HELPING STUDENTS TO COMMUNICATE STATISTICS BETTER

DAVIES, Neville and CONNOR, Doreen
Royal Statistical Society Centre for Statistical Education
England

In this paper we present examples of improving student communication skills in statistics by using resources that are created from student and teacher involvement in data production. Collecting real data that the students and teachers can relate to, the creation of data-active worksheets, lessons and projects can help to bring statistics to life. This, in turn, can aid the development of thinking skills that enable students to communicate statistics better, both orally and in writing. We give examples that:

- *encourage students to learn statistical communication skills through collecting real data and using them to produce useful projects that can yield meaningful conclusions;*
- *enable teachers to be confident and competent about the types of data that should be collected and used, and how they should be organised and presented;*
- *show teachers how to support students in Data Handling and Statistics projects.*

INTRODUCTION

In the report into post-14 mathematics education in the UK, Smith (2004) repeatedly makes the point that 'Knowledge of Statistics and Data Handling is fundamentally important for all students'. But he also draws attention to a number of issues in relation to the *role* and *positioning* of statistics within the curriculum. For example, the mathematics examination (called the General Certificate of Secondary Education, GCSE) sat by all students aged 16 contains a compulsory Handling Data project component. This contributes 10% to the overall mark awarded for the exam. A number of issues have arisen over this and Smith (2004, p85) comments: 'There is concern that current (project) requirements lead to a rather artificial approach to analysing and interpreting data, rather than encouraging substantive involvement with "real life" problems'.

Mathematics teachers have experienced great difficulty in coping with both teaching the statistics subject material necessary for students to carry out projects associated with real life problems and finding suitable data sets for them to use. These often lead to badly conceived, inappropriately analysed and poorly communicated projects.

REAL DATA PRODUCTION AND HANDLING

There has been much research in the United States into how to best teach and learn statistics, mainly at the introductory level in Higher Education. However there are parallels that resonate with the UK experience of introducing statistics to school students. The American Statistical Association (ASA) and the Mathematical Association of America (MAA) long ago recommended that '*almost any course in statistics can be improved by more emphasis on data and concepts, at the expense of less theory and fewer recipes'*, (Cobb, 1992). Hogg (1992) argues that courses should emphasise components of *statistical thinking* by incorporating more data and concepts, fewer recipe derivations, more automated computations and graphics, using real data, and foster active learning (group problem solving, projects, written and oral presentations). See the discussion in Snee (1993) and the many references therein.

For improving student skills Garfield (1995) extols the virtue of their gaining real world statistical experience from solving real problems with real data. She also summarises educational research views on statistical learning which suggest: more time be spent developing student understanding; use of small group learning activities; open-ended problems; practice sessions and experimental work.

The theme underpinning the previously discussed US research evidence is that teaching statistics should involve students in *data production* (Cobb and Moore, 1997). Amongst other things, these authors critically examine the role of mathematics in teaching statistics, and how statistical and mathematical thinking differ. In the UK the data production principle is implicit in the Qualifications and Curriculum Authority (QCA) recommendation that data handling and statistics, especially for project work, should be taught through the *Handling Data Cycle* which comprise an iterative sequence of four guidelines:

(a) Specify the problem and plan: formulate questions in terms of the data needed, and consider what inferences can be drawn from the data; decide what data to collect (including sample size and data format) and what statistical analysis is needed;
(b) Collect data from a variety of suitable sources, including experiments and surveys, and primary and secondary sources;
(c) Process and represent the data; turn the raw data into usable information that gives insight into the problem;
(d) Interpret and discuss the data: answer the initial question by drawing conclusions from the data.

The steps in this cycle are illustrated in the following schematic diagram (Figure 1) which also provides a useful way to think about teaching the principles behind each step. In addition, the statistical thinking involved in carrying out all four linked activities provides a helpful way to start to communicate the sequence of outcomes that will help to solve the original problem. In fact the cycle closely matches the steps professional statisticians follow in applied statistical work. See, for example, the statistical problem-solving paradigm discussed by Stuart (1995).

