

**UNCOVERING AND DEVELOPING STUDENT STATISTICAL COMPETENCES  
VIA NEW INTERFACES**

**Jim Ridgway and Sean McCusker  
University of Durham  
UK**

**James Nicholson  
Belfast Royal Academy  
UK**

*Statistics ... the most important science in the whole world: for upon it depends the practical application of every other science and of every art; the one science essential to all political and social administration, all education...* Florence Nightingale

The paper reviews the nature of statistics in the UK National Curriculum. While there is great opportunity for statistics to inform thinking in a range of disciplines, there is little coherence in the planning of activities for children. Formal assessment of statistical ability focuses on procedural knowledge, applied to univariate and bivariate problems. These provisions are inadequate for informed citizenship. We report work on World Class Tests, where students aged 9 and 13 years are presented with both paper-based and IT-based problem solving tasks in mathematics and science. Many of these tasks require students to work with 3 or more variables, often non-linearly related to each other. Students perform rather well on these tasks. Tasks are shown, together with performance data disaggregated by school and sex. Features of the displays that make them well suited to the exploration of complex patterns are discussed, and our design principles are described. Our current activities set out to embed test items in curriculum materials suitable for use in mathematics and science classrooms. A good grounding in handling complex data is essential for informed citizenship; we are in the process of developing powerful methods to help students develop appropriate skills.

**The Nature of Statistics in the English National Curriculum**

*Statistical thinking will one day be as necessary for efficient citizenship as the ability to*

*read and write.* H.G. Wells

The English National Curriculum sets out programmes of study for every curriculum subject for students aged 5 to 16 years. Statistics occurs as a topic for study under Data Handling, one of 4 strands in Mathematics (the other strands are: Using and Applying Mathematics; Number and Algebra; Shape, Space and Measure). There are rich opportunities across the curriculum to develop an understanding of complex issues with the help of statistical techniques (notably in Science; Geography; History; Personal, Social and Health Education; and Citizenship). Unfortunately, the National Curriculum was written subject by subject with no attempt to produce coherence across subjects; as a result, the same topic in statistics is to be presented to students of different ages in different subjects, and some important topics do not appear at all (see Holmes (2000) for a detailed analysis). Another consequence is that teachers with no background in statistics education are making decisions on which concepts to introduce to students, with little appreciation of the inherent difficulties of those concepts. Advanced-level subjects are available for students aged 16 to 18 years. Here students choose a range of more specialised courses. Mathematics is one of the three most popular subjects (along with English and General Studies). Data for the year 2003 from the UK's largest Examinations Authority show that more than 85% of students took at least one statistics module, and that more than 50% of students took 2 statistics modules. We set out to analyse the content of these statistics modules. Two raters independently analysed the examination papers used to assess statistical knowledge by each of the six Examination Authorities at AS level, the first level of advanced study. Raters were asked to categorise the questions according to whether they assessed computational or procedural knowledge, or could be broadly classified as interpretation. This category included: identifying the assumptions which underlie certain models (e.g. use of the Poisson distribution); making a choice about which probability distribution to use in a particular context; interpreting the conclusions of an hypothesis test in context; and all reasoning from data. There was excellent inter-rater

agreement; on no paper did their judgements on the proportion of marks in each category differ by more than a single mark. Differences were resolved successfully by discussion. The proportion of marks awarded for the broad interpretative categorisation on the examination papers were: 3%; 12%; 22%; 25%; 25% and 28%. Raters were also asked to identify questions where students were required to work with 3 or more variables. Not a single example was found on any examination paper.

Putting the national curriculum analysis together with the detailed study of statistics examinations, an unsatisfactory pattern emerges. Statistical education focuses on a rather narrow range of techniques applicable only to univariate and bivariate analyses; student exposure to practical applications of statistical ideas in the curriculum is poorly specified. It is likely that students in some programmes will enjoy a curriculum rich in statistical ideas, while the majority will not. When the role of education to prepare students for informed citizenship is considered, one is hard pressed to show where in the National Curriculum students acquire the skills necessary to understand the data presented in government information intended to explain new evidence-informed policies (such as the recent document on new legislation on crime <http://www.homeoffice.gov.uk/justice/sentencing/correctional/reducingcrime-changinglives.html>). We suspect that neither Florence Nightingale nor H.G. Wells would be happy with our current situation.

