

# Ofqual Board

## Paper 25/16

**Date:**

27 July 2016

**Title:**

Strategy, Risk and Research Update

**Report by:**

Michelle Meadows, Executive Director, Strategy, Risk and Research

**Responsible Director:**

Michelle Meadows, Executive Director, Strategy, Risk and Research

**Paper for discussion/information**

**Open paper**



**Issue**

1. This paper updates the Board on the Strategy, Risk and Research (SRR) Directorate's key work since the last Board meeting.

**Recommendation**

2. The Board is asked to note the range of work undertaken within the Directorate and the progress made. The Board is also asked to consider the Strategic and Systemic Risk Registers contained in Annexes A and B. Finally, the Board is asked to endorse our approach to continuing work on inter-subject comparability as described in paragraphs 17 to 20.

**Regulatory Strategy**

3. Board members are invited to consider a separate discussion paper on our regulatory strategy. The outcomes of the Board's discussions in the meeting will be used to inform an update to our strategy to be published in the autumn.

### **Strategic Risk**

4. In May the Board provided useful feedback on the format of the Strategic Risk Register. We noted the Board's wish that risks be more succinctly reported, with clearer direction on where to focus attention. We welcome feedback on the new format with a view to making further enhancements. Board members are also asked to review and comment on the current risk position as shown in the register attached at Annex A.
5. The direction of travel is largely positive. In part, the high number of changes to ratings reflects our deeper risk review and clarification of risk definitions. However, it also reflects successful mitigating actions, for example, for risk 15 related to accountability we are bearing the fruit of our stakeholder management.
6. The Board will note two newly developed risks. Risk 17 relates to technical education and the related policy development is set out in a separate board paper. Risk 18 is that ineffective implementation of new rules for the review of marking and appeals undermines public confidence in general qualifications.
7. We propose merging the people and accommodation risks, 2 and 4 respectively, due overlap in the definition and consequences. We also propose closing risk 3 'Replacing the Regulatory IT System' now that the portal is live and the project is approaching completion. Residual risks will continue to be managed at directorate level.

### **Entity Risk**

8. We collect and compile data for each regulated entity to inform judgements of their risk. We have previously committed to issuing entity data sheets to each awarding organisation (AO). Throughout the project, we have noted and sought to mitigate significant risks, particularly that issuing risk assessments would drive the wrong behaviours, for example, that AOs would choose not to submit event notifications to improve their ratings.
9. In light of these risks, the new Chief Regulator paused the project while we reviewed the practice of other regulators. We found the practice of issuing risk information was not unique but it was uncommon. It was more usual to publish the factors that contribute to a regulator's assessment of risk. We therefore intend to take this approach as part of work to refresh our risk framework this year. We plan to publish this information in tandem with our updated regulatory strategy.

### **Risk Analysis**

10. In developing our approach to risk we have designed and piloted a new risk analysis process. This involves making a multi-disciplinary assessment of potential risks related to a specific issue

that has arisen from intelligence. Colleagues from across the organisation contribute assessments of risk arising from the specific qualification lifecycle, how the relevant qualifications are delivered, market analysis and policy changes. We have completed the first of these analyses to look at short course vocational qualifications. This is now informing commissions for further detailed assessment of compliance. We are now using this approach to analyse risks related to current apprenticeship qualifications.

### **Systemic Risk**

11. Systemic risks are “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”.
12. In February we reported to the Board the outcomes of a comparative judgement (CJ) exercise to understand relative perceptions of impact and likelihood for a range of systemic risks. Since then we have created a Systemic Risk Register which is attached as Annex B.
13. The register includes an assessment of our organisational grip on systemic issues through current work. Organisational grip is an evaluation of our ability to act on and influence each systemic risk. The register also links to our Strategic Risk Register to improve understanding of our grip on the risks raised.
14. New systemic risks have arisen since the CJ exercise was undertaken. New emerging risks will be reflected in the next iteration of the register, which will be reported on a six monthly basis starting in November. We will also begin to focus on proactive identification of new systemic risks through enhanced risk and scenario analysis.
15. To note, a systemic risk may become a strategic risk if:
  - the risk may affect our ability to meet our corporate plan commitments in the next 12 months; and / or
  - the risk warrants continual Strategic Management Group and Board oversight.
16. The Board are asked to note the continued development of this work and are invited to comment on the work we have underway to address systemic risks. Members will be invited to contribute to the refresh of the register in September.