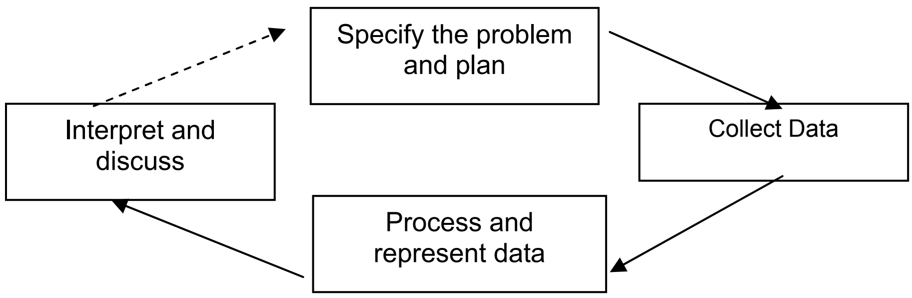


Figure 1. *Schematic representation of the Data Handling Cycle.*

SUPPORTING STUDENTS TO WRITE STATISTICS PROJECTS

Our experience is that students who are taught data handling and statistics through the above cycle of activities are better equipped to *do* and *communicate* statistics, especially when those students need to carry out and write-up their projects. Following adoption of teaching methods that use the activities specified in the Handling Data cycle, in section 5 we provide evidence of improved attitude to data handling, a less fearful approach to project work and more differentiation in examination results. In supporting students to follow this cycle and for designing, writing and communicating statistics projects using real data, the Royal Statistical Society Centre for Statistical Education (RSSCSE, 2003) has produced a booklet *A Toolkit in Data Handling for Projects*. We now summarise some of the advice contained in the booklet.

First, we strongly recommend that teachers remind students that data should always be collected *for a purpose* and that a key goal is to *get information* from the data in order to illuminate the original project undertaken. Second, we emphasis that there are a number of milestones that should be followed *within* the handling data cycle of activities. These are summarised in Table 1.

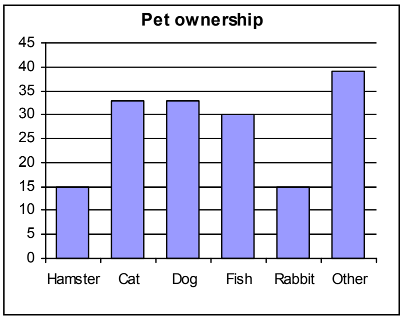
Table 1
Milestones within the Data Handling Cycle

1.1 Specify the problem and plan
• *Decide* on the aim/purpose of the investigation and include a reason for this choice. Try to give a clear hypothesis, for example. 'Children grow more rapidly during ages 7-11 than between 11 and 16.' For higher-level projects you can give a set of inter-related hypotheses. • *Design* Show how you are going to investigate your chosen topic. Include: ▪ how you are going to conduct your investigation (e.g. survey, experiment, secondary data); ▪ what data are needed (what data are relevant; ensure that there are sufficient data); ▪ what method is to be used for collecting the data; ▪ sampling techniques that are clearly described; ▪ assumptions you are making about the data; ▪ where any secondary data used come from.
1.2 Collect data
• *Data* - Decide on an appropriate sample size; give reasons for your decision. Decide exactly what needs to be measured and/or counted and state these. • *Collect* the data from a suitable source or sources and consider the best way of recording them. • *Develop* - A pilot survey can be used to find faults and problems in the questionnaire and harden up ideas on sample size and make up, spot potential pitfalls or particular areas to investigate. Revise your plans in the light of your pilot survey.
1.3 Process and represent data
• *Display* - A wide variety of relevant graphs, charts and statistics is important for getting insight into what the data are telling you. Experiment with these and decide which ones are the best to use for your particular project. Ensure there is no redundancy in representations, or they will be penalised. Try to show your data in the most appropriate way that yield the most information and state reasons for the choice. Ensure graphs and diagrams are clearly labelled. • *Choose* the best representative value for your data (e.g. the mean). Do not be repetitive. Ensure a good range of techniques. • *Details* - Write up your work as you go along rather than leaving it all to the end. Keep brief rough notes of any decisions, discoveries, thoughts, observations or ideas which occur at any time.
1.4 Interpret and discuss data
• Adaptations that you make in the light of a pilot study should be part of your final report. • Make sure that you interpret all diagrams and calculations in the context of the original aim or hypothesis. Relate everything back to your original hypothesis/ideas/purpose. • Some of your results may lead you to "change direction" and investigate a particular aspect further, or to gather more data for comparison. • Only make statements or draw conclusions that your work entitles you to do. Your conclusions should link your results and interpretations back to your original aims and hypothesis. • Write a summary of your results at the end of your project and make suggestions for further investigations. Also acknowledge any limitations due to the type of sample possible, availability of data or other factors. It can be useful to discuss your findings with others. • In your report discuss points for possible further investigation.

