

Regulatory Summary
GCSE Maths Sample Assessment
Materials



May 2015

Ofqual/15/5691

Contents

- Summary of the research 2
- Summary of the research findings 3
- The actions exam boards are taking..... 4
 - Pearson..... 4
 - OCR..... 5
 - WJEC..... 5
 - AQA 6
- What teachers can expect to happen next 6
- Figure 1 7
- Figure 2 8

Summary of the research

Following the accreditation of the reformed GCSE maths specifications, concerns were expressed about differences in the difficulty of exam boards' sample assessment materials and in their approach to the assessment of problem solving. We conducted a programme of research to evaluate whether the concerns were valid and the differences sufficient to undermine the teaching, learning and assessment of GCSE maths.

The research comprised four complementary evaluations of exam boards' sample assessments.

Study 1: Compared the expected difficulty of all items (questions) from exam boards' sample assessments, including with items from their current GCSE maths papers and from similar qualifications internationally. PhD mathematicians judged pairs of items and indicated which item they thought was the most mathematically difficult. These judgements were drawn together to allow a comparison of expected difficulty across exam boards¹ and across jurisdictions. [See figure 1 below for comparisons between England's GCSEs.]

Study 2: A sample of 3,865 Year 11 students (15-16 year olds) sat the exam boards' (non-calculator) sample assessments. Random allocation of students to exam papers meant that the difficulty of items and whole assessments could be compared across exam boards. It was possible to examine whether the assessments would differentiate between students of different abilities [see figure 2 below] and whether they would support reliable grading (by making approximate adjustments for preparation and motivation).

Study 3: Employing the definition of mathematical problem solving used during the creation of the new GCSE maths qualifications, GCSE maths teachers judged the extent to which exam boards' items elicited problem solving. It was possible to identify the items which were judged to best elicit mathematical problem solving and to explore whether this was linked to item features such as expected difficulty.

Study 4: Five GCSE maths teachers took part in an in-depth analysis of the ways in which items testing mathematical problem solving varied across exam boards' sample assessments. The teachers took an entirely independent view of the items, first describing the dimensions upon which they varied and rating each item on those dimensions. Using these dimensions we were then able to compare across boards the approach taken to the assessment of mathematical problem solving.

¹ WJEC Eduqas papers were not included for reasons of timing

Summary of the research findings

In Study 1, we found that, overall, the sample assessments were judged to be more difficult than the current exam papers. The expected difficulty of the higher tier assessments was more in line with similar assessments in international jurisdictions. There was, however, a difference in difficulty across the boards. Our evidence suggests that AQA's assessments are of similar difficulty to their existing papers. These were judged the hardest of the three boards included in this study, when aggregated across papers (although these differences in difficulty are accounted for when papers are graded). However, their planned assessments exhibit lower expected difficulty than the other exam boards' sample assessments.

In Study 2, students' performance on the assessments highlighted a significant risk that the sample assessments from OCR, Pearson and WJEC Eduqas would fail to differentiate effectively across the full range of ability of the cohort for which the qualification is intended. This is due to the assessments being too difficult for students. Of course, it is difficult to predict what the impact on students' performance of good teaching, exam preparation and heightened motivation will be in 2017. However, even once approximate adjustments for these factors were made, there remains a significant risk that the assessments would not differentiate effectively between students of the range of abilities that are expected to study GCSE maths. Were this to be allowed to continue, it would be likely to undermine the reliable grading of students when assessments are sat for the first time in 2017.

Studies 3 and 4 focused on the assessment of mathematical problem solving. This is defined by an assessment objective (AO) summarised as 'Solve problems within mathematics and in other contexts' (AO3). Prior to our research, some stakeholders suggested that differences in exam boards' approaches to testing this AO were a source of perceived differences in the difficulty of the assessments. Studies 1 and 2 showed that this was not the case. Differences in difficulty stemmed from differences in items testing the two other assessment objectives measured in GCSE maths. These can be summarised as 'Use and apply standard techniques' (AO1) and 'Reason, interpret and communicate mathematically' (AO2).

