *Beyond GDP* conference). This change of measures, and a deliberate attempt to get more citizens involved in the democratic process, is leading to the creation of new measures of social progress (which are conceptually and technically problematic – such as 'happiness'), and to new forms of data communication (such as dashboards, and interactive displays of OECD data). A number of Statistics Offices are attempting to increase the uses made of their data by displaying data on public sites such as *Swivel* and *Many Eyes.* The need for citizens to become statistically literate has never been greater, and is increasing in tandem with the increased globalisation of communication about evidence.

In the UK, the Statistics and Registration Service Act (2007) gives the Statistics Board the objective of promoting official statistics that serve the public good. While there is no definition of exactly what this entails, there does seem to be an expectation that better explanation of the messages contained in official statistics will be an integral part of the development of the service; this will raise a number of significant challenges.

The move towards evidence informed policy makes the challenge of SL particularly acute – citizens (and politicians) are expected to be able to understand complex evidence that has been gathered from a number of different sources, and analysed in different ways. These data are multivariate; there are non-linear relationships between variables, and complex interactions.

Statistical illiteracy poses considerable problems for this agenda, and statistics education in the school curriculum (in the UK at least) does not equip students to deal with anything but the simplest of relationships (for example, in post compulsory education in England, Wales and Northern Ireland, students taking the most advanced courses in statistics are asked to reason with at most two variables, and these variables are always linearly related – see Ridgway *et al.,* 2007a).

SALVATION? PEOPLE NET SITES FOR DATA SHARING AND EXPLORATION

There has been a huge increase in both the technological capacity of the internet and its use by people as a social networking environment. Sites such as *YouTube* and *Facebook* offer individuals the opportunity to present information about themselves (and indeed about anything) in a variety of forums. People *are* interested in reasoning about evidence; sites such as *Snopes,* which addresses urban myths, receives a huge number of hits. Viégas *et al*., (2006) reported a study on the use of email where histories of communications were investigated through a visualisation tool called *Themail*. The researchers put a lot of effort into ensuring the confidentiality of results. They were surprised to find that many of the participants in the study wanted to share the outcomes with family and friends, and to engage in discussions about what the data visualisations could reveal.

Sites such as *Swivel* and *Many Eyes* present a wide variety of shells into which data can be loaded and displayed. They offer internet users the opportunity to upload data sets and graphical representations to the web, and to enter into discussions about them. Major data providers (including OECD and a number of national statistics offices) have posted data sets, and offer professional critiques of comments on the data. PN software should, in principle, become an excellent vehicle for inculcating SL. If PN technologies are to make a significant contribution to the development of SL, they need to be usable without compromising the notion of community engagement and co-development. Key features for active citizen engagement leading to enhanced SL include:

- high quality data sets from reputable sources;
- high quality interactive displays, appropriate to the data being displayed;
- lively and engaging commentaries – even if (perhaps, *especially* if) they contain reasoning errors;
- some 'professional' mentoring, where misconceptions are explored;
- effective search facilities.

At the time of writing, PN sites largely fail on each of these criteria:

- many of the data sets have dubious provenance (the *Official Source* data sets are a notable exception);
- many of the data displays are poor, and have little interactivity;
- there is no facility to identify data sets by the size of the discussions they have generated. One might imagine that topic hubs, where members join a special interest group would be a place to pick up interesting discussions, but a search of all 105 *Many Eyes* topic hubs on March 7th 2008 showed no topic hub with more than 4 comments. There are very few substantive comments about data sets on *Swivel*. These results should be compared with the traffic on *Snopes*, or real-time discussions around sporting events;
- professional mentoring is fine where it occurs (e.g. OECD responses to comments on their data sets), but at present there is little to comment on;
- it should be easy to search resources first by theme, and then using more detailed criteria. On December 7th 2007, the OECD site within *Swivel* had 220, 952 graphs, but no way to search them other than 'most viewed' or 'recently uploaded'. Within the site there are 4 'groups' which are communities looking at *population; the environment; energy; labour*. Once a selection has been made, no further selection is possible by topic. The ease with which material can be uploaded and stored raises many management and accessibility issues, such as how you navigate through such a vast catalogue of material.

Any data representation carries an implicit theory about the nature of the data (and sometimes contains an implicit theory about the way the data should be modelled). The ease with which material can be uploaded can lead to a number of conceptual problems. Naïve users are free to display data in inappropriate ways; some of the features of the graphical tools that are available facilitate impressive misrepresentations of data. We explore these problems, below.