## **Standards**

### **Inter-subject comparability**

17. At the May Board meeting members considered paper 7/16 which provided an update on the evidence gathered about inter-subject comparability in GCSEs and A levels, and indicated a preferred policy position for discussion. At that time, we were still working on the implications of an analysis that had been prompted by a letter we had received in April from prominent science organisations. A paper based on that within-subject value-added analysis was considered by Ofqual's Standards Advisory Group (SAG) at its meeting on 1 July.
18. SAG's discussion has confirmed our view that this new analysis should be explored further to see whether it might provide either a compelling case for the adjustment of grading standards in some A level subjects (the sciences and modern languages) or even a justification for a wider re-alignment of A level grading standards.
19. Our plan is to involve some members of SAG as we continue that analysis over the summer, to provide a public statement in September so that it is clear what position we have reached following the conference and survey held earlier this year, and to meet with some of the science community to consider the implications of the within-subject value-added analysis for science A levels. We would want to consider carefully both the validity of the analysis itself and what it means for how we define A level standards, as well as the implications for the system were we to apply grading corrections based on the data. A possible step after that would be to have a similar meeting with modern linguists.
20. If the Board is content with the approach described above, we will have a further discussion at the SAG meeting in October and provide a substantive update for Board members in November or at the latest in January 2017.

### **Summer 2016 awarding**

21. At the end of May we published an open letter to schools and colleges<sup>1</sup> alongside provisional entry statistics for summer 2016.<sup>2</sup> Although these data are not complete, we collect them at the same time each year, which means we can make meaningful comparisons. Overall entries for 2016 are generally stable compared to 2015, with some notable exceptions.
22. For GCSE, overall entries are stable but there are some changes at subject level. Subjects that count for the EBacc and/or specific progress/attainment 8 categories have entries that are increasing,

---

<sup>1</sup> <https://www.gov.uk/government/news/ofqual-publishes-open-letter-and-figures-on-summer-2016-exams>

<sup>2</sup> <https://www.gov.uk/government/statistics/summer-2016-exam-entries-gcses-level-1-2-certificates-as-and-a-levels-in-england>

while other subjects are declining. For example, the following subjects have increased entries: science (22%), additional science (11%), separate sciences (5 to 6%), geography (8%), history (7%) and computing (87%). The significantly increased GCSE science entries are probably because progress/attainment 8 incentivises a shift back from BTEC Applied Science to science GCSEs.

23. In contrast, these subjects have decreased entries: citizenship (26%), D&T (10%), performing/expressive arts (11%) and ICT (28%). MFLs are the exception to the trend. Spanish entries are up by 3% but both French and German entries are down 8% despite counting for EBacc and progress/attainment 8. Given that the more traditional subjects tend to be those that students find more challenging, outcomes may be lower in these subjects and overall national results may reflect this.
24. Entries for CIE's IGCSE in first language English are relatively stable, with an increase of 2% compared to 2015.
25. Entries for reformed AS subjects are lower than in 2015. We expected this, although we were unsure of the likely size of the decrease. Overall entries from Year 12 students for the new AS subjects are down by just over 20% but there are some differences between subjects. For example, entries for English language and literature are down by 32% but computing entries are only down by just 10%.
26. In June we published the data exchange procedures for 2016 as a Regulatory Document.<sup>3</sup> This sets out the detail of the predictions to be used by exam boards. As usual, we will receive and review data from all awards, but we will be paying particular attention to the following: new AS, A level MFL and CIE's IGCSE first language English.

#### New AS

27. The Board has previously agreed that we will maintain existing grade standards in new AS (and new A levels in 2017), using predictions based on prior attainment to guide awarding decisions. Although it will be difficult to awarders to compare work on the new AS to previous work, it is important that awarders check the plausibility of the grade boundaries suggested by the predictions. We have been working with the exam boards to operationalise that check in a consistent way, so that all boards are starting from a common position in their scrutiny of candidate work.
28. Appendix 3 of the data exchange procedures sets out the principles exam boards should follow in these new AS awards. As part of our monitoring, we will be attending 11 AS awards in 6 subjects (biology, chemistry, physics, computer science, psychology and English

---

<sup>3</sup> <https://www.gov.uk/government/publications/data-exchange-procedures-for-a-level-gcse-level-1-and-2-certificates>

literature) across the four exam boards, to observe how they put these principles into practice.

29. We expect to see more year-on-year variation in individual school and college results in new AS. We are planning to publish information in early August on previous levels of variation in AS, as part of our communications around summer awarding and results. We will then publish similar information for the new AS on results day.

