

Ofqual Board

Paper 74/16

Date:

25 January 2017

Title:

Strategy, Risk and Research Update

Report by:

Michelle Meadows

Responsible Director:

Michelle Meadows

Paper for discussion/information

Paper is part open, part closed, annexes are closed.



**Strategic Policy and Risk Team
Strategy and Corporate Planning**

1. In December we identified a list of the top 30 regulatory priorities for the year ahead. These were informed by the priorities discussed at the October Board strategy day and recent systemic risk work. We assigned to each priority an indication of the size and scale of the work necessary. In addition, we identified a range of enabling activities needed – both in the work of corporate services teams, and other underpinning functions such as strategic relations, risk, and regulatory framework development (Annex A).
2. This work is providing a clear steer to what we are *not* prioritising, as well as what we *are* prioritising and is now being used to direct business planning. Further, we have redesigned our approach to business planning to set out clear milestones and resource requirements. This will improve our ability to identify and manage resource pressures across the year, and make prioritisation decisions, if required, when new activities emerge in the year.

OFFICIAL – SENSITIVE

3. Work is underway to prepare our published 2017-20 Corporate Plan. This will present key strategic priorities and business plan activities, with stakeholders in mind. The Board will have the opportunity to comment on the draft plan at the February Strategy day, before formally agreeing the plan for publication at the March Board meeting.

Strategic Risk

4. The Strategic Risk Register is shown in Annex A. The ratings have changed little since the last review with a few notable exceptions. The judgement of the Board Transition risk (Risk 1) has been amended as the result of the successful appointment of a new Chair. We have made positive progress in managing two of the risks associated with vocational qualifications reform, Apprenticeships (Risk 6) and Functional Skills (Risk 8). The National Assessment risk (Risk 15) has risen mainly due to the current lack of dedicated resource to support this area of work. Our review of the risk associated with technical qualifications will be split into two in future. This split will draw out the risks associated with our ability to develop and implement our regulatory approach in line with government's reform timeline.

Entity Risk

5. In December we completed our six-month refresh of entity risk profiles for each awarding organisation. These set out the potential impact of failure based on the number of certifications and the stakes of the qualifications; and the likelihood of failure, using a range of indicators, such as rapid growth or contraction in certifications.
6. We intend this to be the last time that entity risk profiles are produced in this way. We are testing with General Qualifications a new approach to more real-time entity risk profiling, compiling risk information from a variety of sources to produce a 'single view of risk' for each entity. This single view of risk will be one of the key sources of information in improved day to day risk based regulation of entities.

Regulation Development and Impact Team

Proposed changes to the European Qualification Framework

7. The European Commission has proposed changes to the European Qualifications Framework (EQF), including: regular re-referencing of national qualification frameworks to the EQF; direct referencing of some qualifications to the EQF; developing common principles for credit accumulation and transfer.
8. We support periodic re-referencing but not the other proposals. Referencing qualifications directly to the EQF, rather than to the national qualification frameworks, would fundamentally change the nature of the

OFFICIAL – SENSITIVE

EQF and undermine national frameworks. Similarly, while common principles for credit accumulation and transfer could be beneficial, our view is that they should not form part of the EQF. We also question the Commission's vision for a system of credit accumulation and transfer across all qualification types.

9. Our thinking is aligned with colleagues across Government and other member states have also expressed concerns. We will continue to engage with colleagues in DExEU and other departments as negotiations progress. Ultimately, none of the proposed changes would be compulsory. We can choose not to adopt any problematic aspects of recommendations.

Legal Team

This section (paragraphs 10 – 14) has been redacted, as its publication would be prejudicial to the effective conduct of public affairs.

Standards Team National Reference Test

15. We have commissioned NFER to include a student survey with the NRT when it is taken. This will provide some information on the motivation of students when taking the NRT, and in their GCSEs, as well as indicators of how much they value GCSEs in English/maths and how useful they believe English/maths skills are to them in their lives. This information will provide a useful backdrop to the interpretation of NRT results.
16. We have investigated the relationship between KS2, the Preliminary National Reference Test results and GCSE performance to establish how likely the NRT is to be a good predictor of GCSE performance. The correlation is similar to that between GCSE and KS2 despite the fact that KS2 is taken several years earlier. However, this may well improve when we examine the NRT relationship as students will have studied the new curriculum and the sampling will be better.

MFL native speaker research

This section (paragraphs 17 – 18) has been redacted, as its publication would be prejudicial to the effective conduct of public affairs.

Preparation for summer 2017 awards

This section (paragraphs 19 – 21) has been redacted, as its publication would be prejudicial to the effective conduct of public affairs.