### **Addressing the Problem**

*Everything should be made as simple as possible, but no simpler'. Albert Einstein*

For the democratic process to work effectively in a complex world, citizens need to be able to understand arguments based on complex evidence that uses both quantitative and qualitative data. This places an important duty of care upon educators world wide. There is extensive research which highlights the statistical misconceptions of adults (e.g. Tversky and Kahneman, 1983) and young students (e.g. Batanero, Godino, Vallecillos, Green, and Holmes,

1994) which need to be overcome. Much of what is taught in schools focuses on univariate and bivariate analyses and on linear relationships. Such knowledge provides a useful grounding for statistical knowledge, but has a rather limited application in the face of the data encountered outside school settings. Technologies (IT) are emerging which make it possible to present complex data in new ways, and which give users considerable control over the ways they interact with data sets. Here, we consider some results from large-scale trials where IT has been used to present exciting challenges to students, based on complex data sets. IT reduces the time spent doing calculations and manipulations which characterise much of the school statistics curriculum. However, greater apparent procedural competence may mask a lack of understanding of underlying concepts. Experience in classrooms (e.g. Nicholson, Mulhern, and Hunt, 2002) suggests that increased automation has reduced ‘contact’ with the actual data. As a consequence, a grasp of the meaning of the data and the sensibleness (or otherwise) of possible relationships is not well formed. Many software packages do not prevent inappropriate graphs being drawn, unjustified statistical procedures being applied to data, nor do they provide either the depth of inferential insight or the background knowledge (i.e. context specific specialist knowledge) needed to make reasonable interpretations of the analysis. Greer (2000) observes that access to data is increasing dramatically, but human skills in data handling lag far behind. Hawkins (1997) observed that ‘Having the vision to see what technology can, or might, do is not synonymous with knowing how to take advantage of this in a teaching context.’

The use of technology in teaching statistics has a number of inter-related strands:

- issues regarding the use and usability of software packages;
- new opportunities to present ideas dynamically and interactively to students, rather than in more conventional, ‘static’ environments;

- the need to develop robust models of new sorts of conceptual learning and understanding which will underpin teaching in IT-rich environments.

One of the challenges to the introduction of statistics coursework into GCSE mathematics is that pupils are being expected to deal with complex data sets, where multiple variables are to be considered (Nicholson, 2003). There has been little tradition of data interpretation within the mathematics curriculum. The main emphasis has been on the mechanical skills of constructing specified graphs correctly, or ‘reading the graph’ to extract detailed information. We have little subject-specific pedagogic knowledge to guide our teaching – such as a sensible hierarchical structure of statistical concepts which allows us to develop interpretative skills in pupils effectively. Exposure to large numbers of data sets and graphs that have something to say which is both accessible and of substance offers a way forward. The advent of new technology makes this approach practical, if the teaching resources are available when the hardware becomes available in classrooms. A further challenge is that IT can actually facilitate the development of new cognitive processes; we need to develop a pedagogy which capitalises on the new affordances of IT (see Ridgway and McCusker, 2003).

### **World Class Tests**

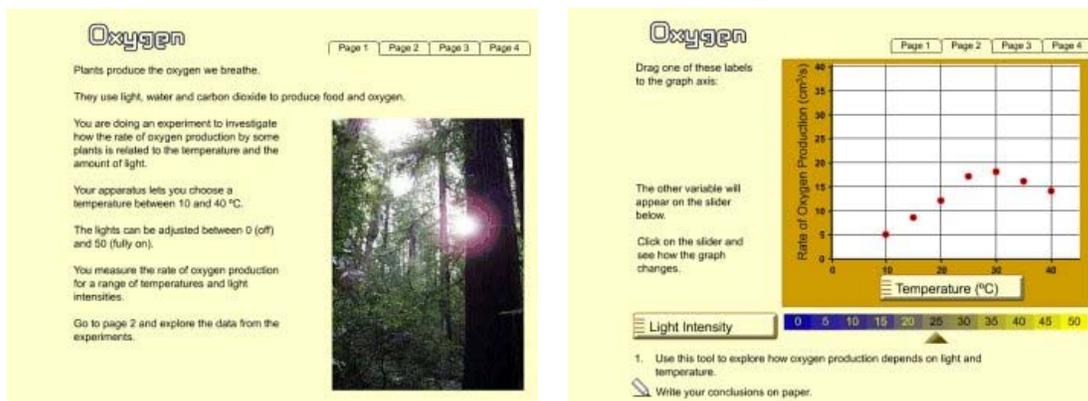
A large scale project funded by the UK Department for Education and Skills set out to assess the problem solving skills of high attaining students in mathematics, science and technology. The MARS Group, based at Durham and Nottingham, has been responsible for the development of the problem solving tests. Much of the testing has been computer-based. Our focus has been on students aged 9 and 13 years; so far, over 10,000 students in 20 countries have taken our tests. Each test has a paper component and a computer-based component. The computer-based component comprises 5 tasks, and students have 60 minutes (age 9 years) or 75 minutes (age 13 years) to work on these tasks. Because of the focus on

mathematics and science, reasoning from data has provided the basis for many of the tasks.

