

STATISTICS ASSESSMENT: THE GOOD, THE BAD, AND THE UGLY

Sean McCusker, James Nicholson and Jim Ridgway
 SMART Centre, Durham University, United Kingdom
 sean.mccusker@durham.ac.uk

Assessment tasks and scoring schemes convey information to students and teachers about the nature of each discipline, and what is valued. It is important that tasks provoke desirable classroom practices, and motivate students to continue to engage with the discipline. Tasks from high-stakes national examinations are presented that are likely to have the opposite effect; Most are 'toy' problems where a technical exercise is presented in a context which is uninteresting and oversimplified. In contrast, we show tasks that engage students in reasoning with authentic data from large scale surveys, and that require statistical reasoning to challenge assertions about matters of fact or plausible courses of action. We believe that better tasks are an important element in addressing Hand's (2009) concerns about the public image of statistics.

INTRODUCTION

High stakes assessment has a very strong effect on what happens in the classroom (Ridgway, McCusker & Pead, 2004; Burkhardt, 2009). The availability of past exam papers allows teachers to align their lessons with the exam papers which they might reasonably expect pupils to face in the future. As such, for many children in England and Northern Ireland in the 14-16 years age group <http://www.educationaldesigner.org/ed/volume1/issue3/article9/>, their experience of statistics or statistical methods, is heavily influenced by the content of the examinations taken at the end of compulsory education - GCSEs. These experiences will shape pupils' ideas about the nature of statistics and the kind of problems which it can solve, both in terms of context and complexity of information. A review of a range of GCSE statistics papers and the statistics content of GCSE maths papers from 2008, the most recent year where a full set of papers is available, has identified a large number of contexts which are at best mildly interesting and at worst obscure or tedious - and sometimes both. Such tasks do nothing to teach pupil about the power of statistics to solve real world problems.

The mathematics curriculum is compulsory at GCSE. It includes a strand dealing with both probability and statistics. In this paper, the term 'statistics' will be used to include probability based items. For most children the only statistics they will encounter as part of their education will be the statistics items in the GCSE maths papers (a small proportion of children also take GCSE Statistics as a separate qualification).

EXEMPLIFYING THE PROBLEM

Current Statistics items, both on the GCSE Statistics and within the GCSE maths papers can usually be grouped into one of 4 categories.

Real world but irrelevant or uninteresting	
	<p>A food processing plant manufactures pots of organic yoghurt. The label on each pot reads 'weight 150 g'.</p> <p>(a) Give one reason why the company does not check the weight of all pots produced. (b) To check the weights of the pots, samples of five are selected at random from the production process.</p> <p>The mean weight of each sample is recorded. The company requires that the sample mean weights must fall within the range $150\text{g} \pm 0.8\%$</p>

AQA GCSE Statistics Higher Tier 3311/H June 2008

Real world but not age-appropriate

The table shows the Price Index Numbers for the average monthly rental of two-bedroomed properties in Manchester between April and August.

The base month is April.

Average monthly rental of two-bedroomed properties in Manchester

Month	April	May	June	July	August
Price Index	100	97	110	70	96

(a) In which month was the average monthly rental higher than it was in April?

The average monthly rental was £720 in April.

(b) Calculate the average monthly rental in July.

...

Edexcel GCSE Statistics Higher Tier Paper 1H 1389/1H June 2008

Context irrelevant

Clare is reading a poetry book.

(a) The shortest poem in the book has one verse and the longest poem has six verses. The poems in the book are from ten to fifteen lines in length. Design a two-way table for Clare to record the number of verses and the number of lines for the poems in the book. [3]

(b) Clare counted the number of letters in each word of one poem. There were 28 words in the poem. She drew this table to show her results.

Number of letters	Frequency	
1	2	
2	5	
3	6	
4	8	
5	2	
6	3	
7	2	
Total	28	

Calculate the mean number of letters per word in the poem.

...

OCR GCSE Mathematics Syllabus A Paper 2 (Foundation Tier) June 2008 J512/02

Unrealistic context

There are 8 pencils in a pencil case.

1 pencil is red.

4 pencils are blue.

The rest are black.

A pencil is taken at random from the pencil case.

Write down the probability that the pencil is black.