**Research Team
Marking reliability studies**

22. We are now able to routinely create marking consistency metrics for GCSEs and A levels. However, the metrics are based on data from exam boards' quality control mechanisms (seeds and sample double marking). As we have previously discussed, publishing such metrics might have perverse consequences for the monitoring of live marking (by corruption of the seeding process). Thus, to be able to publically report on marking reliability we are conducting empirical marking reliability studies in the first quarter of 2017. These will involve the blind re-marking of scripts. The subjects to be included in these studies are GCSE English Language, GCSE Biology; AS English Literature; AS History.
23. These marking studies will be informative in a number of ways. For example, we will explore which boards/units are associated with higher levels of marking consistency within the same qualifications; we can also triangulate the data obtained in the marking reliability studies with that generated for the metrics from the seeding data.
24. In addition to obtaining reliability measures, we are also conducting two strands of experimental work. In the first strand, we are looking at whether some methods of expert judgement other than traditional marking might yield greater consistency. We are investigating paired comparative judgement as well as an adaptive rank ordering method.
25. In the second strand, we wish to help develop an understanding of the difference between legitimate difference in examiner judgement versus illegitimate difference in examiner judgement. This work will progress the typology of mark disagreement proposed by Black and Newton (2016)¹ which identified four categories of marking disagreements procedural error, attentional error, inferential uncertainty, and definitional uncertainty. This empirical work will help contribute to guidance and requirements in post-results processes. Early findings will be available by the July Board meeting.

Reviews of Marking and Moderation – evaluation of 2016

26. The Official Statistics for Reviews of Marking (formerly known as Enquiries about Results) were published in December 2016² and show, compared to 2015, a decrease in the number of grades challenged by centres. Of those challenged, however, a fairly similar

¹ Black, B. and Newton, P. (2016) *Tolerating Difference of Opinion*. Paper presented at the 17th AEA-Europe Annual Conference, Limassol, Cyprus.

² <https://www.gov.uk/government/statistics/reviews-of-marking-and-moderation-for-gcse-and-a-level-summer-2016-exam-series>

OFFICIAL – SENSITIVE

proportion resulted in a grade change (18% in 2016 compared to 19% in 2015).

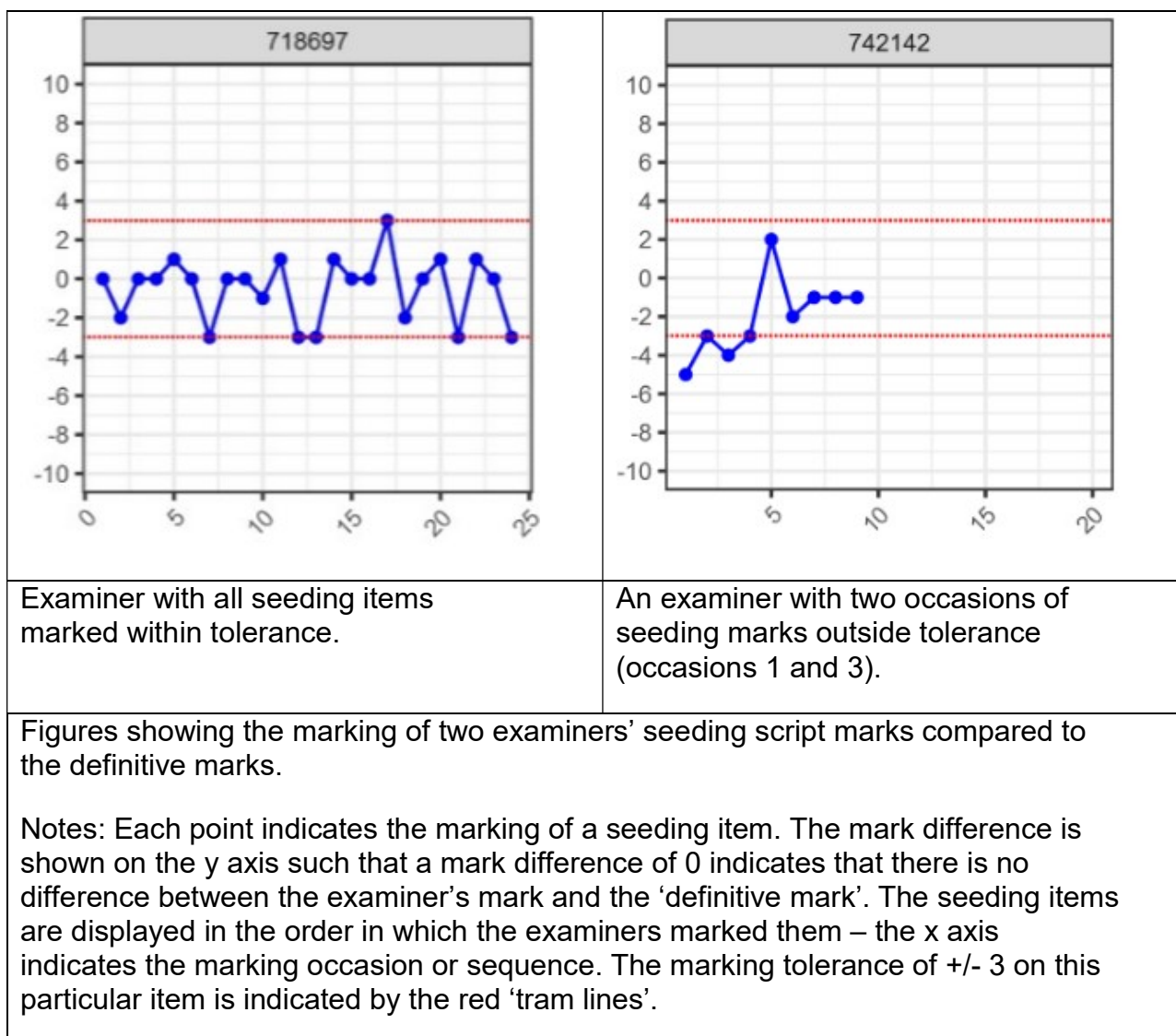
27. It is not clear from such data the extent to which boards have fully implemented the new requirements around reviews of marking and moderation – for example, in the requirements for training and for reviewing examiners to only change marks where there is an error or an unreasonable application of the mark scheme.
28. Thus we are conducting an evaluation. The first evaluation activity entails administering a survey to all reviewing examiners and moderators to find about their experiences of training in the new system and the circumstances under which they might change marks. To date, WJEC and Pearson have circulated the link to these online surveys and they have been completed by several hundred reviewers.
29. Another evaluation activity involves scrutinising a small but representative sample of examination scripts which have been subject to a review of marking to help understand the extent to which the new rules and requirements were enacted. Such scrutiny has a carefully designed procedure and will be conducted by both exam board senior examiners and Ofqual subject experts. The subjects included in this scrutiny include GCSE English Literature and GCSE Mathematics.
30. Finally, we will also look in more detail at the data from the boards, in terms of mark changes as well as grade changes. Where a unit/component, such as those of GCSE Mathematics or Sciences, contains short answer, relatively objectively marked questions, small mark changes may indicate ‘errors’ which have been corrected as part of the review process. In contrast, in units such as A level English Literature or History, which contain entirely extended response items, we might not expect to find small mark changes since these would indicate a legitimate difference in marker judgement.
31. The work evaluating Reviews of Marking and Moderation should conclude in early April in order to ensure sufficient time for boards to address any issues in time for the 2017 post-results session.