Nonetheless, there were differences between the exam boards in their approach to testing AO3. WJEC's AO3 items were judged to be best at eliciting mathematical problem solving. Their items tended to require more open-ended answers, to require candidates to justify their answers and methods, to allow multiple approaches to solving problems, but also to require understanding of more demanding language. In contrast, in general, the judges believed that some of AQA's foundation tier items were less effective at eliciting AO3 responses compared to other boards' items.

The actions exam boards are taking

We have set out our intent that all exam boards review their sample assessment materials (SAMs) to achieve the following:

- We have confidence that live assessments set in 2017 are able to function well.
- The policy aims for reformed GCSEs are met.
- A more comparable level of difficulty between the assessments of the four exam boards is found, so that we are not relying overly on differences in the positioning of grade boundaries to achieve comparability.
- Their problem solving questions draw on the examples of good AO3 items in the research report and our proposed revisions to statutory guidance.
- Specimen materials are made publicly available that reflect each of these four aims, in place of those previously released.

Exam boards must have due regard to our research findings and have agreed to take the actions outlined below.

Pearson

There was clear evidence from the research that if the approach taken in the live exam papers matches the approach taken in the SAMs, there is a material risk that they would not sufficiently differentiate across the range of student ability and would not support the setting of reliable grade boundaries. On that basis, Pearson are redeveloping their specimen materials so that they are more likely to differentiate between students when the assessments are live. The problem of a lack of differentiation exists on both tiers but is more extreme on the higher tier.

We expect Pearson to make sure their higher tier papers continue to stretch the most able students by maintaining the level of difficulty of the most difficult questions but also to ensure that the assessments are accessible to the full range of ability of students entering this tier. We expect Pearson to modify their foundation tier assessments to ensure they are accessible to the least able students that sit the current foundation tier so that the new GCSE is accessible to the same proportion of students that currently sit GCSE maths.

When considering difficulty, Pearson will focus particularly on items testing AO1 and AO2. Pearson will scrutinise items with very low facility values and consider what changes could be made to make them more accessible without necessarily reducing the mathematical challenge. They will consider whether there are features of the item design that contribute to their difficulty and whether it would be possible to test the underlying mathematics in a more accessible way.

There is evidence in the report to illustrate high quality AO3 items which Pearson will use to inform item writing.

OCR

There was clear evidence from the research that if the approach taken in the live exam papers matches the approach taken in the SAMs, there is a material risk that they would not sufficiently differentiate across the range of student ability and would not support the setting of reliable grade boundaries. On that basis, OCR are redeveloping their specimen materials so that they are more likely to differentiate between students when the assessments are live. The problem of a lack of differentiation exists on both tiers but is more extreme on the foundation tier.

We expect OCR to make sure their higher tier papers continue to stretch the most able students by maintaining the level of difficulty of the most difficult questions but also to ensure that the assessments are accessible to the full range of ability of students entering this tier. We expect OCR to modify their foundation tier assessments to ensure they are accessible to the least able students that sit the current foundation tier so that the new GCSE is accessible to the same proportion of students that currently sit GCSE maths.

When considering difficulty, OCR will focus particularly on items testing AO1 and AO2. OCR will scrutinise items with very low facility values and consider what changes could be made to make them more accessible without necessarily reducing the mathematical challenge. They will consider whether there are features of the item design that contribute to their difficulty and whether it would be possible to test the underlying mathematics in a more accessible way. In particular, OCR will consider whether the wordiness of items is reducing their accessibility.

There is evidence in the report to illustrate high quality AO3 items which OCR will use to inform item writing.

WJEC Eduqas

There was clear evidence from the research that if the approach taken in the live exam papers matches the approach taken in the SAMs, there is a material risk that they would not sufficiently differentiate across the range of student ability and would not support the setting of reliable grade boundaries. On that basis WJEC are redeveloping their specimen materials so that they are more likely to differentiate between students when the assessments are live. The problem of a lack of differentiation exists on both tiers but is more extreme on the higher tier.

