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 four strands that ask the following questions:

- **Strand 1:** Based on the views of mathematicians, what is the 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:** Based on maths experts' views of students' model answers, which questions from the new maths GCSE sample assessment materials elicit the best mathematical problem-solving?
- **Strand 4:** Based on the views of maths experts, how do problem-solving questions vary across the sample assessment materials for the new maths GCSE?

All four 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 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, and how reliable the judgements were.

Strand 2 – Pilot testing to evaluate question difficulty

Methodology

Whereas strand 1 considers the perceived demand of questions, this strand will evaluate the actual 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 and further evidence regarding the technical functioning of the assessments.

Strand 3 – Judging model problem-solving responses

Methodology

This strand uses a comparative judgement framework, in a similar way to strand 1.

High-achieving Year 11 maths students will be asked to provide model responses to questions drawn from the boards' sample assessment materials that contain four or more AO3 (problem-solving) marks. The responses will include a description of the pupils' reasoning in coming to their answer.

Around 30 maths teachers will then be asked to judge which responses best exemplify problem-solving, given a specified definition. Their analysis will allow us to order the questions by the extent to which they elicit good problem-solving.

Strand 4 – Differentiating interpretations of problem-solving

Methodology

This strand will seek to identify any differences in approach to mathematical problem-solving across the new GCSE maths specifications.

The same questions as used in strand 3 will be presented to five maths experts in randomly selected triplets. They will be asked to pick the two questions that are most similar and to specify why the third item is different. This difference is known as a 'contrasted attribute'. The procedure will be repeated until no new attributes are revealed. The experts will then be asked to note those attributes they consider relevant and irrelevant to the definition of problem-solving. Each question will subsequently be independently rated by each expert with regards to the various identified attributes on a five-point scale.

This approach will highlight any differences between exam boards, and the dimensions across which those differences occur. The extent to which experts' ratings are consistent will be considered.