

Ofqual Board

Paper 9/15

Date:

20 May 2015

Title:

Strategy, Risk and Research Update

Report by:

Michelle Meadows, Executive Director, Strategy, Risk and Research

Paper for discussion and information**Open paper**
#####
Summary

1. This report updates the Board on work being conducted within the Strategy, Risk and Research Directorate. Often the work of the directorate underpins activities across the organisation and so is reported in other updates. However, to give the Board visibility of the scope and pace of activities within SRR we will provide this separate update biannually.

Recommendations

2. The Board is asked to note the range of work undertaken within the Directorate and the progress made.

Research**GCSE Maths Research**

3. Following the accreditation of the reformed GCSE Maths specifications, concerns were expressed regarding differences in the difficulty of exam boards' sample assessment materials and in the approach they had taken to the assessment of problem solving. Indeed, the latter was recognised

during accreditation and the diverse approaches were considered valid. However, stakeholders argued that the differences were a source of variation in difficulty.

4. Four complementary evaluations of exam boards' sample assessments are now complete. The first study compared the expected difficulty of all items from exam boards' sample assessments, including comparison with items from recent GCSE Maths papers and from similar qualifications from other international jurisdictions. The second study compared the actual difficulty of items from the non-calculator sample assessments. The third study compared the extent to which items were judged as eliciting the mathematical problem solving construct. The fourth study investigated the ways in which problem solving items varied across exam boards' sample assessments.
5. The research has been reviewed by an independent advisory group (Tina Isaacs, Alastair Pollitt and Chris Sangwin). The feedback we received was helpful and positive, and did not lead us to question the conclusions we had drawn from the research. It is important, however, to acknowledge the limitations of the work. The research was conducted on sample assessments and not live papers. Further, the students who sat the assessments had not been prepared for this qualification and were not as motivated as they would be in a live exam setting. As you would expect, our analyses seek to overcome these limitations and we have born them in mind in drawing conclusions. The exam boards have also conducted some research but nothing as rigorous as this programme.
6. Overall, experts believed the new sample assessments would be harder than the question papers from the current specifications. Indeed, the expected difficulty of the higher tier papers was more in line with that of similar international assessments.
7. While the experts expected AQA's current papers to be harder than those of OCR and in particular Pearson, they judged AQA's sample assessments to be easier than OCR's or Pearson's. In other words, the order of perceived difficulty has reversed in these new specifications. Indeed, the expected difficulty of AQA's new sample assessments was almost identical to that of their current papers. Modelling suggests that the difference in difficulty across the boards is comparable to that in the current papers and could therefore be dealt with in the setting of grade boundaries. However, given that we have this 'pre-test' of the papers it is questionable whether this would be the right approach given the possible wash back effects on teaching and learning.
8. The correlation between experts' judgements of the expected difficulty of papers and how hard students actually found them was surprisingly

strong. Indeed, students scored much better on AQA's non-calculator sample assessments than those sitting the other boards' assessments. Eduqas (WJEC) and Pearson higher tier papers were particularly difficult for students and there is a significant risk that there will be difficulties in setting reliable grade boundaries on papers such as these in 2017.

9. The research also showed differences between the boards in the quality of their approach to the assessment of problem solving (AO3). Eduqas's AO3 items were judged to be best at eliciting problem solving. However, there was no evidence to suggest that the differences in approach to problem solving contributed to differences in difficulty. Indeed, there was no difference in the difficulty of exam boards' AO3 items. The differences in difficulty lay in items testing AO1 (use and apply standard techniques) and AO2 (reason, interpret and communicate mathematically).
10. At time of writing, our regulatory response to the findings is under discussion. Exam boards must have regard to the research evidence and if in light of the evidence they create live exam papers that do not function then we can take regulatory action. We are concerned by the failure of the Eduqas and Pearson papers to sufficiently differentiate between students. These papers, and to a lesser degree those of OCR, need to be made more accessible to students otherwise there is a significant risk that the exam boards will not be able to set reliable boundaries. Boundaries may be too low and too bunched, and so there will be an increased risk of students being misclassified. We are also concerned that by not increasing the difficulty of their sample assessments relative to their current papers, AQA have not met the policy intention behind the reforms in maths and are an outlier. We will seek to have AQA increase the difficulty of their papers while not undermining their technical functioning.
11. Our experiences of GCSE Maths have led us to consider what extra steps we can take to support the accreditation of other key qualifications. The first study in which experts judged the expected difficulty of items was so successful that we plan to employ a similar methodology to compare the expected difficulty of the GCSE Science sample assessments submitted for accreditation this summer.
- Teacher Ethics in Assessment**
12. Part of our strategy in managing malpractice has been to influence the sector's thinking on ethical issues in assessment. On March 26 we held a joint symposium with Oxford University on Teacher Ethics in Assessment. This brought together teacher leaders, teachers, academics and exam boards to have an open exchange about the pressures facing teachers, evidence on the level of malpractice / dubious tactics, and what could be done to better support teachers faced with ethical dilemmas. The findings