Exam boards’ implementation of quality of marking monitoring systems

32. As part of an audit of how boards implement their quality of marking monitoring systems, we are contributing to the testing plan by making use of the marker monitoring (seeding) data. We can use this data to sample examiners from a number of units to examine patterns in seed marking, in and out of tolerance, over the marking period.

OFFICIAL – SENSITIVE

33. Sample graphs are displayed below.



34. We can compare the 'stopped markers' (those who the board determined to be too aberrant to continue marking) with those who were allowed to continue. We will also present some such data back to boards and ask for evidence that they have implemented their own processes for dealing with instances when markers mark seeding items outside tolerance. These processes usually state that markers will be given additional guidance or training and/or be subject to additional monitoring.

Publications

35. Publishing in academic journals is key to the credibility and quality of the research team. There have been several papers accepted for publication since the Board were last updated, these are listed below for information (papers can be made available to any members on

OFFICIAL – SENSITIVE

request). In particular, it is worth noting that Ofqual’s approach to validity has been published in Practical Assessment, Research & Evaluation.

Holmes, S.D., He, Q. & Meadows, M. (in press). An investigation of construct relevant and irrelevant features of mathematics problem-solving questions using comparative judgement and Kelly’s Repertory Grid, *Research in Mathematics Education*.

Newton, Paul, E. (2016). Macro- and Micro-Validation: Beyond the ‘Five Sources’ Framework for Classifying Validation Evidence and Analysis. Practical Assessment, Research & Evaluation, 21(12). Available online: <http://pareonline.net/getvn.asp?v=21&n=12>

Opposs, D. (in press). Whatever happened to school-based assessment in England’s GCSEs and A levels? *Journal of Education*, 66.

He, Q., M. Meadows & B. Black (in press, 2018). Statistical techniques used for detecting anomaly in test results. *Oxford Review of Education* (Special issue: The Consequences of Metrics in Education: Gaming, Malpractice and Cheating).

Taylor, R. C. (2016). The effects of accountability measures in English secondary schools: early and multiple entry to GCSE mathematics assessments. *Oxford Review of Education*, 42(6), 629-645.

Paper to be published	YES
Publication date (if relevant)	After the meeting but not in full – redact paras 10-14, 17-21
If it is proposed not to publish the paper or to not publish in full please outline the reasons why with reference to the exemptions available under the Freedom of Information Act (FOIA), please include references to specific paragraphs	<i>See guidance on exemptions below</i>

OFFICIAL – SENSITIVE

ANNEXES LIST:-

ANNEX A Closed

ANNEX B Closed