ANALYSIS OF GRAPHICAL DISPLAYS ON PN SITES

Swivel and *Many Eyes* both allow general users to upload data onto a publicly accessible site. Users have to register before they can upload data or make comments - but anyone can view data. *Many Eyes* has more types of graphical display available than *Swivel*, with matrix charts offering some multi-dimensionality, and the user has more control over the functionality by defining variable types. However, neither site has resolved issues about how to automate the production of graphs. As a result many of the graphs on display are meaningless.

Figure 1: A Graph with 3 Distinct Scales on the Vertical Axis.

Figure 1 is from the UN Statistics Division; the key identifies the units of the vertical scale for each plotted variable. *Unemployment* is measured in thousands; *School Life Expectancy* is measured in years, and the *Adolescent Fertility Rate* is a percentage. Meaningful comparisons are extremely difficult, if not impossible to make. If *Population* (measured in thousands) is also selected, only a single bar is distinguishable – because the scale on the vertical axis now stretches to 14,000k.

Search facilities on *Many Eyes* are somewhat rudimentary. On March 5th 2008, there were 14,604 data sets on the site; 666 of these were selected by a search on *population*. A search on *UK population* resulted in 3 hits. One was Yukon (a Canadian territory); another was the Ukraine, listed in a data set giving ethnic populations in Russia; and the third gave the incidences of Mad Cow Disease in the UK.

Visualisations and datasets can be rated – a user clicks an up or down arrow to change the rating incrementally by 1. The range of these was from -2 to +3 on the 14,604 data sets, presented, suggesting that this does not provide a tool for discriminating quality.

Some of the visualisations offer really striking insights into datasets, for example, the matrix chart in Figure 2a shows the survival statistics for the Titanic sinking, by gender, age and ship status (crew and class of passenger).

Figure 2a: Survival Data from the Titanic

Figure 2b: An Alternative Representation

In Figure 2a the rows show people who did and did not survive; the columns show their status on board, the colours in the pie charts give gender breakdowns; and the size of the pie chart is determined by the size of that group (or it could be represented as stacked bar charts in each of the cells). Interchanging the positions of the sex and survival variables, see Figure 2b, gives an alternative perspective, where the proportions in the pie charts are now giving the survival rates in each of the eight groups. There is a different emphasis in these two representations, but both are very powerful in communicating key messages from a complex data set.

Swivel's interface allows the user to choose between *absolute values*, and values relative to either *range* or *average*. Individually, these three graphs give quite different sorts of information, yet the vertical axis does not identify whether values are relative to the range or to the average. This is compounded by the default being *relative to range* when working with a single data set, but *Swivel* automatically switches to *relative to average* when a comparison is asked for with another data set, without the interface giving any indication that a change in structure has been initiated by the software. A dataset was created and uploaded into *Swivel* to investigate the various representations.

Figure 3a: Temperature in ^{0}F against Time

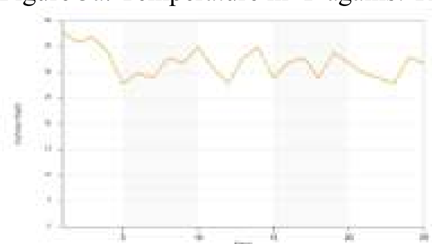
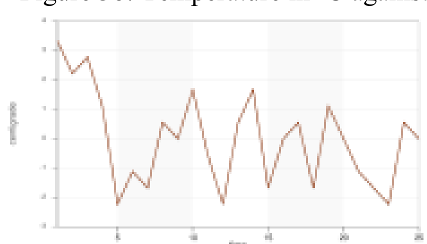


Figure 3b: Temperature in ^{0}C against Time



Figures 3a and 3b show the same temperature data recorded in Fahrenheit and Centigrade. The autoscaling feature means the same data looks quite different in terms of how variable the temperature has been.

Figure 4a: Relative to Range (^{0}F)

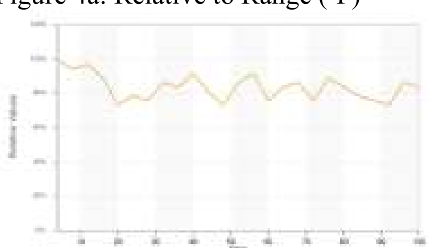


Figure 4b: Relative to Range (^{0}C)

One might expect that graphs showing the same set of data which has been linearly transformed (here, between the Fahrenheit and Centigrade temperature scales) would necessarily look identical when the vertical scale is transformed to be relative to scale, but Figures 4a and 4b show that this is not the case.

