ASSESSING STATISTICAL PROBLEM SOLVING

GIBSON, Liz, MARRIOTT, John & DAVIES, Neville
Nottingham Trent University
United Kingdom

In this paper we report the results from a major UK government-funded project, started in 2005, to review statistics within the school mathematics curriculum for students up to age 16. New teaching materials that explicitly use a problem-solving approach through other subjects have been developed. We will report extensive trialing of these, the development of corresponding new assessment regimes and how these work in the classroom. The new ways of assessing are particularly poignant since, in September 2006, the UK government announced that coursework is to be dropped for mathematics exams sat by 16-year-old. As a consequence, areas of the curriculum previously assessed by coursework, are now being ignored. We will provide some new and useful ways of assessing this content. Our findings have implications for teaching, learning and assessing statistics for students of the subject at all ages.

INTRODUCTION

The state of mathematics education in the United Kingdom has been an issue of concern for some years. Following a commission by the UK government Smith (2004) published a report into post-14 mathematics provision. His report made wide-ranging recommendations for improving mathematics education in schools in England. Following one of the recommendations, in 2005 the Qualifications and Curriculum Authority (QCA) commissioned the Royal Statistical Society Centre for Statistical Education (RSSCSE) to review the position of the *teaching* of statistics in the national curriculum for England. Indeed Smith had suggested that statistics education might be improved by teaching it through other subjects, such as science and geography, rather than from its current position in the mathematics curriculum.

In autumn 2005 the RSSCSE/QCA Review carried out a national survey of heads of mathematics, geography and science to determine their views, needs and capabilities within the statistics area. The results showed that teachers of science and geography appeared more confident than mathematics teachers when teaching school students to understand and interact with statistical concepts and ideas - skills that are fundamental in developing statistical literacy.

An examination of the approach to the teaching and learning of statistics within the science, geography and other curricula revealed that the key differences between the approach adopted within these subjects and that used within mathematics were in the application of statistics to real contexts and problems arising from within each subject. The RSSCSE/QCA Review project recommended the teaching of statistics should be seeded through real world examples drawn from science, geography and other subjects but should be retained within the mathematics curriculum. This corresponded with the growing body of opinion that statistics is best taught through problem solving. Consequently the project developed a set of learning and teaching resources for use in mathematics lessons by mathematics teachers which use a statistical problem solving approach (PSA) and draw on real problems in real contexts. Alongside the teaching materials a new assessment regime was designed to test the effectiveness of the materials.

The teaching materials and supporting documentation can aid teachers in incorporating formative assessment into their teaching. The teachers' notes contain suggestions for discussion and so support teachers in their use of questioning for formative assessment. There are also worksheets which allow teachers to get quick insights into pupils' understanding throughout the lesson. Later we describe the approach we adopted to more summative assessment methods.

It was considered that an assessment regime that graded students' knowledge and skills in problem solving could not just rely on asking, for example, questions about the mechanics of calculating statistics or the ability to draw bar charts and/or histograms. The regime needed to focus on the thinking behind each stage of the process, the links between the stages and an ability to entertain the fact that there may be more than one solution to a statistical problem. In the next

section we describe the way in which we developed the assessment regime taking into account these key points and the thinking needed to learn through statistical problem solving.

DEVELOPING A NEW ASSESSMENT REGIME

The PSA consists of a four stage cycle using the four activities: specify the problem and plan; collect data from a variety of suitable sources; process and represent the data; and interpret and discuss the results. This approach has been advocated as part of the English national curriculum in mathematics since 2000 where it is referred to as 'the handling data cycle'. Using this approach together with real data in real-world problem contexts draws on a number of different cognitive skills. Thus, in devising an assessment regime that could be used to grade students' problem solving skills, these *cognitive* skills needed to be identified. Different forms of learning also needed to be recognized. Bloom et al. (1956) published a taxonomy of educational objectives which was later revised by Anderson and Krathwol (2001). The six categories of Anderson's revised taxonomy are: remembering; understanding; applying; analysing; evaluating; and creating. These categories are considered to be a hierarchy of skills – although there is some educational debate as to whether they are also progressive.

In considering the development of an assessment regime we completed a mapping from the PSA as specified within the English national curriculum for mathematics onto the categories of the revised taxonomy. This mapping revealed that *each* stage in the cycle demands the use of at least four levels of the taxonomy. For example, at the first stage of the PSA, the only category that is not used to any extent is 'evaluating'.