of Ofqual's ethical survey of teachers were presented, allowing a nuanced and delicate treatment of this sensitive topic.

13. Materials relating to the day have been collected together to provide a valuable resource to those working in this area (<http://oucea.education.ox.ac.uk/events/invited-symposium-teacher-ethics-in-assessment/>). Board members can find a copy of the presentation on the findings of the ethical survey here. It is not our intention to publish the full research report but to refer interested parties to these slides which make clear the dangers of drawing firm conclusions from a survey of a small and non-representative sample of teachers.
14. While some stakeholders were initially concerned by our ethical survey and our plans to hold the symposium, these concerns have now dissipated. Feedback from those attending has been unanimously positive. HMC and ASCL in particular welcomed the seminar. ASCL had referenced the need for an ethical framework for teachers in their Blueprint for a Self-Improving System. The symposium spurred their interest and they are now working with CIEA to produce a framework. They have committed to keep Ofqual informed. We will continue to support this work such that the teaching profession retains full ownership.
15. As a follow-up to the ethical survey, later this month we will conduct focus groups with teachers of a range of subjects. The intention is to gather richer data regarding: how exam tactics might change once reformed GCSEs and A levels are introduced; where in the assessment lifecycle opportunities for malpractice arise; and which deterrents and sanctions might be most effective.

Regulating for Validity

16. Researchers have supported an organisation-wide programme of work to develop Ofqual's regulatory strategy in a post-QCF world. In particular, this has involved exploring the implications of the principle of regulating for validity. Much of this work has been analytical – intended to help Ofqual to present a narrative which is both informed by longstanding debates in the academic literature and tailored to the particular context of regulating qualifications in England. It has included the production of a new definition of validity and related technical terms; the elaboration of a generic qualification lifecycle – focused on the design, provision and evaluation of qualifications – with which to highlight critical features and processes related to validity; an analysis of qualification purposes and their relationship to validity; an exploration of the idea of qualification ‘types’ within the landscape of regulated qualifications. This work will continue over the next few months, as the regulatory strategy becomes more tightly defined, and will support communications with awarding organisations and other stakeholders.

Inter-Subject Comparability

17. The comparability of subjects is an issue that continues to concern stakeholders. We are forming a pack of evidence upon which to begin a discourse with stakeholders about the complexity of the issue and the pros and cons of the different responses. A literature review of research on the comparability of different GCSE and A level subjects is almost finalised, along with a research paper comparing the difficulty of subjects using Rasch analysis. Earlier drafts of these papers were considered by the Standards Advisory Group. We have also been writing a paper on policy – why this matter deserves our attention and what might be done – and another one on the approaches taken to this issue around the world. The intention is to publish these materials as a package and then to host stakeholder events in the autumn so that we can engage a variety of people with the question of what actions, if any, Ofqual might take. A paper recommending a way forward is due at the Board at the start of next year.