We expect WJEC to make sure their higher tier papers continue to stretch the most able students by maintaining the level of difficulty of the most difficult questions but also to ensure that the assessments are accessible to the full range of ability of students entering this tier. We expect WJEC to modify their foundation tier

assessments to ensure they are accessible to the least able students that sit the current foundation tier so that the new GCSE is accessible to the same proportion of students that currently sit GCSE maths.

When considering difficulty, WJEC will focus particularly on items testing AO1 and AO2. WJEC will scrutinise items with very low facility values and consider what changes could be made to make them more accessible without necessarily reducing the mathematical challenge. They will consider whether there are features of the item design that contribute to their difficulty and whether it would be possible to test the underlying mathematics in a more accessible way.

WJEC have created high quality items to test AO3 and should continue this good practice.

AQA

AQA's sample assessments were judged to be a little more demanding than their current exam papers. However, the research suggests that in a live setting, AQA's foundation and higher tier sample assessments would differentiate well and would be accessible to the full range of abilities.

We expect AQA to continue to ensure that their assessments will function well in 2017 whilst also lifting, to some extent, the expected difficulty on the foundation tier. AQA will scrutinise items with higher facility values and consider how the mathematical difficulty could be made more challenging without making the items inaccessible or damaging the ramping of the assessments. They will review AO3 items, particularly on the foundation tier, in light of the evidence in the research report. These improvements will ensure that AQA's assessments better reflect the policy intentions of Government at the foundation tier.

What teachers can expect to happen next

Exam boards will be submitting sample assessment materials to us in coming days which we will scrutinise. We have processes in place to review new materials from exam boards so that they can be in schools from the end of June.

Figure 1 - Box plots showing the relative expected difficulty, as judged in Study 1, of assessments from the current specifications and the sample assessments