Our central concern is to assess problem solving ability, not curriculum knowledge.

A challenge for task designers is to create tasks which are intellectually rich, using interfaces which pose no problems during the testing sessions. A number of new task ‘types’ have been developed. Just 4 interfaces will be shown in this paper, to illustrate the scope for developing new interfaces which are easy to use, and which encourage an enthusiastic engagement with data exploration and data analysis. Three assessment tasks are shown below and one curriculum task is shown later. Scoring rubrics for the assessment tasks are included in the Appendix.

### Oxygen:



**Figure 1: Screens from the Oxygen Task (age 13 years)**

**Oxygen** provides data on oxygen production by plants as a function of light level (roughly linearly related to oxygen production within the range of variables used) and temperature (which shows a curvilinear relationship). At age 9 years, students are shown a graph of ‘plant growth’ as a function of temperature, and have ‘light level’ as a slider. At age 13 years, we refer to oxygen production, ‘light intensity’ and use a denser data set. Later pages in the task allows students to choose the axes to be plotted (light level can be plotted, and temperature used as a slider) and to use the slider to control the setting of the other variable.

Students must engage actively with the display, and make decisions about exactly which aspects of the data they wish to see. The task involves three variables within a real world setting. The data used within the task is realistic, though somewhat 'cleaner' than real data, and the task is placed in a context such that students perceive that the exercise and results may be of use to someone (notably horticulturalists).

Students are required to explore linear and polynomial relationships within a data set of three variables. Having understood the nature of the relationships, they are expected to find variable settings that produce the highest oxygen yield. A number of hypotheses are presented to the students, which are partially true, and partially flawed. Students are asked to use their understanding of the data to evaluate and reconcile them.

An interesting informal observation from watching large numbers of students working with the task (and during development trials) was the extent to which students appeared to code the data enactively. If asked to describe the data - and often spontaneously when asked for explanations (or when observed solving the problem on their own) - students would move their hands and arms in shapes which mapped out the 3D data surface which could be deduced from the sequential analysis of all the 2D data plots. Students used the kinetic code seemingly easily and unconsciously – they saw no need to explain what they were doing – the listener was expected to understand the data surface being sketched in the air immediately.

## Chortlers:



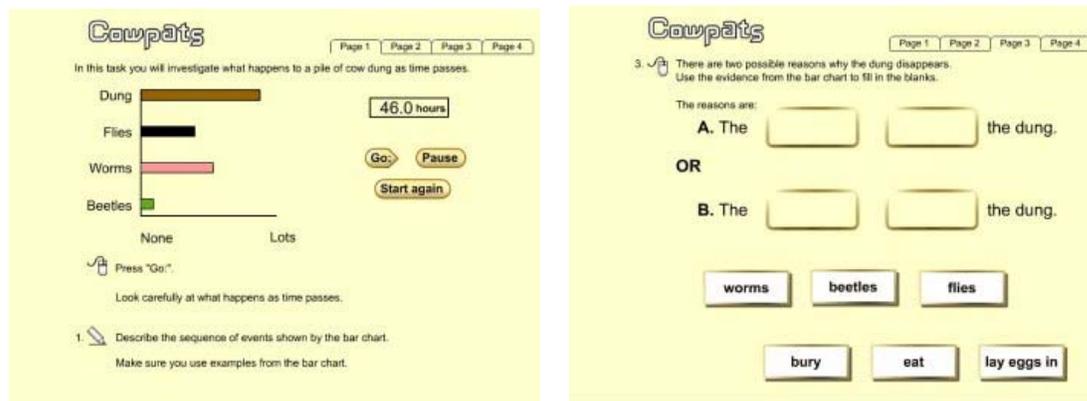
**Figure 2: Screens from Chortlers Task**

**Chortlers** provides data on the association between a bird population and 3 possible factors that might be associated with declining numbers. This task involves students in an active process in which a number of choices about data presentation need to be made. Students can choose to sort, group or graph the data. Within each of these choices there are further options of choosing which variables to use as a primary discriminator, and which variables to plot on a graph. Data can be explored in any way and in any order: students are allowed to investigate as many data representations as they wish.

Students are required to explore a data set comprised of 4 variables. The data has been invented, however may be described as realistic in that they tell a credible story about factors influencing the decrease in a fictional bird population. Within this context, the task of exploring possible plausible factors can be seen as useful and productive.