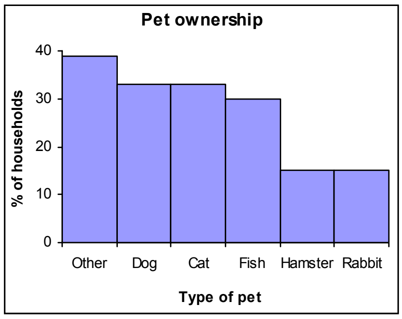
Within the context of 1.1 we are often asked by teachers ‘What types of data should I encourage students to collect/use?’. Rather than deciding these beforehand we stress that it is far more important to specify the problem and plan the project carefully first. For example, starting with a general idea such as ‘I am interested in comparing heights of boys and girls and whether they own pets', the nature of the investigation needs to be made more precise, including the reason for collecting appropriate data. There is also a misconception with some teachers that 'continuous data are better than discrete data' - we have seen examination marks awarded accordingly! What matters is the *scale of measurement* for data. From a *practical* point of view, of course, *all* data are discrete, they are simply measured on different scales. Confusion over 'types' of data often lead to flawed projects and this can be seen from scrutinising the graphs and diagrams presented by students. For example, the data collected about pet ownership would be measured on a discrete scale, but once the responses are recorded in software such as Excel, it is tempting for students and teachers to simply dip into the wonderfully easy way that Excel produces its charts without thinking about the variable and how the responses *should* be displayed. Amongst other things, Excel cannot distinguish between scales of measurement.

In phase 1 (2000-2001) of the *CensusAtSchool* project we collected data about pet ownership. Figure 2 shows six different ways that the responses can be communicated using Excel. Underneath each chart we list errors of presentation that can help teachers and students discuss important general issues of communicating data through graphs and charts. Figure 1 is taken from the RSSCSE (2003) *Toolkit in Data Handling for Projects* and the charts were produced by Neville Hunt of Coventry University.

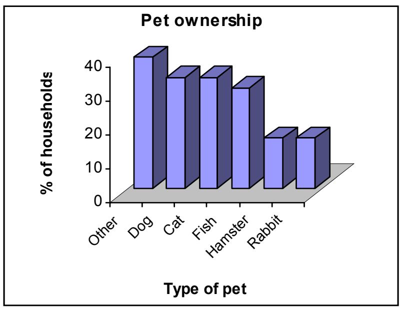
The errors in the charts show that software flaws and limitations must be borne in mind when teaching students how to communicate data. These can help focus what types of charts and graphs *should* be used. They also demonstrate that students should develop a healthy scepticism to automatic 'button-pressing' production of charts and graphs.



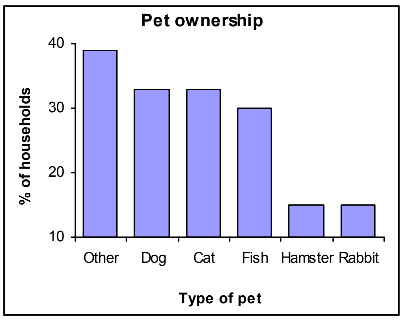
The columns are not in order of size.
Arguably the horizontal axis needs no title.
The vertical axis has no title and too many marked points.



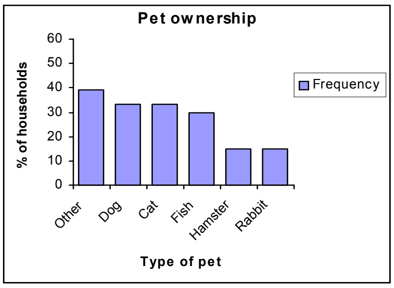
There should be gaps between the columns.



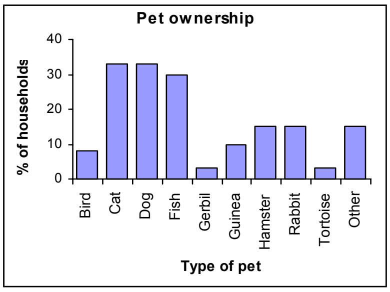
The axis labels are not horizontal.
The 3-D visual effect distorts the picture.



