GCSE maths

Summary of research programme

Overview

Our GCSE maths programme of research has been designed to further explore features of the sample assessment materials that have been developed by exam boards as part of the current reforms. This current design of the research programme is made up of 3 strands that ask the following questions:

Strand 1: Based on the views of mathematicians, what is relative mathematical demand of questions in sample assessment materials for the new maths GCSE?

Strand 2: Based on a sample of current year 11 students sitting the sample assessment materials, what is the relative difficulty of questions in the new maths GCSE?

Strand 3: Using students' descriptions of their approaches to tackling questions, are there different approaches to mathematical problem solving in the new mathematics GCSE sample assessment materials?

All three strands are scheduled to report at the end of April 2015.

Overview of research designs

Strand 1 – Comparison of item demand

Methodology

The methodology applied in this strand of the research uses a comparative judgement framework based on question demand. This means that mathematicians are presented with a series of pairs of maths questions and asked, for each pair, “Which question is the most mathematically difficult to answer fully?” The judges are asked to make this judgement many times for randomly paired questions. Based on these many judgements, a statistical model is fitted (the Rasch model) that places the questions on a scale from the least to most mathematically demanding.

The questions used for this strand of research are those from sample assessment materials for the new GCSE maths specifications, questions from the current GCSE maths papers and also questions from 12 international jurisdictions. This leads to a pool of over 2,000 questions. More than 40 judges will be asked to perform 1,000 judgements each.

Once these judgements have been made, the statistical model will allow an evaluation of:

- which questions are perceived to be of greater and lesser mathematical demand;
- how reliable the judgements were.

Limitations

This strand of research provides an evaluation of perceived demand of questions and does not, necessarily, reflect the difficulty of questions as experienced by students of the appropriate age and experience. Maths experts, by their nature, tend to see beyond any context within which the maths is set and/or any complexity introduced due to the specific numbers that appear in a question. This, in addition to the judgement that that judges are being asked to make, will be helpful in isolating the perceived mathematical demand from any other features.

This is, however, only part of the picture giving rise to the need for strand 2 (outlined below) that will access the actual difficulty of items.

Strand 2 – Pilot testing to evaluate question difficulty

Methodology

Whereas strand 1 is considering the perceived demand of questions, this strand will evaluate the actually difficulty of questions as experienced by students.

To evaluate the actual difficulty of questions, current Year 11 students will be asked to sit one of the question papers from the sample assessment materials as a mock exam. As an incentive to participate, the scripts will be marked and teachers provided with student and item level analyses following the study. Once schools have been recruited to participate and have nominated their candidates as higher or foundation tier, question papers from the different exam boards from the appropriate tier will be randomly distributed between the students. One non-calculator paper from each exam board will be included in this strand and will be sat under exam conditions.

A minimum of 500 students will be required to sit each question paper at each tier from each exam board. This number is necessary to achieve a level of randomness that will provide meaningful comparisons between the groups sitting the different papers. Following marking of students' scripts, analysis of the data will provide details of:

- the relative difficulty of questions as experienced by candidates;
- further evidence regarding the technical functioning of the assessments.

Limitations

The primary limitation of this strand of the research programme is that the students taking part in the pilot will not have followed the course of study relevant to the newly reformed specifications.

The potential effects are that students are not suitably prepared, leading to very low performance and/or disengagement with the exam. This could lead to floor effects with a large number of students achieving very low (or no) marks on a large number of items limiting the usefulness of the analysis. This may also affect different questions in different ways as some material will be new to the content.

While these risks are not insignificant, this cohort of students is the most representative group available and this potential risk will born in mind during the analysis phase.

The sampling of schools/students will not be nationally representative. Given the relative, rather than absolute, nature of the analysis, this feature does not

compromise the analysis. Attention will be paid to whether the profile of recruited schools may lead to wholesale floor or ceiling effects.

Strand 3 – Approaches to problem solving

Methodology

This strand of the research programme will explore how students go about solving questions. To elicit the thought process students go through, potential higher tier year 10 students will be asked to articulate how they would explain to a fellow student how to go about answering a particular question. These explanations will then be used in a comparative judgement design (similar to strand 1) with judges deciding which explanation demonstrates the best problem solving ability as encapsulated by the definition of A03. These judgements will then be used to place all students on a scale from those with the lowest to highest problem solving ability. Students will then be asked to respond to these questions in a normal manner and their responses will be marked. Analysis of the relationship between the judges' views of students problem solving ability with the marks students actually achieve on these questions will provide information on the extent to which items are measuring problem solving ability. Comparisons between exam boards' approaches to problem solving can then be made.

Limitations

This is a challenging research design and it is difficult to know if all students will be able to perform the explanatory task outlined above. To test this, we will be running a pilot phase to assess viability of the approach and will adjust the design accordingly.