Looking at the comparison feature does not help to clarify the situation. In Figure 5a the *relative to average* comparison is used: the 'straight line' graph shows the Fahrenheit temperatures - the vertical scale goes to 10^{13}(%), because the mean of the centigrade temperatures (which should be exactly zero on the data put in) is recorded as $4x10^{-15}$. Figure 5b shows the *relative to range* comparison discussed above, where the visual impression of the variability of the two data sets is misleading, but not to the extent of the display in Figure 5a where one data set now appears to have no variability at all.

Figure 5a: Relative to Average Comparison

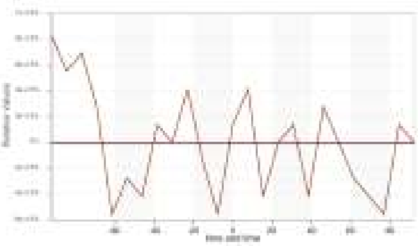
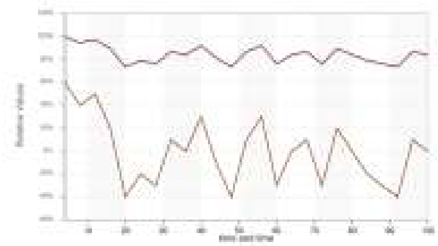


Figure 5b: Relative to Range Comparison



The behaviour of these three types of scale are very different, and while everyone is familiar with absolute values, they pose a huge problem if you want to overlay different types of information, as the representation of the UN survey data in *Many Eyes* shows. Using relative values does overcome this difficulty, but introduces other problems and the interpretation of the graphs requires considerable effort in an environment where superficiality and speed appear to be dominant influences.

These shortcomings imply that currently PN sites do not do much to enhance SL. This does not mean that they do not have the potential to become powerful tools for the development of SL, if some of the structural deficiencies can be overcome, and worthwhile discussions of complex data can be undertaken and mediated by expert contributions. There is an obvious need for tutorial support for users of MV data displays. This could take the form of advice that users could call upon when deciding how to present data. An alternative could be to automatically trigger a set of questions about the data, when (or before) a particular presentation format is chosen. Some form of automated critique of the choice of presentation could also be offered.

STATISTICAL LITERACY AND MULTIVARIATE DATA

The above section began with an optimistic note on the possible contributions that PN software could make to SL, but then identified some serious problems with some features currently built into PN websites for data display. In this section, we present evidence on the potential of carefully designed interfaces to make MV data accessible people with a range of abilities.

Garfield and Ben-Zvi (2007) revisited the principles of how students learn statistics, and restate 8 statements from Garfield (1995):

1. *Students learn by constructing knowledge;*
2. *Students learn by active involvement in learning activities;*
3. *Students learn to do well only what they practice doing;*
4. *It is easy to underestimate the difficulty students have in understanding basic concepts of probability and statistics;*
5. *It is easy to overestimate how well students understand basic concepts;*
6. *Learning is enhanced by having students become aware of and confront their errors in reasoning;*
7. *Technological tools should be used to help students visualize and explore data, not just to follow algorithms to pre-determined ends;*
8. *Students learn better if they receive consistent and helpful feedback on their performance.*

Many of these summary statements (1,2,3,6,8) can be seen as a blueprint for constructivist teaching that has been advocated by many authors, going back at least as afar as Dewey (e.g. 1916, Chapter 6). Statements 4 and 5 refer to the inherent difficulty of the subject matter, and statement 7 advocates the use of ICT. The inherent difficulty of basic concepts in statistics presented via paper and pencil is well documented (e.g. Batanero *et al.,*1994), and one might predict that the interpretation of MV data would be impossible for most people. However, our early studies suggest that this is not the case. Pupils aged 13 years found reasoning with 3D data presented via computer to be no more difficult than reasoning with much simpler 1- and 2D data presented on paper (Ridgway *et al.,* 2007b).

