Mathematics teachers teaching statistics: What are the challenges for the classroom teacher?

James Nicholson
Belfast Royal Academy, Department of Mathematics
Cliftonville Road
Belfast BT14 6JL, Northern Ireland
j.r.nicholson@qub.ac.uk

Catharine Darnton
Qualifications and Curriculum Authority, Mathematics Test Development Team
83 Piccadilly
London W1J 8QA, United Kingdom
DarntonC@qca.org.uk

1. Introduction:

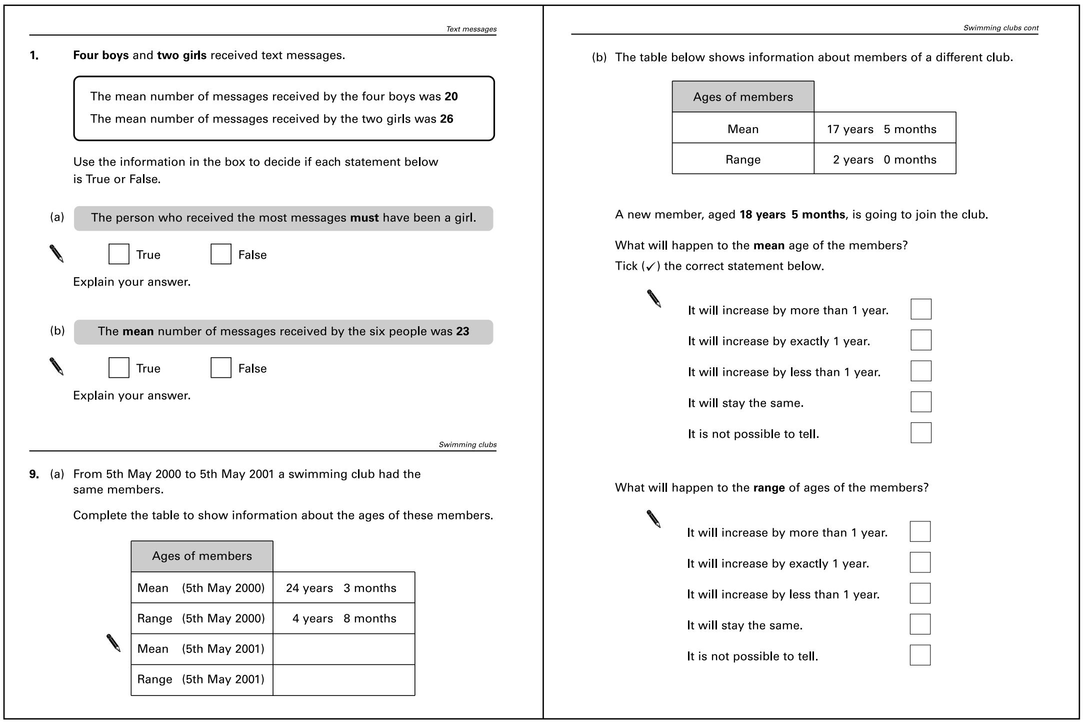
Data Handling and Chance are now integral components of many national curricula in mathematics. However, they require some different approaches from both teachers and learners from other components of the curriculum. This may pose particular problems if the teacher has not studied statistics as part of his or her mathematics course, or indeed if the teacher is not a subject specialist. We consider two broad areas in which conceptual difficulties arise across the whole secondary age range. We seek to identify particular classroom issues from analysis of some questions used in the statutory national tests for key stage 3 [taken by 14 year olds] in England, and attempt to draw parallels with difficulties seen in the later stages of secondary statistics.

The statutory tests now routinely ask for an explanation of reasoning in data handling questions, as well as in other curriculum areas. While pupils at this age can struggle to articulate their reasoning processes explicitly, we can gain some insights into their thought processes from analysis of large numbers of scripts. The A-level examinations at the end of secondary education reward computational accuracy and procedural competence much more heavily than interpretative skills. Typically only 10 to 20% of the total assessment focuses on interpretation. The incentive for teachers to spend the time needed to develop interpretative skills to a high degree may be lessened as a result.

2. Essentially mathematical statistics:

In most school mathematics there is a single 'correct answer' to the question posed, although there is often more than one way of arriving at that answer. Calculations of statistical quantities such as the mean and range of a data set would be examples of this type of situation that students meet at key stage 3. There is a tendency for teachers to focus on procedural aspects of calculating the correct answer, placing less emphasis on why the measures are useful to represent a data set, and what can be inferred about underlying data sets by comparing the measures. Indeed the simple understanding that summary statistics do not uniquely identify the complete set of values can present difficulties for pupils when assessed in context, although we might suspect that few would make this error if asked directly.

The difficulties that pupils have reasoning with the mean could be seen in responses to the question *Text messages* shown in Figure 1 below. Pupils had to use information about the means of two data sets to decide whether two statements were true or false and then explain and justify their conclusions. This required pupils to go beyond knowing how to calculate the mean, and enabled those who could do so to demonstrate a deeper understanding of the concept.