Edexcel GCSE Mathematics B (Modular) – 2544 / Mathematics (Modular) – 2381
 Foundation Tier June 2008 –5542F/8B 5381F/5B

There are other items which, plausibly, are more interesting to 14-16 year olds, such as Bats and their diseases, birds and their flight speeds or Opinions on the European Union.

STUDENT INTERESTS

Determining the interests of 16-year olds can be much like star gazing; by the time the light reaches us, the star is long gone. Nevertheless examination of popular media provide some hints.

Of the most popular books on Amazon in the 12-16 year old range, released in the 90 days previous to 26/2/10, seven of the top ten were vampire stories; the remaining 3 were from the Percy Jackson series - the adventures of a 12 year old boy who discovers he is a demi-god. The category of Hobbies and Interests is dominated by exam revision guides, to the exclusion of just about all else. Something can be inferred about this domain by grouping the 3927 books into categories. Science dominates, with 40% of the titles, but perhaps unsurprisingly these are predominantly curriculum or exam guides. The next 40% is made up from three categories—Arts & Crafts, Sport and Animals & Pets.

Analysis of the most popular stories on the BBC Children's News website in February 2010 shows the greatest interest has been in celebrities, disasters and in the peculiar (e.g., 'Band backs Cheryl after split', 'In Pictures: Chile Earthquake', 'Dinosaur-eating snake discovered').

Development of test items (many of which were statistical in nature) for the World Class Arena involved extensive classroom trialling. Informal observations showed that children are particularly interested in the strange and the gross. They also showed a preference for problems which are real and relevant, if not to themselves then at least to someone. Carlson (1998) makes similar observations.

High stakes assessment drives the curriculum. In an attempt to avoid possible negative consequences, Birenbaum et al. (2006) argue that pupil assessment should be derived from valuable curriculum activities. One virtue of this approach is that pupils can be assessed on their performance on extended tasks, rather than on short examination tasks where it is difficult to assess much of the spectrum of statistical reasoning. Their Learning Integrated Assessment System is a model bridging the space between summative assessment and formative assessment. They highlight the importance of assessing process as well as product. Here, we offer a sketch of how this might be done in the context of statistics education. Students are presented with authentic data on topics of direct interest, and are asked to draw conclusions that can have important consequences for both individuals and for social policy.

STATISTICAL UNDERSTANDING IN STUDENTS' BACK YARDS

In his Presidential Address to the Royal Statistical Society, Hand (2009) argues that statistics as a subject area needs to improve its poor public image, and to do this, must show that it has the power to solve interesting, relevant and realistic problems. Few 16 year olds have had experience of statistical solutions to important problems, and few know the kinds of questions which statistics might help to answer. John Tukey said "The best thing about being a statistician is that you get to play in everyone's back yard" (in Hand, 2009). It is reasonable to look for contexts and questions that are relevant to the back yards of 16 year olds.

The European School Survey Project on Alcohol and Other Drugs (<http://www.espad.org/>) shows that, in the UK 70% of students that turn 16 years old during the calendar year of the data collection, had drunk alcohol in the 30 days previous to the survey. Within the last year, over 50% had had a heavy drinking episode and over 50% had been drunk to the extent of staggering or being unable to speak properly. The same survey showed that 30% of 16 year olds had tried cannabis at least once. Recent surveys (BBC) have shown that nearly a 1/3 of teenagers report having lost their virginity below the legal age of consent, which in the UK is 16 years. Amongst 16 year olds, sex, drugs and alcohol are likely to be major topics of interest. It seems perverse to ask them about pencil cases and the weight of yoghurt.

As well as issues of context, there are important issues around the complexity of data and the sophistication of the reasoning required. A review by Ridgway, Nicholson and Mccusker (2007) looked at the content of the statistics papers used for university admissions (A-level Mathematics). Questions involved mostly uni-variate data, and on the few occasions where bi-

variate data was encountered, the relationships were linear. However, to imagine that most of the interesting issues in the world are uni-variate or bi-variate in nature, and can be solved by the simple statistical models encountered in the UK syllabus is to be naive. The major issues of relevance to youth today are multivariate in their nature. For example the extent to which one is exposed to a Sexually Transmitted Disease (STD) is dependent on its prevalence in the age group and socio-economic status of one's sexual partner. Any attempts to deal with STD as an issue must take into account the way in which it has changed over time. (see Nicholson, Ridgway & McCusker, 2010 for a fuller discussion). Analysis of some 2009 GCSE statistics papers is more encouraging; there are some instances of non-linear bi-variate relationships, some set in reasonably interesting contexts.

