

Ofqual Board

Paper 53/15

Date:

18 November 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. The work of the Strategy, Risk and Research (SRR) Directorate underpins activities across the organisation and so is often reported in other updates. However, the biannual report aims to ensure that the Board has full visibility of the scope and pace of activities within SRR.
- Recommendations**
2. The Board is asked to note the range of work undertaken within the Directorate and the progress made.
- Strategic Policy and Risk**
- Strategy**
3. The Board provided valuable feedback and challenge to our regulatory and organisational strategies at the October Board strategy day. We have carefully considered that feedback. In particular, we recognise the Board's concern that our regulatory strategy properly reflects a focus on gathering and assessing evidence of validity.

4. In our thinking we have found it useful to distinguish between a regulatory focus on the existence and following of processes (e.g. does the AO have a process for moderating teacher assessment and is that process followed?); the outputs of processes (e.g. does the process of moderation create the correct rank ordering of candidates?); and the outcomes of the qualifications (e.g. how well does the qualification support progress to employment as intended?).
5. The first two categories of process related work can be more or less strongly related to validity. We recognise that auditing, investigating and enforcing conditions not directly correlated to validity will sometimes be necessary - for example, those conditions related to financial probity – but this work should not form the bulk of our activity. A blended approach is necessary, both to meet our statutory objectives and the expectations of our stakeholders. A blended approach will also require us to use the full range of evidence gathering and regulatory tools available.
6. We recognise that we must be clearer as to our approach to enabling and aiding compliance with the conditions. We have produced guidance in relation to some but not all conditions, and we will consider the gaps in that guidance.
7. We also recognise that we have a wider role in informing the education and training system, for example in the quality of marking area. Indeed, there will be occasions where we will adopt ‘shine a light’ and influencing techniques rather than focusing on AO compliance with conditions.
8. The strategies for regulating GQs and VQs will be refined in light of this thinking. These will be underpinned by strategies for research and evaluation, stakeholder management and communication, and corporate services strategies for people, resource and information management.
9. Our efficiency objective will be addressed by a new efficiency strategy, also being developed now as part of our overall strategy work. We have found that the strategy development process has prompted valuable consideration of each area – and others too, such as regulatory burden.
10. The strategies we are developing now will directly inform business plans and associated financial / human resource plans for each team in the organisation. These culminate in the next iteration of our Corporate Plan, to be published at the end of March 2016. We will use the February Board Strategy Day to capture the Board’s input to the Corporate Plan.

Risk

Entity Risk

11. Our bi-annual review of entity risk profiles is underway, with revised profiles to be completed by the end of November. These provide information to inform the planning of future regulatory activity. We continue to review and enhance these profiles, and at our conference in December we will consult with AOs on the indicators we use to inform risk profiles.
12. We have previously committed to issuing risk profiles to each respective AO, and intend to do this in the new year. We are preparing now the necessary communications to minimise the risk of unintended consequences.

Risk Appetite

13. Following the Board's previous discussions we have continued to refine our approach to categorising and describing risk appetite. We found that the approach used by HM Treasury did not describe well the categories of appetite that were relevant to us in Ofqual. We will use our proposed updated categorisation to redraft our risk appetite statement, and also consider ways to build this into our corporate planning and strategic risk management.

Systemic Risk

14. Our approach to capturing and reporting strategic risk is well developed and regularly refreshed and improved. We currently do not, however, have a process for the regular capture and reporting on systemic risk, that is events or changes in circumstances that could impact on multiple AOs or multiple qualifications, in ways that could affect our ability to meet our statutory objectives. Of course, a systemic risk may also be strategic.
15. We are now collating and analysing systemic risks to create the first draft of a new systemic risk register. The intention is that the Board will consider this bi-annually, with a first iteration to come to the Board in the new year.

Financial Stability

16. We are in the policy development stage of a new approach to monitoring financial risk in awarding organisations. Following a report we commissioned from Deloitte, we are developing a new approach that proposes all AOs provide a base level of financial and other information about their business planning processes, with certain AOs providing enhanced information where the base level of information shows relatively higher financial risks. The Board will receive a paper outlining the proposed policy and approach in January, prior to consultation.