Other Research Highlights

18. In order to investigate how the Enquiries About Results (EAR) process might be improved a research project comparing different potential EAR processes has been conducted.
19. The annual ‘perceptions survey’ which explores attitudes to general qualifications (and vocational ‘equivalents’), as well as awareness and perceptions of Ofqual is complete. At time of writing the findings of the survey of over three thousand heads of schools, teachers, students, parents, the general public, employers and universities are confidential (subject to National Statistics reporting rules). The survey is due to be published by the end of May. We may be in a position to inform members as to the high level findings at the meeting.
20. Part of the strategy for building Ofqual’s research capacity and capability is to move away from outsourcing research to actively collaborating with external research experts. As part of this strategy a three year research collaboration with the University of Oxford has begun. This is to investigate the impact of qualification structures on teaching and learning, that is, a comparison of the impact of modular versus linear GCSEs on outcomes. The evidence as to the impact of modular and linear structures is surprisingly thin. The project will mean that more informed policy decisions regarding qualification design can be made, should interest in modular structures return. This project involves a review of the literature, interviews with teachers and large scale data analysis over decades. Working with experienced researchers including Jo-Anne Baird, Tina Isaacs and Gordon Stobart on this multi-method project will help up skill our research team, and deliver much needed evidence.

Standards

Ofqual Educational Assessment Seminar

21. For 2015 we changed the focus of what used to be called the Ofqual Technical Seminar. Previously the delegates/presenters were from exam boards or other regulators, and discussion were generally focused on operational issues relating to GCSE and A level awarding. This year's seminar was focused on more strategic issues and we invited external speakers and delegates to get wider and more expert input. We had representatives from Cambridge Assessment, Cambridge University, STA, NFER and Qualifications Wales. Presentations and discussions focused on three key themes – setting standards in reformed qualifications, quality of marking and malpractice. Including external organisations had a positive effect and the discussions benefited from alternative perspectives. The 2016 seminar is already being planned.

Review of summer 2014 GCSE/A level awarding documentation

22. The Code of Practice sets out detailed procedures that exam boards must follow when setting grade boundaries. Nevertheless, recent discussions with boards about particular awards alerted us to possible differences in the way those procedures are operationalised. So we are reviewing the awarding documentation for a sample of summer 2014 GCSEs and A levels to see whether there are systematic differences in the way grade boundaries are set. We are considering, for example, how senior examiner recommendations are used, what range of marks is considered. We plan to report our findings to exam boards ahead of the summer 2015 awards. The findings will inform our approach to regulation this summer and our work to develop high-level General Conditions about awarding/grading to replace the Code.

Evaluating inter-board screening of grading standards at GCSE and A level

23. Following each summer series, JCQ carries out statistical inter-board screening exercise to compare the grade standards between exam boards in a subject. This comparison is based on a student's average GCSE performance and is often held up as the best way of statistically comparing standards. It is therefore a key evaluatory tool. Indeed, we are currently exploring how IGCSEs can be included in this exercise in future. However, the data from this exercise is sometimes at odds with other evidence regarding comparability e.g. KS2 predictions. This project is documenting the strengths and weaknesses of the approach so that we can better understand its limitations.

Evaluation of question re-use in A level Government & Politics

24. In autumn 2014 we were alerted to a blog commenting on an example of an essay question (and mark scheme) from a previous exam series being re-used in a summer 2014 Government & Politics paper. We were concerned that this could advantage or disadvantage some students, and that it could undermine confidence in the qualification. We reviewed a number of A level Government & Politics papers from AQA, OCR, Pearson and WJEC and found varying numbers of questions being re-used. We reported our findings to exam boards and we will be following up this issue with them.

Quality of Assessment Metrics

25. Work to collect metrics on the quality of GCSE and A level assessments continues. While these projects are currently centered on General Qualifications, focus will widen to appropriate Vocational Qualifications over time. There are two strands – one focused on quality of marking, the other on internal test reliability. Currently, there is a reliance on EAR data to measure quality of marking. This is unfortunate as grade changes following enquires are a flawed measure. Enquiries are targeted at key boundaries and there is evidence to suggest that whether a candidate's grade changes or not is influenced by the re-marking examiner's approach to the review. Having more robust metrics will allow meaningful comparison across exam board's specifications in the same subject. Outliers with low reliability will be the focus of regulatory attention. This project will report in June and the recommendations discussed at the Standards Advisory Group in the autumn.
26. For the first time following this summer's exams, boards will report coefficients measuring the internal reliability of their exams. Internal reliability is important because it affects the reliability of grading. A test with low internal consistency may need to be longer to ensure that candidates' grades are reliable and are not unduly impacted by the subject content that happens to be tested. Sometimes a test with high internal consistency could be shorter, although the wash back effect on teaching and learning needs to be considered. We will compare the internal reliability of exams across exam board's specifications in the same subject. Again outliers with low reliability will be a source of attention.