Anderson and Krathwohl (2001) also introduced a second attribute/category which they refer to as the *knowledge* dimension. The categories of this dimension, representing the outcomes of the thinking process, are factual, conceptual, procedural and metacognitive. The learning objectives of the statistical PSA were mapped onto the resulting two way classification that combines the cognitive process dimension with the knowledge dimension. The resulting table is not reproduced here, but can be viewed at www.rsscse.org.uk/qca. This process then naturally identified our starting point for the development of the assessment and associated questions.

After considering various options it was decided that an *online* assessment regime which involves students setting themselves up as advisors and critiquing the work of others should be developed. Having recognized that high level skills are demanded in applying the PSA it was felt that any assessment needed to occur in a supported environment, particularly when considering that the initial design of the learning materials was for use with students aged between 12 and 14. However, we now believe that older students, for example taking pre-university examinations or first year university statistics modules, will benefit from the approach we have adopted, both for learning and assessment.

It was also decided that the assessment would be time-constrained and should be as accessible as possible. By necessity, statistics has a high level of literary demand and so by using an online tool, features such as audio commentary and zooming can be used to improve accessibility. Similarly, the choice of a familiar context for the problem at the core of the assessment puts students at ease and helps to generate genuine interest in the problem they are trying to solve. A computer-based tool has the added advantage that large sets of data can be used where appropriate, and are easily stored on a computer database. We believe that the assessment environment not only draws out students' understanding in these areas but also presents statistics in a manner similar to the way it is used in practice by professional statisticians. There is a gradual move towards the use of online assessment from within the QCA, which was also one of the recommendations of Smith (2004) which contributes to our belief that online assessment is the future.

We chose as a context for the assessment a scenario associated with the purchase of a new mobile phone: *Getting the best deal* .It comprises three sections, A, B and C, and was designed to take one hour to complete. The questions used are described in detail in Table 1 and we now summarise the online assessment. Section A is designed to test the candidates' knowledge of the PSA. Candidates' knowledge of the stages of the cycle is tested using a drag and drop facility (see figure 1). They are also asked about the understanding of the cycle as a whole.

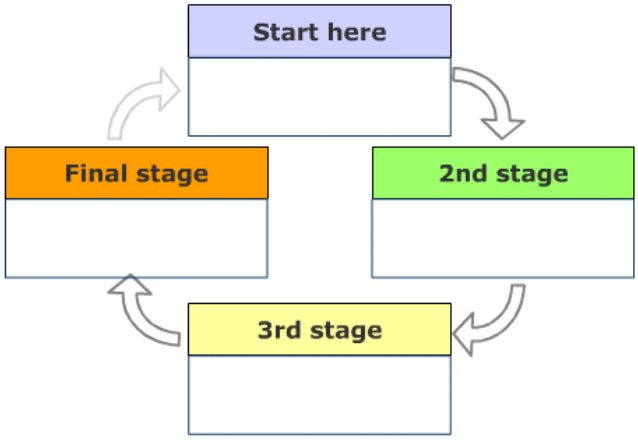

Figure 1: Template for the first task in the assessment

Within section B candidates are required to work with an investigation completed by a fictitious student. They are asked questions about how activities contribute to the whole process as well as being asked to critique the students' approach. This is followed by eight further questions that carefully probe the candidates' understanding of what the student (Ayesha) did at each of the different stages, and the extent to which she was correct in her work.

The same problem (getting the best deal for a mobile phone) is addressed in Section C with another student, 'Andy', who has not yet completed the problem solving cycle. Candidates are led through eight slides in which available information is presented and the candidates are invited to help make decisions on how to proceed. After some of these decisions the candidates are told what was *actually* done but they cannot return to previous pages to change their responses. Amongst other things, the candidates are asked to undertake calculations, make comments on the data and to consider suitable graphical presentations and summary measures. They must type in responses as part of a discussion of what has been found and finally make a decision as to which phone should be purchased. The last three questions explore the fact that the data used by Andy was not all that was available and leads them to comment on how the problem solving cycle could be revisited if different questions are posed.

Our approach to designing a marking scheme and allocating marks for the assessment involved adapting the grading scheme for UK A level coursework in statistics provided by the UK curriculum development organisation Mathematics in Education and Industry (MEI). This scheme allocates the assessment questions to domains for grading and uses a very simple mark allocation scheme. In our regime we use five domains: the first to allow for the holistic view of the problem solving approach and the remaining four to correspond to the four different stages of the problem solving cycle. Following suggestions by Garfield (1994, example 2) we allocated marks that correspond to the responses candidates make to each question being incorrect (0 marks), partially correct (1 mark) or correct (2 marks). Each of the five domains could also be given different weightings if an examiner so wishes.