Text messages

1. **Four boys** and **two girls** received text messages.

The mean number of messages received by the four boys was **20**
The mean number of messages received by the two girls was **26**

Use the information in the box to decide if each statement below is True or False.

(a) The person who received the most messages **must** have been a girl.

True False

Explain your answer.

(b) The **mean** number of messages received by the six people was **23**

True False

Explain your answer.

Swimming clubs

9. (a) From 5th May 2000 to 5th May 2001 a swimming club had the same members.

Complete the table to show information about the ages of these members.

Ages of members	
Mean (5th May 2000)	24 years 3 months
Range (5th May 2000)	4 years 8 months
Mean (5th May 2001)	
Range (5th May 2001)	

Swimming clubs cont

(b) The table below shows information about members of a different club.

Ages of members	
Mean	17 years 5 months
Range	2 years 0 months

A new member, aged **18 years 5 months**, is going to join the club.

What will happen to the **mean** age of the members?
Tick (✓) the correct statement below.

It will increase by more than 1 year.
It will increase by exactly 1 year.
It will increase by less than 1 year.
It will stay the same.
It is not possible to tell.

What will happen to the **range** of ages of the members?

It will increase by more than 1 year.
It will increase by exactly 1 year.
It will increase by less than 1 year.
It will stay the same.
It is not possible to tell.

Figure 1: Exemplar questions relating to essentially mathematical statistics.

The most common correct explanations for part (a) either referred to the range not being known, or showed data for which the statement was false, or referred, for the boys, to the possibility of smaller values being counterbalanced by bigger values. Some pupils were able to calculate the total text messages received by boys as 80 and by girls as 52, but then concluded incorrectly that it must have been a boy who received most text messages. However a much more common error was to believe that the statement was true. The following example of a student response indicates the type of explanation that commonly accompanied the incorrect decision:

'Out of two girls the mean was 26, which means each received 26. Four boys, each must have received 20.'

This response suggests that the pupil may have had insufficient experience of the range of possible different data sets that can give rise to the same mean. Extensive pre-testing of items also indicates that asking for an explanation of reasoning increases the level of demand of the item. The effect of contextualising a problem is more variable – in some cases the context gives support but, in others, the embedding of the mathematics can cause difficulties for pupils.

Pupils had to consider how the mean and range change when the underlying data change in the question *Swimming clubs*, shown in Figure 1 above. Two types of incorrect response were commonly seen. Many pupils assumed either that both the mean and the range would increase by one year or that both would remain unchanged, and so seemed unable to consider the possibility that the behaviour of the two measures would not be consistent. These errors suggest that these pupils lacked experience in transforming simple data sets and comparing measures of location and spread. In particular, the errors suggest that the pupils have not fully appreciated that the two measures relate to quite different aspects of the data set and that there is no reason why both should be affected in the same way. The common error for the first mark in part (b) was to indicate that it was not possible to tell what the effect on the mean age would be. This was also the common response for the second mark. However for the second mark it was the correct response. It seems reasonable to surmise that pupils were indicating in both cases that they could not tell, probably as a result of lacking any strategy to visualise the possible characteristics of the underlying data set.

At the upper end of the statistics curriculum in secondary schools, pupils work with more sophisticated measures. Probability calculations involving distributions such as the Binomial and

Normal, or finding the regression line for a set of bivariate data, would fall into the same broad category for them as the above examples would for pupils at key stage 3.

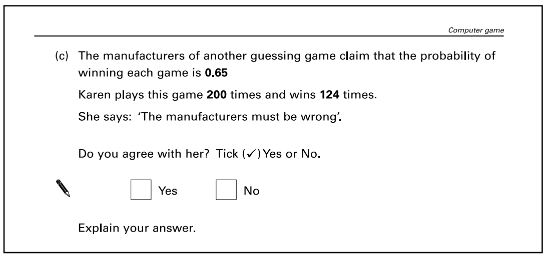
In the current assessment system at A-level the requirement to demonstrate a depth of understanding of key concepts is limited. For example, students will be required to calculate lines of regression, and use them for prediction. Often they will be asked whether prediction is appropriate for a value of the independent variable that requires extrapolation, but full credit can be gained by a simplistic response that identifies that extrapolation is required and that extrapolation is always problematical. It is not necessary to observe anything about why – that the existing model may no longer apply if the range of x is extended – or to give any evaluation of how likely it is that the model will change in the context of the particular set of data. A similar observation could be made about the level of sophistication required in giving the outcome of an hypothesis test. If the correct decision is reported i.e. whether the null hypothesis should be accepted or rejected, and if any other statement made is not contradictory to that decision, full credit is likely to be awarded even where inappropriately assertive language such as *'the mean has increased'* is employed.