Awarding the Reformed Qualifications

27. Work has gone into plans for awarding the new GCSEs beyond summer 2017 and how the comparable outcomes approach should be applied to the new AS/A levels. This is reported in a separate Board paper and so will not be reported here.

Summer 2015 Awarding

28. Preparations for the summer awards are underway and are reported fully in the update from the General Qualifications Directorate (paper 7/15).

Strategic Policy and Risk

Regulatory Strategy

29. The Regulatory Strategy Statement was published alongside the Corporate Plan at the end of March. This sets out for the first time in one place the fundamentals of our regulatory approach. As we implement our strategy to make sure the qualifications we regulate are sufficiently valid, this statement provides a useful explanation of our regulatory tools and how we currently use them. We will review and refresh this statement as our approach develops.

Risk

30. We include in the Statement a high level explanation of our risk framework. We are embedding a more systematic approach to evaluating intelligence coming into different teams in Ofqual about awarding organisations and specific qualifications alongside our current risk indicators. This is being used to commission audit activity. This includes a current round of audits for a group of awarding organisations offering high impact qualifications of the systems and processes in place to make sure qualifications have clear objective and user support and that assessment is fit-for-purpose and can be delivered. These audits will be completed during May.
31. The risk team is also currently working with a small group of awarding organisations to understand the different approaches they take to manage their financial resilience. It is intended that this evidence, alongside analysis of the potential impact of longer term funding cuts particularly in the adult skills budget on specific awarding organisations, will enable us to implement a proportionate approach to monitoring financial viability.
32. The approach to corporate risk management including how we report and manage strategic risks is strengthening as we roll out guidance across teams on how to assess, manage and escalate risks and provide a common template for risk registers to support this. The common template strengthens the link between risks and corporate objectives, and focuses on key controls that Ofqual puts in place to manage risks. Following on from the Board Strategy Day at the end of February we have also taken steps to develop a risk appetite statement and aim to bring this for discussion at the Board over the summer.

Markets

33. Following the publication of the Frontier Economics report on the impact of GCSE and A level reform on the operation of the market, we identified a number of specific threats to standards and confidence. These include the design and content of marketing materials, the development of new products (which might, for example, increase the predictability of exams), whether fees would be known sufficiently in advance to meet schools and colleges planning needs, and whether any increases in fees would be defensible.
34. Marketing materials are being monitored, with a particular focus on GCSE Maths and English Language. We have completed a first round of 'mystery shopping' where we have commissioned subject experts who are also teachers to attend exam board training events for reformed GCSE Maths and English Language. The shoppers have reported back on their experiences and on how the GCSEs are being promoted. We are rolling this out to future events for other subjects.
35. Exam boards are experiencing high uncertainty regarding future market share and delivery costs. We met with them individually to explore their strategies for managing this uncertainty and the finalisation and advance publication of fees. None of the exam boards have yet published fees for the reformed qualifications due to be taught from this September. They said they are waiting for more certainty, for example, regarding the moderation arrangements for spoken language. Up until the election, the possibility of recoupling A levels will have also been a significant uncertainty. We now expect them to publish fees shortly and before the end of the summer term. They are forecasting a range of different increases across different subjects to recover the cost of development as well as higher ongoing delivery costs, in particular marking (due to the reduction in teacher assessment). We are sense checking these increases against the changes in assessment, the costs of development and against a backdrop of a reduction in income from re-sitting, early entry, modularity and so on.

Paper to be published	YES With the section on GCSE Maths withheld until the research findings and regulatory plans are public
Publication date (if relevant)	After the meeting