#### A level MFL

30. For 2016, we have re-baselined the predictions, to use 2015 outcomes as the reference year. We hope this will avoid inadvertently 'undoing' any impact that the changes to the 2015 papers had on overall results.
31. Following a commission from GQ colleagues, earlier in the year we reviewed the live A2 question papers and mark schemes for AQA, OCR and WJEC (the three boards whose assessments were identified in our September 2014 report<sup>4</sup> as insufficiently demanding at the top end of the ability range). We identified some changes that could be made to the mark schemes for summer 2016 that had the potential to improve the way the papers discriminate between good and very good students.
32. In response to our native speaker survey, we have had returns from 620 schools, out of a total of 2,400 with entries this summer. The response rate is disappointing, as it will limit any conclusions we can draw. However, we will be able to investigate the impact of native speakers on the predictions for these centres. We are currently entering the data and will report our findings in the autumn.

#### CIE IGCSE first language English

33. As in recent years, we will monitor the grading of this qualification closely this summer, due to the size of the entry and the challenges in awarding in 2015. We also expect that there will be close scrutiny of the results from schools and teacher associations. We have telephone meetings already scheduled with CIE at appropriate points in the grading process.

#### **Summer 2017 awarding**

34. We have previously presented to the Board details of the risks we have identified around summer 2017 awarding, and we have started to discuss with exam boards the detailed arrangements that we will need to put in place. This includes a two-day Maintenance of Standards/STIG meeting in November with responsible officers and technical representatives. We will finalise the agenda with exam boards in September.

---

<sup>4</sup> <https://www.gov.uk/government/news/improvements-to-be-made-to-a-level-foreign-languages>

### **Other standards work**

35. There are a number of other projects underway related to the maintenance of standards in 2017 and beyond. We have nearly completed interviews with a small sample of GCSE maths teachers on how they plan to make tier entry choices for the new GCSE. Inappropriate tier entry decisions are a threat to exam boards' ability to set tier level standards. We will report our findings later this year.
36. We are also embarking on a project to consider the impact of decoupling AS on teaching and learning to better understand the implications for the maintenance of standards in AS and A level in the coming years. It has proved difficult to speak to any schools this side of the summer break but we have arrangements in place to speak to year 12 teachers early in the autumn term. We are aiming to go into 16 to 20 schools and talk to 2 to 3 teachers from each.

### **Research**

#### **Annual Perceptions Survey**

37. The Annual Perceptions Survey (Wave 14) was published in June. In general, the survey indicated little change in perceptions of GCSEs and A levels. These were largely seen as well understood, trusted and valued qualifications. Some differences were perhaps of note, for example, young people were more likely to disagree that A level standards are maintained year on year and indicated less confidence in the A level system. This may be due to direct experience or the increasing role of social media debates in forming young peoples' views.
38. This publication (as well as the Wave 13) was produced by YouGov on behalf of Ofqual. We worked with YouGov on the presentation of the material to ensure greater emphasis on changes over time and on changes in overall perceptions to give a more balanced report. There was little media coverage.
39. SMG will shortly consider the future of the survey, in particular, how best to measure ongoing confidence in the system.

#### **GCSE Science and A level Mathematics accreditation**

40. To support the accreditation of GCSE Science two rounds of a comparative judgement exercise were conducted in which science teachers compared the difficulty of pairs of items from the sample assessment materials and from the current exams. Now that the science specifications are accredited we will publish the research report.
41. As reported previously, we are supporting the accreditation of A level mathematics in a similar way. A level mathematics items are more complex than items included in previous studies. For high tariff items the method for allocating intermediate marks will play an important role

on overall item facility. We conducted some pilot work to test the impact on the reliability of judgements of including the mark scheme, different types of judges (teachers and PhD students) and the exact question asked of the judges.

42. The pilot indicated that inclusion of the mark scheme had a positive impact upon judging reliability and teachers produced more reliable judgements than PhD students. We found that the question “Which item is more difficult overall?” produced the most reliable judgements for the longer response, higher tariff items found in A level.
43. The live study is currently underway with all the findings of the pilot incorporated into the methodology. We are on track to feed the outcomes of the study into the accreditation panel meetings which start 25<sup>th</sup> July.