The non-zero vertical origin exaggerates any differences



The vertical axis extends too far.
The axis labels are not horizontal.
The legend is unnecessary, cramping the plot area.



The axis labels are not horizontal.
The columns are not in order of size.
Use horizontal bars rather than vertical labels.

Figure 2. *Problem Bar Charts using some of the Pet data from phase 1 of CensusAtSchool.*

FEEDBACK FROM TEACHERS

Teachers from several countries who have taken part in the *CensusAtSchool* project have commented that when students can relate to data, it can help them communicate better. The following quotes, received from teachers in Canada, New Zealand and Australia during 2004, provide anecdotal evidence.

> "We worked on measurement, data management, graphic displays of data, estimating, and different ways of recording data. It's a lot more fun to use data of a personal nature." (Kimberly Burstall, primary teacher, Halifax, Nova Scotia).
>
> "I thought the whole thing was really well prepared and very user friendly. The students were really positive about the task and naturally love to share stuff about themselves!" (Secondary school teacher, New Zealand)
>
> "The children really enjoyed it - something meaningful about them - great interest from all students completing the questions." (Secondary school teacher, New Zealand)
>
> "They found it very absorbing and some were quite realistically creative about the contexts in which their questions would be relevant" (Secondary school teacher, South Australia)
>
> "My students got more out of this project than any text book or teacher could communicate." (Larry Scanlon, primary-intermediate special education teacher, Waterloo, Ontario).
>
> "I like the idea of large statistical projects such as this to heighten the focus on the purpose of surveys and the associated ways one can present the information and draw out conclusions of the findings." (Secondary school teacher, New Zealand).
>
> "Through this project, students see statistics (and math in general) as a set of conceptual tools that help them better understand the complexity of the world in which they live in". (France Caron, Education Professor, Université du Québec à Montréal, Québec).

The evidence appears to be that if you can motivate the students with the data and its context, then they are stimulated to want to communicate better. This even appears to have a knock-on effect between students and their parents. As Mary Townsend, Education Outreach Manager at Statistics Canada comments: "By participating in an on-line survey at school, students bring a helpful message home to parents and caregivers", and she goes further: "*CensusAtSchool* allows learning by discovery with real world questions and an on-going and iterative outreach to schools; it not only supports math and technology curricular, it also builds a teacher community of interest that is continuously connected to Statistics Canada". She feels that linking a government statistical service with a country-wide statistical educational project will benefit everyone.

Finally, the following comment, from Trevor Cole, Head teacher of a primary school in South Wales, indicates his primary students improved over a wide range of skills:

> "The *CensusAtSchool* project is a springboard for enriched data handling experiences and it encourages pupils' thinking skills; it is a lot easier (for them) to interpret tables, graphs, charts; question accuracy and to develop a critical questioning approach, even to question construction. We used the data to bring about change!"

ACHIEVING BETTER COMMUNICATION SKILLS

We now present two examples that demonstrate improved writing skills for statistics projects after students produced and used data in line with the discussion in section 2. These examples were presented at the December 2004 meeting of the General Applications Section of the Royal Statistical Society (Davies et al., 2004).

In our first example, Emma Knights, Curriculum Team Leader of Mathematics and ICT at Upper Avon School, Wiltshire, UK, has developed a strategy for teaching data handling and statistics that involves the whole school in the *CensusAtSchool* project. This takes place over several years (years 7 – 9) before doing the Handling Data project, compulsory for all UK 16-year olds (in year 11) as part of their mathematics examination. The approach is seen as a multi-stage process to prepare the students for this all-important examination activity. For example, in year 8 the current years' *CensusAtSchool* questionnaire is completed in the first term and subsequently, in the summer term, the same students use their *own* data and compare themselves with students in other countries using the databases on the *CensusAtSchool* web site. Citizenship issues, such as reasons things might be different in other countries, are always addressed.

In year 9 the strategy is to develop *investigational coursework* to 'skill-up' the students so that they can tackle a full-blown data handling project in year 10. Data handling skills are developed using the *CensusAtSchool* materials through curriculum activities or (their own) raw data. The emphasis is on interpretation and, by using their own data they *want to* communicate their results effectively. In years 10 and 11 the school encourages students to expand their horizons and consider problems and projects that interest *them*. They take this approach because they are used to working with data that are relevant and real. Some students still choose to use the *CensusAtSchool* data and resources they have already met, but others develop their own ideas with confidence gained from work in years 8 and 9.