The assessment was given to a range of students who had experienced the problem solving teaching/learning resources developed at the material development stage of the RSSCSE/QCA review. The assessment materials were intended to test the students' ability to approach statistics in a holistic manner.

The assessment was well-received by the students and staff who trialled it. In addition to the factors which led us to choose an online approach to the assessment, we feel, and have evidence that teachers support us in our view, that this form of online assessment has several educational advantages over paper-based equivalents. We believe that the assessment regime developed for this project has much educational potential for students of all ages.

Garfield (1994) stresses the need for assessments that measure the understanding of a PSA that can also be viewed as an integral part of the teaching and learning process. The

assessment we have produced is a prototype of a tool which could be used to judge the effectiveness of teaching materials at different levels. The structure and approach we have adopted is appropriate for a variety of ages and ability of students. The template could be adapted for a wide range of individual needs. It is our belief that there is potential for further developing the assessment and providing feedback for teachers to use with their students in different ways. There would be the possibility of merging responses to create a completed exam paper – accessible via paper or on screen. Similarly, there is the potential that feedback could be provided through checklist of skills which have been covered within the assignment. This is an aspect that would need to be developed further.

CONCLUSIONS

In September 2006 the UK Department for Education and Skills made a policy decision to abolish all coursework in the national mathematics examinations sat by 16 year olds. In their October 2006 report the MEI observed that

> "Effective teachers use a variety of methods to encourage interest and understanding but the current high-stakes results culture in education encourages less confident teachers to "teach to the test"."

Consequently there is a danger that there are now key areas of the English mathematics curriculum that will not be assessed and therefore may not be taught. For this reason it is important to have a replacement for the assessment of the part of the curriculum that was formerly assessed through the GCSE coursework.

In this paper we have presented a new assessment regime for grading school age students after being taught aspects of statistics through a problem solving approach. Statisticians are increasingly arguing for such an approach to be adopted in teaching at all levels: if this is done then the assessment methods used need to match the new way of teaching and learning. As problem solving itself involves a range of different levels of cognitive skills, the actual questions posed to students within the assessment need to take these into account.

Students of all ages and knowledge levels can benefit from skills acquired through problem solving. We shall be using similar templates to those developed for teaching, learning and assessing school-age students with those who need to be taught and assessed though solving more difficult problems. See, for example, Chatfield (1995) and Stuart (2003). The questions posed within each domain may depend upon more complicated scenarios. For example, the methods and techniques used in the 'process' domain of the assessment may well refer to cutting-edge statistics, and the questions posed would need to reflect this. Similarly the 'discuss' domain may contain questions that reflect decisions about the efficacy of large drug trials. The templates we have produced can, in principle, handle all these scenarios and we will be reporting their implementation elsewhere.

REFERENCES

Anderson, L. W., Krathwohl, D. W. (2001) *A Taxonomy For Learning, Teaching, and Assessing: A Revision of Bloom's Educational Objectives*; New York Longman.

Bloom, B. S., Englehart, M., Furst, E., Hill, W. and Krathwohl, D R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of college and university examiners. Handbook I: Cognitive Domain*. New York, Longmans, Green.

Chatfield, C. (1995). *Problem Solving: A Statisticians Guide.* (2nd Edition). London: Chapman and Hall.

Garfield, J. (1994). Beyond Testing and Grading: Using Assessment To Improve Student Learning. *Journal of Statistics Education [Online], 2(1).* *www.amstat.org/publications/jse/v9n3/stephenson.html*

Smith, A. (2004). Making Mathematics Count – The Report of Professor Adrian Smith's Inquiry into Post-14 Mathematics Education. London: The Stationary Office Limited.

Stuart, M. (2003). *An Introduction to Statistical Analysis for Business and Industry*. A problem Solving Approach. London: Arnold.