Research

Inter-Subject Comparability

17. The Board considered our work on inter-subject comparability at the October strategy day. The six working papers that we are planning to publish before the end of the year are:
 - I. Comparability of Different GCSE and A level Subjects in England: An introduction
 - II. Inter-Subject Comparability: A Review of the Technical Literature
 - III. Inter-Subject Comparability of Examination Standards in GCSE and GCE
 - IV. Inter-Subject Comparability: International Review
 - V. A Recent History of Regulatory Perspectives on Inter-Subject Comparability in England
 - VI. Exploring Implications of Policy Options Concerning Inter-Subject Comparability
18. The papers will be supported by introductory material setting out the intention of the work, links to other resources (e.g. CEM reports, and previous QCA and Ofqual publications), and information on how to engage with Ofqual and so on. We will make clear that this is not a formal consultation exercise but an opportunity to discuss an important, complex topic before we come to a policy position. The plan for stakeholder engagement will be considered by SAG, a key part of which is likely to be a stakeholder conference in the spring.
19. Two other pieces of research continue. One is an outline of public debates on inter-subject comparability over the past two decades. The other investigates the extent to which concerns about inter-subject comparability influences qualification choice. These will provide additional material to present at the conference and will be published with a report on the latter.
20. We will be a position to propose a policy position to the Board in May.

Measuring the sawtooth effect in GCSEs and A levels

21. It is believed that when qualifications are reformed, student performance drops in the first year of testing and then improves over subsequent years as teachers and students become increasingly familiar with the assessment arrangements and as materials (e.g. past papers) become more available. This is commonly referred to as the sawtooth effect. The size and duration of the effect is unknown. It is assumed that at some point during the life time of a specification, improvements in performance due to test familiarity recede. At this point any improvements in performance reflect changes in the underlying ability of students.

22. Being able to estimate the likely sawtooth effect in the reformed qualifications will enable us to better consider our approach to overseeing awarding, and to plan whether and when we could safely use the National Reference Test outcomes to inform awarding. Hence, a programme of research into the sawtooth effect is underway.
23. So far, early and tentative findings suggest that it takes around three years for students and teachers to become familiar with new specifications and assessments. Following this period, improvements in performance are likely to be gradual and small – but the full programme of research needs to be completed before we can draw conclusions.

GCSE Science

24. To support the accreditation of GCSE Science a comparative judgement exercise was conducted in which science teachers compared pairs of items from the sample assessment materials and from current exam papers. Teachers were asked to judge which out of the pair of items was more difficult for a student to score maximum marks on. This kind of approach was successfully used in our work on GCSE Maths.
25. As we anticipated, for GCSE Science the correlation between the judged expected difficulty and the actual difficulty of items was lower than in Maths. Moreover, the gap between expected and actual difficulty varied by exam board. Analysis showed this could be partly explained by superficial features of the items, for example, multiple choice questions tended to be judged as less difficult than they actually were.
26. This limited the extent to which we were able to use the data to support judgements of the first submissions. Nonetheless, the data was presented to the panel members for discussion, both where it supported their own judgements, but particularly where there were discrepancies between the data and the view of the panel. The discrepancies were considered and assessments checked over to ensure the panel was confident in its view.
27. The study was valuable in illuminating some of the factors that can bias judgements of likely difficulty (in all likelihood including the judgements of accreditation panels). Indeed, once these factors are controlled for, the relationship between expected and actual difficulty was more similar to that achieved in GCSE Maths. We are currently considering how best to use the information on bias and whether there is value in collecting more judgements in relation to the second submissions of specifications.

Quality of Marking Metrics

28. The report on potential quality of marking metrics is complete and the findings will shortly be presented to SAG. This autumn we will be calculating a variety of metrics across GCSEs and A levels for all exam boards. Where there are differences in metrics within qualifications but across boards, we will analyse whether that is a feature of the assessment design, the data the metric is based on, or real differences in the quality of marking. We will consider whether these metrics could be aggregated to give meaningful and appropriate indicators of quality that could be published without having perverse consequences.

Dissemination

29. Dissemination of our research work is a key element of our strategy for several reasons. It builds Ofqual's credibility and authority, and moreover it provides a mechanism for receiving critical feedback on our research and allows researchers to develop their research careers, aiding recruitment and retention. What is more, as the weakness of the evidence base with regard to assessment is a long term risk to the quality of the qualifications system, it makes sense to promulgate research that can be used to improve assessment.
30. Arising directly from the Teacher Ethics in Assessment symposium held in conjunction with Oxford University, a proposal for a special journal issue focused on research into teacher ethics has been submitted to the Oxford Review of Education. This will stimulate the publication of research into this topic and a research paper on Ofqual's survey of teachers' experience of and attitudes to methods of maximising results will be submitted.
31. At the conference for the International Association for Educational Assessment, we ran a discussion group on inter-subject comparability. This showcased our work and allowed us to hear from other countries whether this issue causes concern and how it is tackled. We also presented our approach to regulating validity here. At the European Association for Educational Assessment we ran a pre-conference workshop on this topic, as well as presenting work on quality frameworks, GCSE Maths comparability, the comparative judgement methodology and our research into EAR marking.