The task requires students to identify any factors associated with the decline in the bird population, and to provide appropriate evidence, from the data, as justification for their conjectures.

## Cowpats:



**Figure 3: Screens from Cowpats Task**

**Cowpats** presents data in a dynamic display of four variables. It asks students to reason about the causes of the removal of cowpats. The task shows the growth and decline in the populations of different creatures (worms, flies and beetles) that might be the cause of the removal of the cowpat, along with data on the size of the cowpat itself.

The task is based on real world data and the relevance of the process and results of the task are apparent. Students are asked to describe the data in general terms and then perform a sequencing task in which they are asked to place 'snapshots' of the dynamic display, in order. Students are asked to deduce the cause of the disappearance of cowpats, based on the dynamic data and to suggest 2 hypotheses as to the cause of its disappearance. To achieve this, students must be able to identify plausible causal links by looking at the timing of different events in the data. Students are also required to apply what they have learned from the data set, to a real-world setting (the long life of cowpats on concrete) and from this to provide a better hypothesis than the one offered. Overall the student needs to view and understand the data, and from this conceive of a model that accounts for the phenomenon, then apply that model to a real world setting.

Each of these three tasks was designed, then critiqued by an expert group, and refined in the light of small-scale trials with students. Among the criteria applied at these small-scale

trials were that the task should be interesting, challenging and easy to engage with. A modest study (Richardson, Baird, Ridgway, Ripley, Shorrocks-Taylor and Swan, 2002) presented a set of computer-based tasks to individual students, and interviewed them on a number of task features. Students reported that they were immediately interested in the tasks; they experienced no problems with the computer interfaces; and were happy to discuss the advantages of IT presentation over paper presentation. During the trialling process, students were asked about any interface problems they experienced. These were very rare, and students reported that they enjoyed [sic] doing the tests.

### **Performance on the Tasks**

Tests were administered to hundreds of volunteer students who were entered for ‘World Class Tests’ in a number of test sittings. The sample is ill-defined, but entry was welcomed from students judged to be in the top 20% of the ability range. The data from these large-scale trials show that students can make considerable progress on all of these tasks. These results show that young students can work with multivariate data, even where the relationships are non-linear.

All the tasks pose challenges not faced in the English school curriculum by students. Nevertheless, many students performed well on these tasks.

**Oxygen** is a 10 mark task. Data on student scores were: mean=4.12; standard deviation=2.85.

**Chortlers** is an 8 mark task. Data on student scores were: mean=3.09; standard deviation=2.70.

**Cowpats** is a 12 mark task. Data on student scores were: mean=6.89; standard deviation=2.03.

None of the tasks have been subjected to detailed analysis in order to determine the key design features which lead to student success. More systematic work, and further invention of task types is necessary if we are to understand and exploit the capabilities of IT to help students work with and interpret complex data.

Informal evidence from observing students during early trialling and during pre-tests suggests a number of distinct benefits from presenting tasks via these new interfaces:

- the user has control over the appearance of the data in ways that are impossible to achieve via paper and pencil – such as via fast changing dynamic displays;
- data can be re-presented rapidly in a number of static forms under user control in ways that can also be achieved using paper and pencil – but only by considerable effort. This allows the students to use a more creative and playful approach to the investigation of the data relationships, because the time penalty for a null result investigation is minimal. As such, half-formed ideas or hypotheses can be tested and either rejected or investigated further;
- three-dimensional representations of data ‘surfaces’ are enabled;
- most striking is the use of hand movements to encode complex data in ways that make it easy to understand and remember, via kinaesthetic dynamic representations.

These data and observations promise a good deal, and raise some fascinating questions for further research.