Interesting context; realistic modeling

The scatter diagram shows the average wingspan and the average weight of 8 British birds.

Bird	Wingspan (x cm)	Weight (y g)
Wren	~10	~10
Robin	~15	~20
Blackbird	~25	~50
Pigeon	~60	~200
Duck	~100	~1000
Goose	~160	~3500
Eagle	~210	~4000
Swan	~230	~9000

(Source: British Trust for Ornithology)

Put a cross in the box to indicate which one of the equations is the best model for these data.

$y = \frac{a}{x}$	$y = a\sqrt{x}$	$y = ax^2$
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The scatter diagram below shows the body lengths and wingspans of the same birds.
 ...
 The scatter diagram shows positive correlation.
 (b) What can you conclude about the relationship between the body length of a bird and its wingspan?

A line of best fit has been drawn on the scatter diagram.
 The equation for this line of best fit is of the form $y = ax + b$
 (c) Work out the equation for this line of best fit.

A crow has a wingspan of 100 cm.
 (d) Estimate the body length of this crow.

Edexcel GCSE Statistics Paper 1H Higher Tier1389/1H

There is much in the press which seeks to demonise youth, presenting them in a negative light, or pronouncing on issues which are relevant to them. Some recent headlines from UK mainstream media refer to *Labour's maths bungling leaves kids on scrapheap*; *Shock hike in kid coke addicts* (The Sun–Online), *Alcohol and cigarettes are more harmful than Ecstasy and LSD, says drugs tsar*; *Booze Britain: hard evidence links alcohol and youth crime* (Daily Mail – Online).

They can't read, can't write, keep time or be tidy: Tesco director's verdict on school-leavers (Guardian - Online) *Violent games 'worse' than porn* (BBC News Website).

These headlines are illustrative of those which the pupils in the 14-16 age group might meet everyday in the mainstream press, where ill-founded speculation is presented as fact. It is possible to create curriculum activities where pupils deconstruct media myths about adolescents. This has the potential to show the power of statistics. An example is presented in Figure 1 and Figure 2, that shows a mashup (A webpage with multiple links to information sources) comprising recent survey data on alcohol use, presented in an interactive display, with links to recent newspaper articles on alcohol consumption by young people.

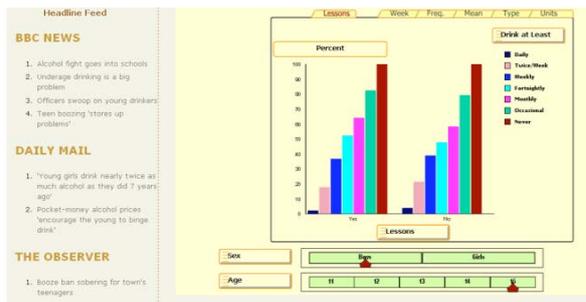


Figure 1.

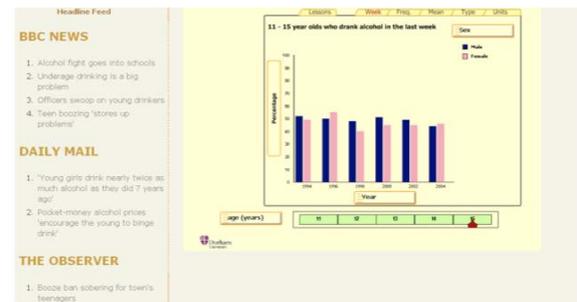


Figure 2.

Students worked with the task for three separate one-hour periods, starting with an introductory piece focused on smoking, which allowed pupils to familiarise themselves with the software. The main objective was for the pupils to create a media report on the topic of Alcohol and Young People.

The structure for student outputs was very open. Pupils were not given strong guidance on the content of what they were required to produce or how to analyse or interpret the data presented to them. They were encouraged to explore the domain and report on their findings, with respect to the news articles. These reports took a variety of forms such as letters to the editor, newspaper articles, interviews and TV news reports. Students were encouraged to produce a report in any format with which they felt comfortable. In total about 100 reports were produced and analysed.

Student responses were categorised primarily in terms of: use of argument (the ways in which students use data to critique or create text media, and the extent to which they synthesise information from different media accounts); and use of empirical data (judgments on the extent to which students have made use of the data in the interactive displays).