Standards

Summer 2015 Awarding

32. Awarding was relatively straightforward in that there were very few changes to qualifications. We challenged a handful of out of tolerance awards but we did not escalate any. Overall results were stable. There was some concern expressed regarding awards in A level MFL and IGCSE English, and these are reported on below. We are now preparing for November 2015 GCSE English and maths awards and for summer 2016 with first reformed AS awards.

A level MFL

33. Following our report in September 2014 we asked exam boards to make changes to their assessments so they were better able to discriminate between students at the top end. We speculated that if these changes were successful, a consequence might be more A* grades, but we were clear that that was not our aim. We did, though, amend the reporting tolerances for A level French, German and Spanish so that the A* rules did not counter any effects of the revised assessments.
34. There was some evidence that assessments were more demanding than in 2013 (the basis of the analysis in the 2014 report) in that grade A boundaries were set at higher marks. However, in some cases grade A boundaries were steady. We await the outcome of the boards' own analyses of the way the 2015 assessments functioned, due in November.
35. We have also conducted analyses to assure ourselves, as far as is possible, that the use of statistical predictions in awarding has not, over time, made A level MFL grading more severe (as linguistically less able students have increasingly chosen not to study MFL at A level).

IGCSE English

36. In May CIE sent us an analysis which suggested that IGCSE First Language English had been leniently graded in 2014, particularly at grade C. It is normal for AOs to routinely review the outcomes of their grading process and to make fine adjustments where appropriate. They proposed to tighten their grade standards and we monitored their grading process closely.
37. The grading of this specification is made more complicated because there are two separate qualifications, with one available to state schools in England and another taken mainly by independent schools and overseas centres. These share some assessments. In the former, specification entries were over 200,000 – twice that of 2014.

38. CIE did tighten their outcomes at C and, to a lesser extent, at A. Following the issue of results there was a good deal of concern expressed by schools that were unhappy with their results. We have evaluated available technical evidence of the award. This suggests the award was slightly more severe than in 2014. Although it was impossible to evaluate the change precisely, there was no evidence to suggest that CIE had tightened the award beyond what was intended and beyond the usual adjustments that exam boards routinely make on the basis of post-award evidence. This supports similar analyses conducted by Cambridge Assessment.
39. CIE have recently issued very detailed explanations to HMC and to other stakeholders, and they have sent us a copy of their initial analysis of the outcomes. They have concluded that the grading was carried out appropriately, although there is evidence that the coursework option is now more harshly graded than the non-coursework option. We have concluded there is no case to re-open the award and we will continue to keep a close eye on November and June 2016 awards.

Double entry for GCSE and IGCSE English

40. Following the increase in entry and concerns over grading of CIE's IGCSE first language English, we have looked at multiple entry in GCSE/IGCSE English/English language. The number of candidates has declined since 2013 (65,000 in 2013 compared to just over 9,000 in 2014 and 2015), largely as a result of the performance table changes where only the first grade counts, and now represents just over 1% of the entry. Most (over 90%) of candidates who are double-entered are entered for a GCSE and an IGCSE, as opposed to two different GCSEs.
41. So far, the analysis shows that the distribution of candidates' best grades is largely similar to the overall distribution, although there is a greater concentration at C – the proportion of double-entered candidates achieving A*-C is 73%, compared to 64% for all candidates.
42. The proportion of double-entered candidates who achieve the same grade in both has remained relatively stable since 2013 (39% in 2013 and 2015). However, there has been a shift in the balance between the better grade being in the GCSE or the IGCSE. In 2013 and 2014 the proportions were roughly equal with 30% getting a better grade in GCSE and 31% and 32% respectively getting a better grade in IGCSE. In 2015, only 19% got a better grade in GCSE while 43% got a better grade in IGCSE.
43. We do not yet know why this change has occurred and analysis continues. It underlines, however, the difficulty of evaluating the appropriateness of the IGCSE award. We will consider the best way to disseminate our findings.

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

44. Following each summer series, JCQ carries out a 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. We have compared this approach to a variety of alternative statistical methods. We found standards were statistically aligned across boards and there was very little difference in the outcomes of the different approaches. However, there was some evidence to suggest that higher grades were more comparable than the lower grades. We will ensure exam boards consider why this may be so.

Paper to be published	YES
Publication date (if relevant)	After the meeting