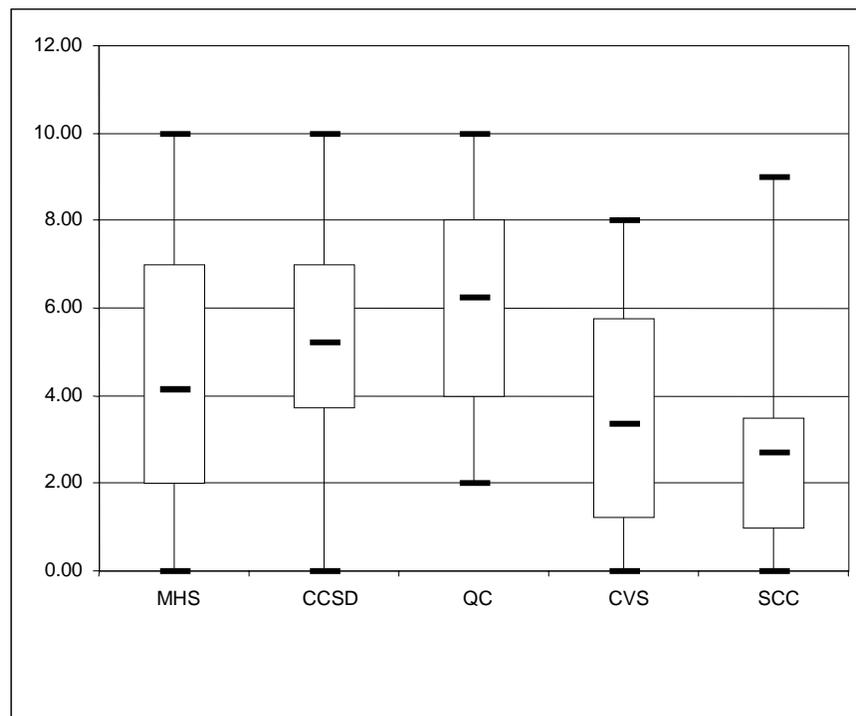
### **Task Performance, School, and Sex**

We set out to look for differences in attainment between schools, and between girls and boys. These judgements are not as easy to make as might first appear. Schools choose whether to enter students, and which students to enter. It follows that a large school entering a small number of its highest attaining students might appear to out-perform a smaller school entering the majority of its students, when this is not the case. Similarly, if schools put forward

a far higher proportion of boys than girls, and there are no sex differences on the test, the pattern of selection could give rise to an apparent sex difference in performance in favour of girls. There could also be an interaction between these two factors; inclusion of students from a high attaining single sex school could distort the evidence on sex differences.

### *School Effects on Performance*

The boxplot below shows data from 5 schools where at least 11 students took the test. The performance of students in the highest performing school is reliably higher than the performance of students in the 2 lowest performing schools.



**Figure 4: Performance on Oxygen in Different Schools**

### *Sex Differences in Performance*

There are no obvious surface features of the task which suggest that the **Oxygen** task will be easier for boys than for girls. However, there are concerns that computer based tasks in general might favour boys. On initial inspection it appears that there is a statistically

significant difference between the mean scores of girls and those of boys, with boys outperforming girls ( $t=2.67$ ,  $df=161$ ,  $p<0.01$ ). However, the highest performing school was a single sex boys' school. To address this confounding effect, two techniques were applied. First, student scores were converted to z-scores based on the performance in each school. The result of this approach showed that the differences in mean scores between the sexes was not statistically significant ( $t=0.67$ ,  $df=161$ ,  $p=0.51$ ). Second, to take account of the fact that different proportions of girls and boys were entered by different schools, only certain schools were entered into the next analysis. For this analysis, we selected schools where the Male/Female ratio is near to 1. In this instance schools were only entered into the analysis if the ratio of girls to boys entered lay between  $2/3$  and  $3/2$ . The results of this analysis showed no significant sex differences in the mean scores on the task ( $t=1.15$ ,  $df=47$ ,  $p=0.26$ ). The same analyses have been carried out on **Cowpats** and **Chortlers**, and in neither case was there any indication that tasks were inequitable in terms of the sex of the students. On the basis of the data we have to hand (which includes data from whole tests, too) there is little evidence of sex differences in performance on IT-based tasks.

### **On the Design of Tasks**

Curriculum tasks are designed to stimulate valuable learning experiences. Assessment tasks define curriculum ambitions, exemplify target behaviours, and show what students know, understand, and can do. Both are central to the success of education. It is interesting, therefore, how little systematic research has been devoted to understanding the task design process (but see Ormerod and Ridgway, 1999; Ormerod, Fritz and Ridgway, 1999; and Samuda, Johnson and Ridgway, 2000), and how few resources are devoted to the development of good assessment tasks. In the UK, this imbalance between the importance of high-stakes assessment and the resources devoted to the development of tasks is particularly marked. The costs of test design for some nationally administered tests are measured in hundreds, not

thousands, of pounds; these tests are used to make decisions which affect the life-chances of students, the status of teachers and schools, and, indeed, the achievements of the education system as a whole. Such a system can work when tests each year are simply rather minor revisions of tests used in earlier years. The system is most unlikely to work when new sorts of test are required to assess new educational goals. Below, we sketch the processes used in the development of the new sorts of computer-based task produced for World Class Tests.

### **The Task Development Process**

Tests have a number of components, notably: tasks; a structure which defines test content; and an administrative system. Here, we focus on task development. Task development follows a particular sequence, which requires creative input and critical analysis. Experienced task designers often blend all these stages into a single process, and have little insight into their own processes. Novice designers experience a multi-stage process characterised by setbacks, extensive revision, and often rejection, and wasted work.

