

# News story: Opportunities for improving quality of marking

Ofqual takes the quality of marking of GCSEs, AS and A levels very seriously. Taken by more than a million students every year, we recognise that trust in the results of these qualifications is essential to public confidence. We have conducted a substantial programme of research over the past 5 years aimed at finding improvements in a system that in many ways already delivers results that are as good as many other systems around the world. Today (Tuesday 27 November), we are discussing 5 new research reports with teachers and education leaders at an event in London that will further understanding of marking in the sector, and how we might work together to drive quality higher in specific areas.

Marking is a complex exercise. It requires exam boards to recruit, train, standardise, and monitor tens of thousands of individuals to review tens of millions of responses each year. Each subject lends itself to being assessed in different ways, from multiple choice questions to long essays, which we know can have a direct effect on marking reliability but also more importantly the learning experience of students in the classroom.

Reformed GCSEs, AS and A levels reflect this trade-off between the absolute reliability of any assessment and the value of qualifications to individuals. The challenge, therefore, is to make marking as good as it can be in every subject, in the context of the style of the assessment. There is not a single, right mark for every answer given in every subject. For many assessments different – but equally legitimate marks – can be awarded for the same answer by expert examiners. Here we expect mark schemes and training to be of high quality. For other assessments, there will be a single right mark. Here we expect the right mark to always be awarded.

Our new research supports these aims by looking at various aspects of the marking process. We will be publishing our research following today's event, along with results of our recent marker examiner survey. In summary:

## **1. Online standardisation**

Standardisation of markers can be conducted in different ways. We have looked at the processes involved in online standardisation in particular, and have identified some good practices that could be more consistently adopted to improve the experience and performance of examiners. These include receiving personal feedback by phone after being approved to begin marking and receiving confirmation that they are awarding marks on the same basis (as well as the right mark) as intended. It is also important for examiners to take personal responsibility for ensuring they review any feedback received.

## **2. Hard to mark responses**

Previous research has identified that sources of disagreement between

examiners can be categorised as: procedural error (mistakes or not following procedure), attentional error (concentration lapses), inferential uncertainty (uncertainty in drawing inferences from the students' responses) and definitional uncertainty (uncertainty in the definition of what is to be assessed). The first 2 categories can be described as errors, while the last 2 are present in responses for which there can be more than one legitimate mark. Our latest research finds that the frequency of each category tends to vary by subject. For example, in biology, inferential uncertainty is more common, while in English language definitional uncertainty is more likely. We expect exam boards to reflect on these findings to see where they can improve their mark schemes.

### **3. Marking versus comparative judgement**

Our examination system values the use of extended response questions in assessing important higher-level skills. But these responses are harder to mark than shorter, or more constrained question types. This can impact upon the validity of the rank order of candidate work. We are therefore considering rank ordering students' work by means other than marking. This study looked at 2 different alternatives – paired comparative judgement, and rank ordering by placing extended responses in rank order – and comparing these with 'traditional' marking using a mark scheme. The research finds that the 3 methods produce rank orders that are very similar. This work indicates that more research in this area could be worthwhile.

### **4. Marking consistency metrics – an update**

Earlier work has focussed on component level marking consistency and found that results in England are comparable to others internationally. This paper reports new qualification level marking metrics, which are shown to be generally higher than those at component level from which they are comprised. And we note that marking consistency remained stable in England between 2013 and 2017. However, this does not mean that improvements cannot be made. In response, the paper considers how minimum acceptable levels of marking consistency might be defined, which would help exam boards to channel additional resource and support. We note that these thresholds would need take into account the subject and/or forms of assessment, but importantly, would need to be understood and accepted by the public.

### **5. Marking consistency studies**

We measure marking consistency of the 4 exam boards offering GCSEs, AS and A levels in England annually. We have previously said that if we were to publish these metrics, we might compromise live marking monitoring. This new research provides an insight into marking consistency without these drawbacks. We found varying levels of marking consistency across subjects and between individual subject units. The results confirm our belief that marking is generally good across the system, albeit there is room for improvement in some specific areas. We want exam boards to reflect on these results and make appropriate changes to question design and mark schemes for future series.

We have also published the results of a survey of examiners, conducted prior

to the summer 2018 series. This survey – which received more than 18,000 responses – gives a picture of the professional background of examiners, as well as their experiences of the examining process. Its findings include:

- survey respondents had an average of 10 years previous examining experience
- more than 99% of respondents were current or former teachers
- the average age of an examiner responding to our survey was 47 years
- 96% of markers and moderators agreed that they were confident in their ability to mark or moderate accurately and reliably

Sally Collier, Chief Regulator, said:

Our latest research confirms that the quality of marking of GCSEs, AS and A levels in England is good, and compares favourably to other examination systems internationally. But we must not be complacent. We must continually strive for marking in every subject to be the very best it can be. We welcome the input of experts across the education system to challenge the status quo and drive improvements. We will reflect further on our own rules and expectations in the light of this work. And we also want exam boards to consider today's findings and take both concerted and independent actions in response. This will ensure public confidence in these qualifications, that are taken by more than a million students each year, is maintained or enhanced.