Cite as: Jones, K., Pratt, D. & Watson, A. (2014). Key ideas in teaching mathematics: Statistical reasoning in Key Stage 3. *NCETM Secondary Mathematics Magazine*, 107, 3-4.

Secondary Magazine 107

National Centre for Excellence in the Teaching of Mathematics

Key Ideas in Teaching Mathematics – Statistical reasoning in Key Stage 3

In this and other issues, the Secondary Magazine will feature a set of six articles, written by Anne Watson, Keith Jones and Dave Pratt, the authors of the recent publication Key Ideas in Teaching Mathematics. While not replicating the text of this publication, the articles will follow the themes of the chapters and are intended to stimulate thought and discussion, as mathematics teachers begin to consider the implications of the changes to the National Curriculum. This article is the third in the series and focusses on Statistical reasoning in Key Stage 3. Future articles will feature Place Value, Algebra and Probabilistic Reasoning. Previous articles focussed on similarity, ratio and trigonometry in Key Stage 3 and Geometric and spatial reasoning in Key Stage 3.

In 2009, Hal Varian, Chief Economist at Google and emeritus professor at the University of California, predicted, "I keep saying the sexy job in the next ten years will be statisticians…The ability to take data – to be able to understand it, to process it, to extract value from it, to visualise it, to communicate it – that's going to be a hugely important skill in the next decades, not only at the professional level, but even at the educational level for elementary kids, for high school kids, for college kids. Because now we really do have essentially free and ubiquitous data. So the complementary scarce factor is the ability to understand that data and extract value from it."

The above quote emphasises the importance of statistical reasoning in professional lives. We might add to the argument the critical skills needed in evaluating media claims on for example the impact of eating certain foodstuffs, in making personal decisions about for example whether to have an operation, and in seeking to understand policy statements that affect the society we live in. It is not difficult then to see the teaching and learning of statistics as one of the more important and relevant subjects in school.

Hal Varian's quote above also succinctly captures statistics as a process, sometimes referred to as the data handling cycle. Some researchers have called this the investigative cycle and include identifying the system dynamics of the problem, designing how to measure the identified variables, collecting the data, analysing the data including generating hypotheses and finally interpreting in order to communicate conclusions.

Research consistently argues that statistical thinking lies in that process and, whilst knowing facts such as ways to calculate various types of average or spread, how to draw a pie chart or how to conduct a t-test are important, they are subsidiary to appreciating the cycle as a process within a problem context. Engagement with data, seeking to extract value from it, is key. Some have argued that this type of thinking is quite distinct from mathematical thinking and that, if the nature of proof is one of the hallmarks that distinguishes a discipline, proof in statistics is very different from that in mathematics. Nevertheless, the teaching of statistics will in the foreseeable future lie with the mathematics teacher, though there is a need to make the distinction between statistical and mathematical thinking explicit.

Traditionally statistics as a discipline is divided into what are sometimes referred to as the descriptive and the inferential. Descriptive statistics is important when one merely seeks to capture and convey information about the whole set of data available. This is especially relevant when all or very large sets of data are available as in a census. Inferential statistics is important when one seeks to draw a conclusion about a population based on a relatively small set of data. The essential elements of measurement, collection, analysis and interpretation apply in both 'games' – the descriptive 'game 'and the inferential 'game'. However, the detail of the data handling cycle is different depending on which type of 'game' one is playing. For example, in the inferential 'game' one might apply one or more of a battery of tests to evaluate whether the sampled data is unlikely to have been generated were a certain hypothesis true. Such tests are not relevant in the descriptive 'game' and research has shown that it is important that teachers and students are aware which of these two 'games' are being played when trying to address any particular problem.

The programmes of study for statistics in the National Curriculum and in the GCSE subject content are structured around these two types of statistics. At Key Stages 1 to 3, the emphasis is very much on descriptive statistics with statements at Key Stage 3 such as: 'describe, interpret and compare observed distributions'; 'construct and interpret appropriate tables, charts and diagrams'; 'describe simple mathematical relationships between 2 variables'. The GCSE content includes statements such as: 'infer properties of populations or distributions from a sample' alongside the use of descriptive statistics.

There is a considerable challenge to understanding in moving from the descriptive to the inferential, which many students will not meet unless their experiences at Key Stage 3 have prepared them. Because of the importance of statistics to future professional lives and in everyday engagement as a citizen and because of the sudden increased expectation at Key Stage 4, we would argue that, notwithstanding the competing claims of other topics and disciplines, mathematics teachers at Key Stage 3 need to introduce students to the inferential 'game' through the data handling cycle whenever possible.

Research shows how this might be done. There has been a substantial body of statistics education research development in the last ten years which points to the pedagogic value of what has become termed 'informal inferential statistics' (IIR). IIR almost always draws on modern software developments. Indeed much of the research has been conducted using TinkerPlots, as an especially well-designed application for younger students, though much of the work could be done through other tools. The key idea in IIR is for even very young students to play the inferential 'game' through the data handling cycle. Students might be challenged with a problem, designed to pique their curiosity, which will require them to collect data and draw inferences. For example, asking how tall younger students might be when they are fully-grown can lead to the collection of growth statistics and their interpretation. Alternatively, students might be given access to a set of data about which the students pose questions and explore. For example, see CensusAtSchool.

One pedagogic approach developed by researchers is referred to as 'growing samples'. Students might be challenged to decide the answer to a question, such as 'How many hours of television per week do children in your year watch?' on the basis of asking fellow students who sit near to them in class. The variation in responses across different groups in the class could be discussed. The sample could then be grown by joining groups and then further by collecting data across the whole class. The process could continue by amalgamating classes until a census of the year was achieved. At each stage questions could be asked about how confident the students were in their answer. Once the results for the year were collected (when confidence should be 100%), the question can be changed to ask about the whole school or even town. Similarly using CensusAtSchool, samples of gradually increasing size could be taken to note the effect of sample size on the confidence with which the inferences can be drawn. This pedagogic approach places emphasis on the inferential 'game' in a way that is accessible to quite young students (it has been trialled in primary as well as secondary schools) because it does not require the knowledge needed to conduct formal tests as in classical inference.

In this short article, we have proposed the need to teach beyond the requirements of Key Stage 3 statistics on the basis that statistics is extremely important and relevant and yet the Key Stage 3 curriculum will not prepare students for the demands of inferential statistics at Key Stage 4. We have pointed to the research on informal inferential statistics as an approach that makes it feasible to address inferential statistics through the data handling cycle at quite a young age.

Keith Jones, Dave Pratt and Anne Watson