First, an author (or authors) has an idea for a task, then creates a task. Often this will be discussed in a small group, and screen layouts will be sketched. The output from group meetings is often a set of connected ideas which are just too big to form the basis of a single task. The author produces a revised version, which we expect students to be able to complete in 10-15 minutes. By now the task is in a state where critical analysis is useful. The author must describe the key activities, the substantive content, and have ideas on how the task will be scored.

WCT tasks are written in FLASH by professional programmers. However, because the process of task design necessitates a great deal of revision, programmers are sent task specifications at a very late stage of development, in order to minimise change requests. The bulk of the task development process is done in-house, by our design team, working in either EXCEL or FLASH. Typically, once a task idea is well developed, the task can be created in

about a day (excluding data recording). Prototype tasks are discussed by the group, using the criteria set out below, either formally or implicitly.

*Do the task in the way you would expect a good student in the target group to do it.*

Is the purpose of the task clear to the students? Is the language clear?

Are the contexts natural and fully explained? Is the task really sensible in its context? Can the task length be justified in terms of the information gained about student knowledge?

How accessible is the task? Is there an easy entry point? Is there sufficient demand at the end?

Is there sufficient time for students to: complete the task; explore alternative strategies; check their work?

*Describe the Task:* What is being assessed? (which problem solving strategies? which concepts and connections? which skills? what knowledge?) Is the task worth being able to do? (does it stand as a worthwhile task in its own right? what class of worthwhile problems does it belong to?)

Have different solution methods been anticipated or encouraged? (Is there really just one solution method? If so, what is it? What are the alternative approaches?)

*Evaluate the Marking Scheme:* Is it unambiguous? Does it reward important skills appropriately? Does it reward a variety of solution methods?

*Look at Student Work:* Does the scoring scheme cover the variety of student work?

Are there ambiguities? Does the task differentiate between students?

**Figure 5: Criteria for Task Evaluation**

Once the task is judged to be good enough, it is trialled with a small group of students – usually 3 individuals. Students are asked to take the role of ‘co-developers’ – they work on the task, and are observed as they work. We ask them what they think the task assesses, about any

aspects that are not clear, and about ways to improve the task. The task may be rewritten in the light of these trials.

Once a task is deemed to be acceptable, it is sent for (re)programming. It is then trialled with a larger sample, (set in a standard test administration ‘shell’) – usually with between 20-30 students. The purpose of this level of trialling is to ensure that the task is accessible to students, and that a range of student performances is obtained (i.e. that the task is neither impossible not too easy), and that we have a range of student solutions that can be used to modify prototype marking schemes. In the final stage of trialling – the pre-test – tasks are assembled into tests in a way that reflects out test ‘blueprint’ or design specification. Here about 200 students take part. The pre-test provides information on the score distribution of tests and tasks, and allows us to tune the marking scheme. Our ambition is to have to make few changes between the pre-test and the live test.

This design process has a number of important features:

- it is highly iterative;
- criteria for good tasks are considered from the outset;
- changes are made early, before extensive resources have been committed;
- outside groups are asked for well defined semi-final products; their creative input is actively discouraged;
- tasks are trialled extensively before they are used.

### **Developing a Better Curriculum**

We believe the current statistics curriculum is inadequate to equip students to become informed citizens in the 21st century where they will be required to deal with complex data sets.

There is good evidence that appropriate IT support can make the understanding of data far easier for quite young students, for example by presenting new sorts of tasks, where dynamic displays show changes in several variables over time. Interaction makes computers well suited to discovering rules and finding relationships. Furthermore students can work with complex realistic data sets, using methods similar to those used by professionals in the field (Ridgway and McCusker, 2003).

We have only just begun to explore the qualitative aspects of this phenomenon, as we create new interfaces. Our current research has two distinct features. First is the development of curriculum modules (at age 9 and 13 years) on data handling which set out to capitalise on the enabling features of IT discovered so far (this complements two other modules at each age on other aspects of problem solving). Second is the development of new interfaces and tasks, and research into the development of statistical competence.

A module comprises 4 one-hour units. In the case of *Seeking Relationships in Data* (MARS, 2004), designed for use with 13 year-old students, the components are:

### **Lesson 1: Controlling variables**

Two situations are presented. In the first, students explore the effects of 2 types of plant food on plant growth via a simulation, using (hopefully) a systematic search procedure. In the second task, students work with a system which calculates how long a hike will take, given the vertical and horizontal distances travelled.

