

## Ofqual Grades Consultation

March 2020

1. Do you agree that asking teachers and lecturers for an estimated grade and the rank order of students is the fairest approach that can be taken, given the government's decision that exams in May and June will be cancelled?

For the system to be as fair as possible, given the circumstances, it is essential that both grade and rank are collected. The fairness of the approach though rests not only on the reliability of the data collected, but on what is done with it afterwards. Statistical moderation processes run the risk of introducing new unfairness to the process. Teachers may calculate entirely appropriate grades, but these could be changed by a moderation process that is not able to deal with the remarkable complexity of awarding so many grades.

2. We expect exam boards will ask teachers and lecturers to come to holistic judgements about the grade/s they think a student would have been most likely to have achieved in the summer if they had taken their exams. Considering the things we have set out above, are there other factors we should reflect in the expectations or any you think should not be included?

The biggest danger here probably lies in limiting the range of suggested 'evidence' or being over-prescriptive about what constitutes 'sufficient' evidence. Schools and colleges will take very different approaches to the management of (for example) a two-year linear A level course. Some will require regular assessment and frequent data drops across the college, and will have a very standardised evidence base. Others will have allowed subjects to build programmes of learning and assessment that suit the needs of the specification. Both may come to entirely valid judgements about student progress, and it would be a mistake to define too tightly what might contribute to this holistic judgement.

It is important to think about how teacher bias re: gender, race, and class might play a role in thinking about what grade a student 'deserves'. We know it's widespread, unintentional, and can actually be made worse by reminding teachers of its existence, so it would be reassuring to know your thoughts about accounting for it in some way other than asking teachers not to be biased in their guidance.

3. We expect exam boards will ask teachers and lecturers to come to holistic judgements about the grade/s they think a student would have been most likely to have achieved for those subjects with additional grades (GCSE English language and A level science subjects). To what extent do you think this information may already be available?

The science A level practical endorsement is a yes/no judgement, so should not be problematic. This evidence should, in almost all cases, have been assembled by this point.

4. We expect exam boards will ask schools to ask teachers and lecturers to determine a rank order within each grade for all the students in the school or college. Considering the things we have set out above, what guidance in particular do you think will be important in supporting teachers to determine these rank orders? Do you think this might be particularly challenging to achieve in certain types of centre or in certain subjects?

In a small school this would be a relatively simple exercise, particularly if (for example) there is only one class for each subject. The larger the centre and the larger the subject the more challenging this becomes. While there is a parallel in the moderation of internally assessed coursework, and many teachers will have experience of that, coursework moderation has the advantage of being able to directly compare pieces of work produced to a common specification assessed against a standardised mark scheme. While ranking within a class is straightforward, ranking within a cohort is much less so. At the larger sixth form colleges, there are over 500 students in some subjects. Given the implications of the rank order are significant (that those who are at the bottom of the grade are the most likely to go down a grade in moderation) there is a danger here that the rank order has a spurious accuracy, and a student's fate hangs on whether they were placed twenty-seventh or twenty-eighth in a class. Indeed, twenty-seventh and twenty-eighth may be identical in their actual ability.

5. We are mindful there will be a lot of interest in teachers' judgements from students and parents. We also know that teachers must be supported in making these professional judgements without being subject to undue pressure, and that the grades they determine may be changed following the standardisation process. Given this do you think teachers' estimated grades should be able to be shared with parents or students?  
Note: it will not be possible to share rank order information without releasing personal data, so this will not be allowed.

There is huge danger here. Schools and colleges have already (in the absence of the actual guidance of how this will be done) come under significant pressure to change 'predicted grades', as parents fight to get the best possible outcomes for their children. Presumably some teacher grades will be a grade U, or grade 3, and this will not go down well with parents. Schools and colleges already have an extraordinary job in the months ahead, and to create an unnecessary point of conflict with parents (which would at least be demanding and potentially impossible in terms of management time), would be unhelpful at what is a remarkably challenging time.

Our view is that teacher predicted grades should not be subject to student challenge. Indeed, it would not be impossible to make them confidential – so as not to raise and dash hopes before the actual final grade is awarded.

6. To support a process that is fair to all students, we expect exam boards will ask the head of centre to confirm that the expectations set for this process were followed to determine the grades and the rank order submitted and that in their judgement what is submitted is a true representation of student performance. Do you think such a step will be manageable for schools and colleges?

I think it is essential that the head of centre has to confirm this, and a good head of centre would work through the decisions made in each subject area, to confirm that the expectations set for the process were followed.