We have designed curriculum materials based around sophisticated interfaces which are coherent with Garfield's principles (although this alignment comes about by overlapping world views about the nature of knowledge and learning, rather than by direct influence). These curriculum materials are based on large scale, authentic, MV data about obesity, alcohol and drug use, poverty and sexually transmitted infections – data have often been collected from the age groups of the students taking part in our studies. Interfaces can be viewed at www.durham.ac.uk/smart.centre. Interfaces engage users in the active exploration of data on topics of immediate interest (sex, drugs and alcohol). Students face open challenges to explain the data, and work in small groups. Students decide which variables to plot, and it is very easy for them to explore trends over time simply by sliding the cursor along the *time* axis. These activities are strongly 'constructivist' in nature (so are consistent with Garfield's principles).

Elsewhere (Ridgway *et al.*, 2007c; Ridgway *et al.,* 2008) we have reported on successful classroom trials using these materials. Teachers reported that students engaged more actively in reasoning and discussing than in other lessons. The increased level of engagement shown by most students and the willingness to participate in discussion makes it far easier for teachers to have a realistic view of what each student is contributing, and of their level of understanding of difficult statistical concepts.

CLASSROOM TRIALS

Here, we report on classroom trials with students in different schools which produced very positive outcomes in circumstances that might seem unpromising. First, we report on a lesson about sexually transmitted diseases (STIs) where 14 year-old girls were taught by an unfamiliar male teacher, with a male observer whom they had never met before. The second was a lesson on pensions for students with special educational needs.

The data on STIs was used with a class of 14 year old girls attending a single-sex grammar school i.e. a school where students are accepted on the basis of academic ability. The lesson was taken by a male Personal, Social, and Health Education (PSHE) teacher who did not normally teach the group. This group of students would not usually deal with STIs at all until two years later within their current curriculum. The teacher had a number of books and articles which covered different aspects of STIs.

Initially, the girls were quite awkward in their language (they had two middle-aged males in the classroom, neither of whom they were familiar with) but this lessened rapidly as the lesson went on (it lasted an hour). The last 10 minutes was used by the girls to summarise for the class, what their small group had found in the open investigation.

In the debriefing session with students the end of the lesson, the girls explained that they found it relatively easy to ask questions about things they were concerned about (which were potentially embarrassing) by relating questions to the data – this provided a neutral reference point which made it easier to raise questions relating to sexual behaviour and practices, and provided a focus for reading different sections of the reference material the teacher had brought with him.

The teacher's commentary produced similar reflections. He found the data interface provided a focus which allowed a lot of the major issues to be developed quite naturally, and even though these girls were younger than those he had taught about STIs previously, he felt that they had learnt a lot. He also expressed the view that dealing with STIs with 14 year-olds rather than 16 year-olds was probably more appropriate – and that this approach made it more realistic to think of introducing the topic at age 14. He was not a mathematics teacher, but felt that the graphs were easy to use and manipulate, and that the focus was constantly on the information in the display rather than on 'mathematics', and he was very pleased with the level of reporting displayed by the girls in the plenary session at the end of the lesson.

The second study took place in a school that does not select for academic attainment on entry. Its intake has fallen both numerically and in its ability profile over recent years because of demographic trends. This school had successfully trialed the unit on alcohol data with pupils in Year 9 (age 14 years) who were of average ability within the school. We asked the teacher if she would trial a revised version of our unit on pensions and savings with another group. She suggested we work with an older group who were classified as having special educational needs. In the context of this school, most students in this group were expected to leave compulsory education with few, if any, formal qualifications (e.g. UK GCSEs at grade C or above).

She arranged for them to have a double period off normal timetable to work with the materials. Ten students worked in 3 small groups. Listening to their discussions it was clear that there was a very large difference between their articulation of ideas and what they were recording on paper. Focusing their attention back to what they had said about the data in discussion made a substantial improvement (though it was still less than the discussion revealed of their understanding) suggesting that they see quite different processes in these two environments. For example, when trying to describe the way savings increased over time when invested (exponential growth) one group struggled to articulate what was going on, but one boy was using his hands to describe it, and then said 'it goes sort of whoosh at the end' as he made an upward flourish. He did not think that this was an appropriate description, but was obviously pleased when it was accepted that it communicated exactly what was happening – which is precisely the point of language. The group continued to work through the materials, often using the 'whoosh factor' as part of their discussions. After about 40 minutes that same boy said to the class teacher that he had done enough – "it was making his brain hurt". She allowed him to sit out the rest of the lesson. She apologised at the end, saying that he was a very difficult boy, and that she had been so pleased with what he had done she didn't want to have a confrontation – she didn't remember him ever working on anything for longer for 10 minutes on the same topic before this lesson.