### **Lesson 2: Finding relationships in situations with two independent variables**

Here students work with simulations in mechanics and optics to discover inverse proportional relationships.

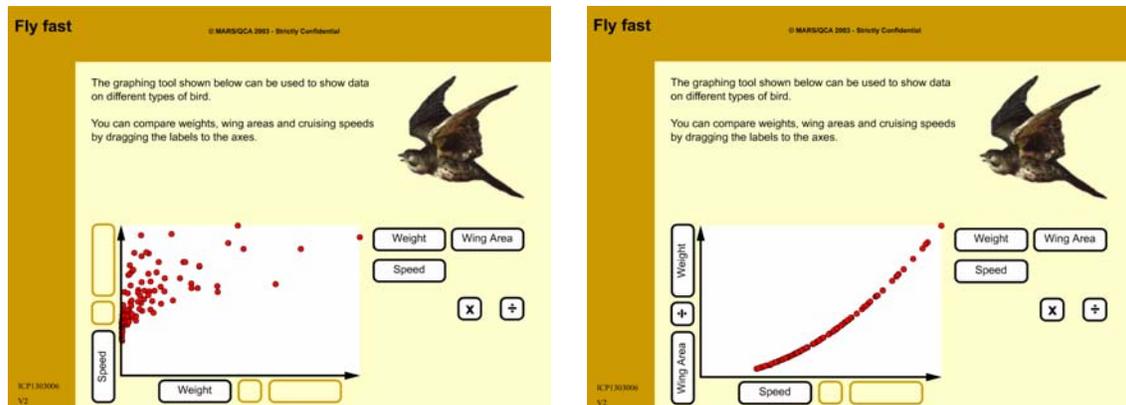
### **Lesson 3: Finding relationships in situations with three independent variables**

The context is the design of bridges. Students can change the length, width and thickness of a plank bridge, and must discover linear, inverse and inverse squared

relationships, then combine them into a single equation to predict the strength of any given bridge.

#### Lesson 4: Finding relationships from graphs

Students are introduced to the power of exploratory data analysis via a unique and novel interface, where the axes are given some of the functionality of spreadsheets (see Figure 6). Students use the computer to generate and study graphs in order to predict the speed at which a bird can fly, given various biometric data. As part of the exploration, students can easily create new variables (such as ‘weight/wing area’).



**Figure 6: Screens of the Fly Fast Task using Functional Axes**

#### Understanding the Applications of IT

We are continuing our explorations of those features of effective interfaces which actually make them effective. This work extends our test development into the classroom where we can develop more extensive task situations and simulations, and can provide far more support for students in the form of teacher guidance, and support materials such as examples of other students' work, check lists and hint sheets.

## Conclusions

Statistics education in the UK falls far short of the ambitions of Nightingale and Wells. There is too little advocacy of the use of statistical technique to inform thinking in key curriculum subjects, and the simplifications made in statistics courses actually go against Einstein's advice, and make things too simple. Multivariate realistic problems (which students can actually solve) are ignored, and the formal statistics curriculum is reduced to practicing algorithms out of context, for which the range of applications is limited. IT can be used to present realistic, multidimensional data to students. Evidence from the World Class Tests shows that students – some as young as 9 years – can work with such data to reason about realistic situations, and produce sensible conclusions. IT has a number of unique features which make it particularly well suited to teaching and assessing skills in handling complex data. Evidence shows strong school effects (which suggests that data handling skills can be taught effectively) but few sex differences. The process of designing effective tasks has just begun; some design principles are emerging. Tasks which are designed for assessment purposes are sufficiently rich to form the basis for valuable curriculum materials.

## References

- Batanero, C., Godino, J.D., Vallecillos, A., Green, D., and Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematics, Education, Science and Technology*, 25(4), 527 – 547.
- Greer, B. (2000). Statistical Thinking and Learning in *Mathematical Thinking and Learning 2 (1&2)* p 1 – 9.
- Hawkins, A. (1997). Myth-Conceptions! In J.B.Garfield and G. Burrill (Eds) *Research on the Role of Technology in Teaching and Learning Statistics*. Voorburg: International Statistical Institute.

Holmes, P. (2000). Statistics Across the English National Curriculum.

<http://science.ntu.ac.uk/rsscse/schools/natcur.html#top>

MARS (2004, in press). Seeking Relationships in Data. London: Granada Learning.

Nicholson, J.R., Mulhern, G., and Hunt, D.N. (2002) Wizardry or Pedagogy?: What is the driving force in the use of the new technology in teaching statistics? *Proceedings of ICOTS VI*, Voorburg: International Statistical Institute.