### **Measuring the sawtooth effect in GCSEs and A levels**

44. 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. Last year we found evidence to support the notion of the sawtooth effect in the pattern of grade boundaries set. This is important because it may help us interpret NRT outcomes.
45. The work on grade boundary analysis has been followed up with a judgemental strand in four subjects: mathematics, history, geography and religious studies using an adaptation of the comparative judgement technique comparing quality of work over time in candidate scripts. This work has found that overall the pattern of judgements of script quality are congruent with the pattern of grade boundaries i.e. increases in candidate performance in the first three years followed by a levelling off. This gives strong support to the notion of the sawtooth effect. This report is in the process of being finalised.

### **Quality of marking**

#### **Research into Online Standardisation and Moderation**

46. In two separate studies we are investigating the judgemental processes and experience in two key parts of the assessment process - online standardisation and moderation.
47. Ofqual’s 2014 Quality of Marking Project found widespread dissatisfaction with online standardisation amongst markers. Limited research has found no difference between online and face-to-face standardisation in terms of marking accuracy. However, such research did not examine the processes themselves or consider whether they contained features associated with learning and skill acquisition. Last year we attended some online standardisation set up meetings and

identified some variability in the number of scripts/items used for training markers and the amount of feedback potential.

48. This study aims to take a close look at the online standardisation experience using a qualitative methodology. A small number of observations, using specialist software, allows us to capture the live authentic experience of markers as they proceed through online standardisation and this is then followed up by a stimulated recall. A survey to all examiners will create more scalable data. Observations have taken place this examination session and are also planned for the November examination session.
49. Moderation is under researched. While work has been done on the front and back end of the process (e.g. optimal sampling methods and scaling algorithms) there is little known about the judgements made in moderation. Arguably, moderation is a more complex task than marking as it involves a greater number of sources of information to synthesise. Over and above candidate work and the mark scheme, moderators also have to take into consideration the original marks given by one or more teachers, the rank order of the candidates in the sample, the marking tolerances and the overall distribution of marks within the centre. Additionally, they may also be aware of the likely grade boundaries.
50. Using similar techniques to the online-standardisation research (observation; stimulated recall; survey) we are currently gathering data on how moderators make these complex judgements. Already the study is revealing interesting facets in moderators' decision making. For both studies, we intend to finalise reports in early 2017.

#### **Quality of marking metrics and marking reliability studies**

51. We now have a variety of metrics which are calculated using double marking data from the quality assurance of live marking. These range from the very simple (e.g. the proportion of responses upon which there is exact agreement as to the mark) to the very complex (e.g. an estimate of subject level reliability of marking with grade boundaries overlaid).
52. During two stakeholder events in June, we reported two simple metrics – the percentage of times there is exact and adjacent agreement between markers across 6 mark items for a range of subjects at GCSE and A level. We contextualised this data by drawing on the marking reliability literature to show the level of exact and adjacent agreement for international assessments. This revealed that the level of agreement for GCSEs and A levels is not out of kilter with that seen internationally. The audiences seemed to understand the data and its implications. There was no negative media reporting.
53. In the autumn we plan to publish more on the metrics but not comparisons between exam boards which might have negative consequences for the quality assurance of live marking and

assessment design. In the meantime, we are using the metrics to inform a series of marking reliability studies.

- 54. These studies will be conducted for assessments about which stakeholders are concerned and which seem to be less consistently marked according to the metrics. They will involve a group of senior examiners re-marking a representative sample of scripts from which marks and annotations have been removed. In the research we will seek to disentangle differences of opinion about the marks awarded from marking errors. This is not possible through the metrics.
- 55. The studies will be designed not only to be descriptive but diagnostic - they will test hypotheses about the potential causes of inconsistency and errors in marking. The rolling programme of studies will begin with assessments from two or three reformed AS subjects across all boards.

**Proficiency specification**

- 56. In the VQ arena, we are scoping a programme of research which is intended to investigate what might count as good or bad practice in specifying the proficiencies (constructs) that qualifications are designed to assess. Proficiency specification embodies two related activities: constructing the proficiency (identifying what needs to be assessed); and representing the proficiency (communicating what needs to be assessed). During the first phase of this programme, the focus will be upon proficiency representation.
- 57. Most VQs in England seem to adopt a ‘learning outcomes’ approach to proficiency representation which can be traced to the National Vocational Qualification model, albeit subsequently shaped by Qualifications and Credit Framework requirements. Given the longstanding critique of the NVQ model, and the abolition of QCF requirements, it is an appropriate time to explore this area. Initially the project will investigate existing practices across a wide range of qualification types, alongside a review of key reports from the literature.

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

**ANNEXES LIST:-**

- ANNEX A** Strategic Risk Register
- ANNEX B** Systemic Risk Register