These reports from very different settings provide 'existence proofs' of the power of interactive displays in promoting student engagement with MV data. There is great deal of work yet to be done to explore the extent to which these existence proofs can be generalized across age and attainment, and data sets. We need to understand the relative impact of constructivist approaches, teaching styles, interface designs, and the choice of MV data sets. There is also a need to map the SL associated with MV data, in some detail, and to take steps to improve SL in this new domain. In particular, we need to move beyond vague advocacy of constructivist approaches, to a detailed understanding of exactly what ideas are core to working effectively with MV data. We need to understand which ideas are easy to acquire and are difficult to acquire, and we need to develop robust classroom activities that lead to effective learning.

As part of our efforts at curriculum development, we have begun to assemble a list of core heuristics that are useful when exploring MV data. These include:

- critique the quality and source of the data;
- describe and explore phenomena before you try to explain things;
- focus on effect size not significance level;
- check that the effect size is a lot bigger than the likely error of measurement;
- identify variables that have the strongest effects;
- look for non-linear relationships;
- look for changes over time;
- look for interactions, and think about 'data surfaces';
- disaggregate the data, and see if the patterns of relationships stay the same as in the aggregated data;
- look for the 'dog that didn't bark' – were there things you expected to see, but didn't?;
- be careful to separate analysis and interpretation – especially in observational data.

REVIEW

MV data is being used increasingly to inform political decisions. Major data providers are keen to present their data, and to have it explored and used by a wide range of people, for a variety of purposes. PN sites present huge amounts of data in a very wide variety of displays that are visually attractive. However, many of the display formats are ill-designed to display complex data. PN sites are hard to navigate, and meaningful discussion about MV data is largely absent. One might conclude from the research evidence on acquiring SL that this is entirely predictable – statistical concepts are inherently difficult, and inculcation of SL will prove to be very difficult. However, another strand of evidence suggests that this need not be the case – classroom trials show that young students can reason effectively with MV data, if presented appropriately. This is an exciting prospect, and suggests that we can provide effective support in a movement towards evidence-informed democracy.

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.

Callingham, R., & Watson, J. (2005). Measuring statistical literacy. *Journal of Applied Measurement, 6*(1), 1-29.

Dewey, J. (1916). *Democracy and Education.* The Macmillan Company. Electronic citation accessed 17 Mar. 2008: http://www.ilt.columbia.edu/publications/dewey.html

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

Garfield, J. (1995). How students learn statistics. *Int. Stat. Rev.*, **63**, 25–34.

Garfield, J.,and Ben-Zvi, D. (2007) How Students Learn Statistics Revisited: A Current Review of Research onTeaching and Learning Statistics. *International Statistical Review*, 75(3), 372–396.

Miller, J. (2005). *Writing about Multivariate Analysis*. London: University of Chicago Press.

Office of Public Sector Information (2007). Statistics and Registration Service Act. http://www.opsi.gov.uk/acts/acts2007/ukpga_20070018_en_1

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

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

Ridgway, J., Nicholson, J. R., & McCusker, S. (2007c). Using multivariate data as a focus for multiple curriculum perspectives at secondary level. Paper presented at the *56th Session of the International Statistical Institute*, Lisbon, Portugal. http://www.isi2007.com.pt/isi2007/index.php.

Ridgway, J., Nicholson, J., and McCusker, S. (2008, in press). Reconceptualising 'Statistics' and 'Education'. In Batanero, C. (ed.). *Statistics Education in School Mathematics: Challenges for Teaching and Teacher Education.* Dordrecht: Kluwer.

Tufte, E. R. (1997). *Visual Explanations.* Cheshire, CT: Graphics Press.

Wainer, H. (2000). *Visual Revelations: Graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot.* Mahwah, NJ: Erlbaum.

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

Viégas, F., Golder, S. and Donath, J. (2006). Visualizing email content: portraying relationships from conversational histories. *Proceedings of the SIGCHI conference on Human Factors in computing systems, 979 – 988.* Montréal, Québec, Canada

Beyond GDP (2007) http://www.beyond-gdp.eu/

Snopes http://www.snopes.com/

Many Eyes http://services.alphaworks.ibm.com/manyeyes/home

Swivel http://www.swivel.com