The iterative development approach to data handling and statistics has allowed the school to acquire a wide range of data in different contexts. The school utilises their own students as a resource, with minimum of effort, and produces a wide range of activities and ideas for teachers. As a result of all these activities and resources, the students are more confident in 'risk taking' in their approach to data handling and the projects have improved dramatically.

For our second example Deborah Stanley, currently Assistant Head Teacher and Science College Coordinator at Rainham Mark Grammar School, Kent, reports successful results from a new strategy for learning data handling that she implemented when Head of Mathematics at Simon Langton Girls' Grammar School, Canterbury, from 2002. The initial strategy was to develop numeracy across the curriculum for years 7 – 9. The subjects involved were not only mathematics, but science, geography, history and food technology as well. The main difference in classroom practice was that all classes took part in the *CensusAtSchool* project and used its resources extensively. This fundamentally changed how data handling was taught across years 7 - 9 and subsequently in years 10 and 11. Pupil motivation was increased and the use of proven *CensusAtSchool* worksheets saved time and effort. They also turned out to be ideal for staff not trained in statistics.

In years 10 and 11 many issues relating to data handling coursework were solved through the real data collected from the project. For example, it was found that *sampling* the data became meaningful and students were able to focus on interpretation and communication. The written projects were much better than they had been in previous years. There may well be positive knock-on effects in differentiating examination performance of students that have been exposed to the new strategy for data handling and statistics. In Table 2 we present the percentage achieving GCSE mathematics grades A* - C by five cohorts of students who sat examinations from 2000 to 2004 inclusive.

Table 2
GCSE Mathematics Grades at Simon Langton Girls Grammar School, 2000 -2004

Year	*A**	*A*	*B*	*C*
2000	15.0%	35.0%	35.0%	15.0%
2001	14.3%	41.8%	34.7%	9.2%
2002	16.2%	38.4%	34.3%	11.1%
2003	21.4%	27.2%	42.7%	8.7%
2004	24.0%	30.0%	40.0%	6.0%

2003 and 2004 were the first two years for which the students had been taught data handling and statistics using a strategy involving the *CensusAtSchool* project. It appears that more differentiation between grades is evident in 2003 and 2004. For example, an extra 5-8 percent of students achieved the highest grade (A*) and an extra 6-8 percent got grade B, but there was a drop in percentage achieving grade A.

CONCLUSIONS

We have attempted to show that improvements in the way students communicate statistics can be achieved by teaching the subject through the use of real data that the students can relate to. When this approach is coupled with the statistical problem solving paradigm implicit in the data handling cycle, a powerful combination can be made to motivate students to think and communicate statistically.

Early evidence from teachers that have incorporated the real data approach into a whole school strategy for teaching data handling and statistics is that the students are more data-aware and show improvement in writing-up statistics projects. Also there is much anecdotal evidence from teachers in other countries that indicate the approach we espouse seems to motivate students of statistics to be more comfortable at communicating the subject.

REFERENCES

Cobb, G.W. (1992), Teaching Statistics. In: L.A. Steen (Ed.), *Heeding the Call for Change*, MAA Notes No. 22, 3-34. Washington: Mathematical Association of America.

Cobb, G.W., & Moore, D.S. (1997). Mathematics, Statistics and Teaching. *American Mathematical Monthly, 104*, 801 – 823.

Davies, N., Connor, D., Knights, E., & Stanley, D. (2004). The *AtSchool* Experience. December 2004 Meeting of the General Applications Section of the Royal Statistical Society.

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

Hogg, R.V. (1992). Report of workshop on Statistics Education. In: L.A. Steen (Ed.) *Heeding the call for change*, MAA Notes No 22, 33-43. Washington: Mathematical Association of America.

Royal Statistical Society Centre for Statistical Education (2003). *A Toolkit in Data Handling for Projects.* Nottingham: Nottingham Trent University.

Smith, A.F.M. (2004). *Making Mathematics Count.* Report of the Inquiry into Post-14 Mathematics Education. London: The Stationery Office Ltd. Available at http://www.mathsinquiry.org.uk/report/.

Snee, R.D. (1993). What's missing in statistical education? *The American Statistician*, *47,* 149-154.

Stuart, M (1995). Changing the teaching of statistics. *The Statistician, 44*, 45-54.