Nicholson, J.R. (2003). GCSE statistics coursework, and dealing with multi-dimensional data. *Mathematics in Schools*, 32(3), 8 - 14

Ormerod, T. C., and Ridgway, J. (1999). *Developing task design guides through studies of expertise*. In S. Bagnara (Ed.), 3rd European Conference on Cognitive Science, ECCS '99, Siena: Consiglio Nazionale Delle Ricerche. pp. 401-410.

Ormerod, T.C., Fritz, C.O. and Ridgway, J. (1999) *From deep to superficial categorisation with increasing expertise*. Proc. 21st conference of the Cognitive Science Society, Vancouver, August. pp.502-506.

Ridgway, J., and McCusker, S. (2003). *Using Computers to Assess New Educational Goals*. *Assessment in Education*, 10(3), pp 309-328.

Richardson, M., Baird, J-A., Ridgway, J., Ripley, M., Shorrocks-Taylor, D., and Swan, M. (2002). *Challenging minds? Students' perceptions of computer-based World Class Tests of problem solving*. *Computers in Human Behavior*, 18(6), 633-649.

Samuda, V., Johnson, K., and Ridgway, J. (2000). *Designing Language Learning Tasks: A Guide*. Working Papers on Task Design. Department of Linguistics and Modern English Language, University of Lancaster, UK.

Tversky, A., and Kahneman, D. (1983). Extension versus intuitive reason: the conjunction fallacy in probability judgement. *Psychological Review*, 90, 293-315.

## Appendix

Q	OXYGEN: Age 13		Points	Section points
	<p>The core elements of performance required by this task are:</p> <ul style="list-style-type: none"> <li>• explore relationships among 3 variables in a scientific context;</li> <li>Identify optimal values.</li> </ul> <p>Based on these, credit for specific aspects of performance should be assigned as follows:</p> <p><i>Mark entirely from paper</i></p>			
1	<p>Oxygen production:</p> <ul style="list-style-type: none"> <li>• is not affected by light intensities below 5</li> <li>• increases with light intensity above 5.</li> </ul> <p><i>Part mark:</i> If pupil states that 'as light increases, more oxygen' :</p> <ul style="list-style-type: none"> <li>• increases as temperature rises up to 30°C</li> <li>• decreases as temperature rises over 30°C</li> </ul> <p><i>Part mark:</i> If pupil states that 'best temperature is 30°C' :</p>		1 1 (1)  1 1 (1)	4
2	<p>They could both be right.</p> <p><i>Reason:</i> They hadn't controlled the light intensity. or Ann was looking at low light intensity Jim was looking at high light intensity.</p>		1  1	2
3	<p>Ann 's idea is better than Jim's.</p> <p><i>Reason:</i> e.g. Jim's idea might overheat the plants.</p> <p>A better idea still would be to have Temp = 30°C, Light intensity =50.</p>		1  1  2	4
	Total Points		10	10



<b>Task Name:</b>	Cowpats
<b>Age Range:</b>	13

## Scoring Guide

Q	box	PC/ Paper	Computer scoring/Manual scoring/Guidance	Point s	Section Points
1	A		<p><i>Student should recognise sequence of events, i.e.:</i></p> <ul style="list-style-type: none"> <li>Starts with lots of dung</li> <li>Flies start to arrive.</li> <li>Worms start to arrive</li> <li>Beetles start to arrive.</li> <li>Dung starts to disappear.</li> <li>Worms flies and beetles start to disappear.</li> </ul> <p><i>Deduct 1 mark for each significant omission or clearly incorrect observation.</i></p>	3	3
2	B		<b>gorder=("g2","g3","g4","g1")</b>	2	2
3	C		<p><b>ans3nouns contains "T2" (Beetles)</b></p> <p><i>i.e. has recognised that only the beetles seem to affect the amount of dung.</i></p>	1	4
	D		<p><b>ans3verbs contains only some combination of "T4" (Bury), "T5" (eat) and "_null" and is not (_null,_null)</b></p> <p><i>i.e. sees that only burying or eating the dung will make it go away, doesn't mention eggs.</i></p>	1	
	E		<p><b>ans3nouns=("T2","T2") and (ans3verbs=("T4","T5") or ans3verbs=("T5","T4"))</b></p>	2	
4	F		Cows drop dung everywhere but the cowpats on the fields disappear quickly...	1	3
	G		<p>...because the beetles can't bury dung in a concrete drive.</p> <p><b>Part mark</b></p> <p>... because worms can't get to the cowpat if it's on concrete.</p>	2 (1)	
<b>Total:</b>					<b>12</b>