Appendix Table 1 Questions in the on line assessment *Getting the Best Deal*

Question	Task
A. Overview of the problem solving approach	
A0	Descriptions of the stages of the cycle must be dragged and dropped into the correct position.
A1	Explain why the process is called a cycle and why that description is important.
B. Ayesha's investigation Ayesha has to pay for her own mobile phone. She wants to use what she has learnt about the problem solving approach to achieve this. She has written up her investigation in rough using ten statements.	
B0	The ten statements must be dragged and dropped into their correct positions in the cycle.
With reference to the **'plan'** statements in Ayesha's investigation	
B1	Look at the four statements and then indicate whether someone else could read this plan and carry it out for themselves (YN)†.
B2	If 'no' to B1, describe what other information is needed (tb)‡.
B3	Indicate which statement states the problem Ayesha is trying to solve ('Don't know' is possible).
B4	Suggest another method to choose a day at random (tb).
With reference to the **'collect'** statements in Ayesha's investigation	
B5	What is the mean length of phone calls (tb)?
B6	Comment on choice of shop (tb).
B7	Would collecting more information on different deals help? Explain answer (tb).
With reference to the **'process'** statements in Ayesha's investigation	
	The process statements used by Ayesha are presented showing her calculations.
B8	Why did Ayesha use the number 30 in the calculations (tb)?
B9	Do you think it's reasonable to expect she'll use her phone the same amount each month (YN)? Explain your answer (tb).
B10	Are there other things you think she should do with the data (tb)?
With reference to the **'discuss'** statements in Ayesha's investigation	
B11	Would you advise Ayesha to change to the deal given that the calculations are correct (YN)? Why (tb)?
B12	Ayesha thinks that £56 is a lot of money and 840 minutes is a long time. It is based on data from one day. What might have happened on this day for it to be higher than normal (tb)?
B13	What could Ayesha do to check if her use on this day is similar to other days or not (tb)?

Question	Task
Andy's investigation Andy considers advertisements for two phones that only differ in the amount of time and number of texts included in the package. The two adverts are displayed.	
With reference to the **'plan'** stage of Andy's investigation	
C1	Andy's friends say that the deals are the same. Do you agree (YN)?
C2	Can you give an example of when one deal might be better than the other (tb)?
Andy decides to investigate his phone usage to identify the better deal. The candidates' task is to help him with his investigation. (Candidates are told their responses will be judged on the basis of their explanations, not on whether they choose the same course of action as Andy.)	
C3	What data do you think he should collect (tb)?
C4	Over what period of time should he collect data? Explain choice (tb).
C5	How could he collect the data (tb)?
A handwritten statement of Andy's problem and plan are then presented for consideration.	
C6	Why does he decide to choose a bill at random (tb)?
C7	Why are his mobile phone bills a good way to collect the data (tb)?
With reference to the **'collect'** stage of Andy's investigation	
A table of the number of calls, total call duration and number of texts for each of 31 days.	
C8	The call durations are in seconds and look messy. Suggest how Andy could make this better (tb).
With reference to the **'process'** stage of Andy's investigation	
	He decides to change the seconds to minutes and round them to the nearest half minute.
C9	Candidates must complete the calculations. (Data boxes are available for the answers. An on screen calculator, with instructions for use, is also available.)
C10	One row (presented) in the table doesn't look right. Can you spot why (tb)?
C11	Can you explain why it's happened (tb)?
C12	Is it a problem (tb)?
C13	Complete the frequency table for the number of calls made each day for the month. (Data boxes labelled 0, 1, …, 12 are there for the answers.)
C14i	Look at the chart (line graph of daily call duration). Mark whether it is helpful or unhelpful (or not sure) in deciding which package to use.
C14ii	As C14i for bar chart for number of calls per day.
C14iii	As C14i pie chart for number of calls per day.
C14iv	As C14i scatter plot of number of calls vs number of texts.
C15	Candidates are asked to undertake calculations for the number of calls made daily. (Data boxes labelled mean, median, mode and range are there for the answers.) On screen calculator.
With reference to the **'discuss'** stage of Andy's investigation	
C16	Candidates must interpret the mode. (tb)?
C17	How long is spent on the phone in the month (tb)?
C18	Which package do you think he should choose and why (tb)?
	Superimposed line graphs of number of texts and number of calls for a different month are presented. An anomaly is highlighted.
C19	Candidates asked for an explanation (tb)?
Referring back to the **'plan'** stage of Andy's investigation	
Andy had noticed that many phone deals seemed to suggest that the more texts sent the fewer calls made. Two sample deals are presented.	
C20	Write a plan for an investigation into whether **the more texts a person sends, the fewer phone calls they make.**

†Where YN is indicated there are 'Yes', 'No' and 'Don't know' check boxes.
‡Where 'tb' is indicated the candidates are invited to "Type your answers here".