A high proportion (> 80%) of the children were enthusiastic about the task. Reports were well presented, with a good sense of audience and structure: only four reports failed to meet all of these criteria adequately.

- Over 80% of the reports used data, with about 60% using it accurately and appropriately to critique the media articles or in the creation of their own articles.
- Just over 20% of the reports described trends in a clear and accurate manner, and a further 10% described trends with at least some degree of success.
- About 15% of the reports made mention of 2-way interactions: this is interesting, because the topic of interaction is usually considered to be too complex at age 16 years. In fact such analyses are not required even within the A Level Mathematics and Statistics curricula in England, designed for mathematically able 18 year olds.

DISCUSSION

Pupils' worlds are filled with complex and interesting problems. Statistical reasoning is a powerful tool which is able to help make sense of these worlds. Pupils need to see the applicability of statistical techniques and how these can relate to their environment and help them to understand the issues which engage and threaten their interests. If we lead children to believe that statistical

reasoning consists of set of techniques which allow them to determine confidence intervals around the volume in pots of yogurt, they might struggle to see how this set of skills might reveal to them the threat of catching an STD, help them to counter the bad press which they receive or reveal to them the fact that the behaviour of their parents' generation was much worse than theirs.

REFERENCES

- Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., Wiesemes, R., & Nickmans, G. (2006). A Learning Integrated Assessment System. *Educational Research Review*, 1(1), pp 61-67.
- Burkhardt, H. (2009). On Strategic Design. *Educational Designer*, 1(3). Online: www.educationaldesigner.org/ed/volume1/issue3/article9/.
- Carlson, W. (1999). "A Case Method for Teaching Statistics". *Journal of Economic Education*, 30(1), 52-58.
- Nicholson, J., Ridgway, J., & Mccusker, S. (2010). Luring non-Quantitative Majors into Advanced Statistical Reasoning (and luring statistics educators into real statistics). *Proceedings of ICOTS8*.
- Ridgway, J, McCusker, S., & Pead, D (2004). *Literature review of e-assessment*, NESTA Futurelab Series Report 10, Bristol: NESTA Futurelab.
- Ridgway, J., Nicholson, J. R., & McCusker, S. (2007). Teaching statistics – despite its applications. *Teaching Statistics*, 29(2), 44-48.
- Hand, D. J. (2009). Modern statistics: the myth and the magic? *Journal of the Royal Statistical Society*, 172, 287-306.

Is she better or worse off? Some might argue that the answer is entirely a matter of opinion, and that one cannot address it objectively. The three premises of standard welfare economics are powerful because they allow us to replace this vexing normative question with a positive one: would Norma choose a salad over a sandwich? The occasional objection to a direct judgment entails nothing more than a difference of opinion between the analyst and the consumer as to what constitutes a good or fulfilling life. Statistics are worse than raw numbers. They reduce numbers into an even smaller quantity of numbers. A mean is a single number that may represent thousands of individual measurements. There is no possible way that statistics can fail to lose some of the information contained in the original measurements. Statistics trap us into analysis paralysis. No amount of statistical analysis can ever produce certainty. Some people have real problems dealing with this. Instead of accepting the uncertainty and acting anyway, they gather more data or apply more analysis to the data they already have. Statistics are confusing. Send me examples of good, bad, and ugly statistical usage. I'd really like to hear about them. bad, or ugly colormaps often relies on these values, partly because. meaningful experiments can be difficult to design [81] or implement. One possible reason for the difficulty in assessing and defining the characteristics that make an effective colormap is the lack of a common framework. Wainer and Francolini stressed the importance of a common language decades ago: "For if we do not have a vocabulary with which to discuss graphic concepts, we cannot discuss these concepts in." efforts, and attempts to automate the assessment, choice, or generation of colormaps. We hypothesize that the goal of creating a cross-discipline. The Good, the Bad, and the Ugly: Rethinking Election Monitoring. International election monitoring has become a prominent tool for promoting election integrity and democracy (Beigbeder 1994; Abbink and Hesseling 2000; Kelley 2008a).¹ But several factors raise questions about the validity and effectiveness of international election monitoring. When organizations have to return to the same countries election after election, what are they accomplishing? Figure 3 shows the assessment of the monitoring organizations in the elections they have monitored between 1975-2004. The figure shows the strictest assessment.