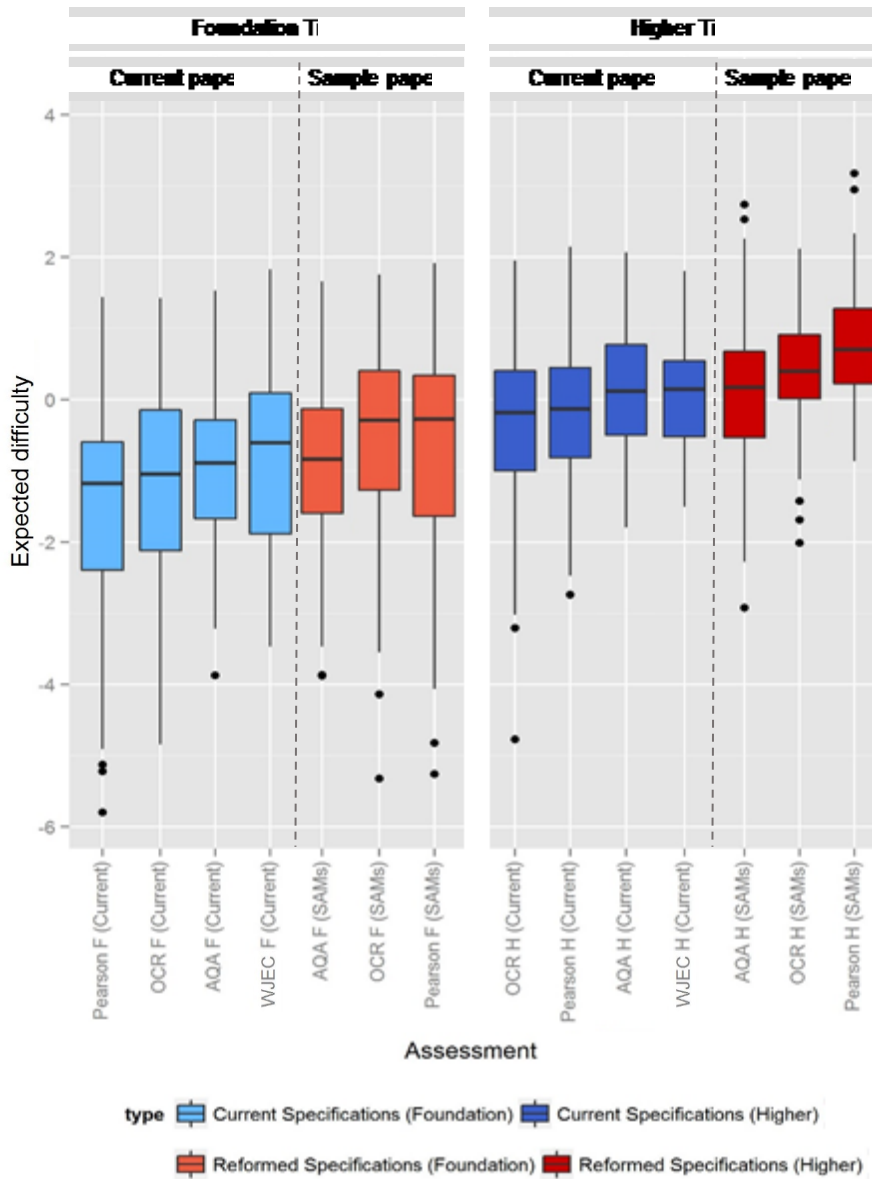
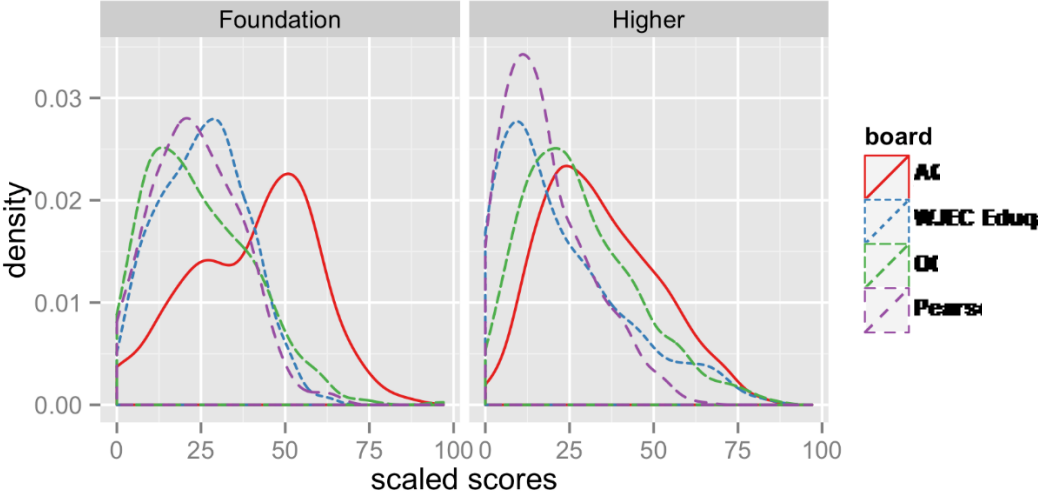


Figure 2 – Mark distributions for the sample assessment materials tested in Study 2



We wish to make our publications widely accessible. Please contact us at publications@ofqual.gov.uk if you have any specific accessibility requirements.



© Crown copyright 2015

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit <http://nationalarchives.gov.uk/doc/open-government-licence/version/3> or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: publications@ofqual.gov.uk.

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

This publication is available at www.gov.uk/ofqual.

Any enquiries regarding this publication should be sent to us at:

Office of Qualifications and Examinations Regulation

Spring Place
Coventry Business Park
Herald Avenue
Coventry CV5 6UB

2nd Floor
Glendinning House
6 Murray Street
Belfast BT1 6DN

Telephone 0300 303 3344

Textphone 0300 303 3345

Helpline 0300 303 3346