There are other concepts underpinning some standard techniques for which little or no assessment is attempted. For example, predictions generated by lines of regression contain more variability as x moves away from $\bar{x}$ because the line is actually a stochastic quantity, depending on the particular sample of observations. The higher the correlation of the two variables, and the larger the number of data pairs available, the more consistent will be the regression lines obtained. However the lines still diverge as x moves away from $\bar{x}$, with consequently more variability in the predicted y values.

3. The effects of randomness and exercising judgement:

These are situations where there is an element of reasoning or judgement to be exercised, and which therefore do not necessarily have a single 'correct response'. Undoubtedly these situations are harder to provide instruction for than situations where the response can be broken down procedurally. Moreover, they are much harder to assess appropriately, since the possibility exists that an 'inaccurate evaluation' (as the examiner sees it) may be due to incorrect reasoning, or to correct reasoning using a different value system. Unless the examiner can see both the conclusion which a student has drawn and some evidence as to the process by which it was reached, it is not possible to assess reasonably the worth of the response.

We live in a world where people sincerely hold very different views, based on the 'same evidence'. Some people may be averse to a particular risk where others are more inclined to take it, and indeed the same person may display different characteristics with respect to risk depending on the context, and in particular depending on the consequences which might follow in a risk situation.


Computer game

(c) The manufacturers of another guessing game claim that the probability of winning each game is **0.65**

Karen plays this game **200** times and wins **124** times.

She says: 'The manufacturers must be wrong'.

Do you agree with her? Tick (✓) Yes or No.

☐ Yes ☐ No

Explain your answer.

Figure 2: Exemplar question relating to the effects of randomness and exercising judgement.

If pupils are to be able to understand the logic of hypothesis testing, and risk assessment strategies generally, it would seem to be wise for them to have had some prior experience in reasoning informally about what can, and also what cannot, be concluded from the result of a probabilistic experiment. This type of reasoning was assessed in part (c) of the question *Computer game* shown in Figure 2 above. Only a quarter of pupils working at the level at which the question was aimed were able to give a sufficiently clear explanation, that random processes are unpredictable, to justify the award of the mark. About the same proportion ticked 'Yes', commonly with the explanation that 124 out of 200 is not equal to 0.65. It seems possible that pupils giving this incorrect response lacked practical experiences of situations involving chance. The better performance on questions requiring similar analysis in a context such as coin tossing would suggest that many pupils have difficulty in transferring a principle from one context to another which is less familiar.

The study of random processes leads to formal inference and decision making in A-level statistics courses, and this is an area in which many teachers trained in mathematics and without a strong statistical background are uncomfortable. In statistics we are faced with taking decisions in the face of uncertainty. Subsequent information, i.e. the outcome of the actual event that was previously uncertain, might show that an alternative course of action would have been preferable in terms of outcome. However, this does not mean that the decision itself was the wrong one to have made, or that a different decision should be made if presented with the same situation again in the future.

In the United Kingdom there is a television quiz show *The National Lottery Jetset* where contestants have to guess whether the number on the next ball in the sequence of draws for that night's lottery was higher or lower. Since only the numbers 1 to 49 are used in the draw, and the maximum history to be adjusted for is when 5 balls have already been seen, it is easy to identify the optimal strategy. A correct guess means that the contestant controls the category of question he or she will be asked, which has some effect on the outcome of the contest. Another similar situation would be in purchasing travel insurance against which I subsequently make no claim – it doesn't mean that I made a bad decision in the purchase.

4. Conclusions:

Reasoning with data requires different skills from much other mathematics. Teachers who are not familiar with common difficulties and misconceptions may miss opportunities to help pupils confront these and achieve a deeper understanding of core concepts. We recognise that it is difficult for pupils to articulate their reasoning, but as technology allows more of the computational work and graph drawing to become automated, there should be scope for greater emphasis to be placed on interpretative skills. Experience of a wide range of contexts and types of data formats is necessary for both pupils and teachers to develop confidence in this area, so it is important that suitable resources become widely available to support the development of these skills. As students come through to A-level with greater experience of interpreting data, we would like to see the development of assessment of the underlying concepts at a deeper level than is currently the case.

5. Acknowledgements and source of questions:

The discussion of issues at key stage 3 is based on work undertaken by the Mathematics Test Development Team at QCA. *Text messages*: Changes to assessment 2003: sample materials for key stage 3 mathematics, QCA 2002. *Swimming clubs*: Key stage 3 statutory tests 2002, QCA 2002. *Computer Game*: Key stage 3 statutory tests 2001, QCA 2001.

RÉSUMÉ:

Le 'traitement des données' et 'la chance' sont des composants intégraux de beaucoup de programmes d'études nationaux des mathématiques. Ils exigent quelques approches différentes des professeurs et des étudiants à d'autres composants du programme d'études en raison de la présence de l'incertitude. Cet article considère quelques questions de la salle de classe appropriés à deux étapes différentes d'éducation: les issues indiquées par une analyse des essais nationaux Anglais pour les étudiants agés de 14 ans et les issues résultant d'une considération des modules de statistiques du programme d'études 16-19.