7. We know some students will have recently joined schools and colleges, and some others may not have studied at the centre but instead will have been planning only to take their exams there. How much contact do you believe a school or college will have to have had with a student in order to make an estimate of their likely grade?

I don't really see how external candidates and recent joiners should or can be part of this process. For external candidates there is no evidence base of performance or improvement, and for a recent joiners the evidence may or may not be strong enough. In these circumstances a centre should be allowed to post a grade of 'insufficient evidence' if that is the right call for the individual circumstance.

8. We are planning to provide information for students on this process. Do you have any thoughts on what this should include?

Information for students should include a clear sense of what teachers have been asked to do. It should emphasise that it is a professional judgment, taking into account all the available evidence assembled about the student concerned, and the professional judgement and institution knowledge developed over time with multiple cohorts of students. Teachers are making a judgement about what they know about one student in the context of what they know about all the students they have ever encountered. The guidance for students should emphasise that the teacher grading and ranking is one part of the process, and that the moderation of that is also important. There should be clarity of just what an 'appeal' will involve, and who will be appealed to, and what circumstances should warrant an appeal. My view is that these early 'appeals' should be about whether an error has been made (the wrong John Smith was put in the D grade band, for example). If candidates really don't like the grade awarded to them, then they should be sitting the Autumn exam.

9. Is there anything else you suggest we should be considering at this stage?

For schools and colleges, the biggest issue is probably what happens after the grades have been awarded. Using what teachers have calculated we end up with a proposed grade profile for each subject in every school and college, and a student ranking of each of those cohorts. What we are missing is an understanding of the factors that led teachers to put forward those particular grades. Is my understanding of an A grade the same as another college's understanding of an A grade? To moderate different cohorts without a consistent basis for the grades they calculated is very difficult indeed.

To get some sense of why the 'post-teacher calculation moderation' phase is critical, consider the following:

- Some centres have better qualified candidates than other centres. To what extent will prior attainment be taken into account when deciding whether a centre's grades are in line with expectation? There is a danger of sticking too tightly to what a value added model would predict for a centre in that...

- In a large College A level entries account for over 2000 entries and we are concerned that the proposed grading method "to ensure that the distribution of grades follows a similar pattern to that in other years" may mean an expectation for this national distribution to also be reflected in a very large college cohort. However, this would not be the case. Take, for example, an outstanding college, where very high and statistically significant value added, with a very high proportion of high grades, have been maintained year on year and are far in excess of the national rates. This is a different context from a small school sixth forms whose value added data, which because of low learner numbers, is not statistically significant or alternatively varies wildly year on year, making reliability of any predictions difficult.
- Some centres are better than others, and in each centre some departments are better than others. To what extent will historical performance be taken into account in moderating the performance of individual centres? For example, I know of a media studies department which has outperformed predictions based on prior attainment by at least half a grade per student for each of the last ten years. If their outcomes were limited to what a regression line would predict, around half of their students would get a grade below what they would have achieved had the exams gone ahead as planned. These students would clearly be disadvantaged (contradicting the stated aim of the whole process).
- But historical performance may also provide a flawed basis for analysis. Some centres are on a different trajectory to other centres, and in each centre some subjects will be on different trajectories. Imagine a psychology department which had a particularly poor year in 2019. If it had made no real effort to make any changes, then historical performance may be a useful starting point for grading. But what about one where the head of the underperforming department had been replaced, and the new leader had formed a new dynamic and highly effective team. It would be unfair if that department was pegged to performance the previous year.
- If prior attainment and prior performance are to be used, at what level can they be applied? At centre level, at subject level, at candidate level? I think it is important to use prior attainment and prior performance as part of the process, but getting this right is a real challenge.
- Some centres will be new and will have no history to go on.
- Other centres will be new to the particular board they are with that year, so will have no history to consider in moderation.
- There is a real danger of grade inflation here (as centres try to push the best possible grades for their students). The danger is that the processes used to dampen the inflation and restore the grade profile to a credible one may dampen performance in centres that was actually correct.
- You could develop a process that had statistical moderation of grades followed by human moderation of grades, but to what extent is there expertise in exam boards to carry out this moderation? They have extraordinary expertise in standardisation and moderation of real grades, using real student work, and they have robust processes to identify and respond to 'rogue' marking. But they do not have the same experience of doing the same with 'calculated' grades. The question is who would have that expertise? The timeframe for this exercise is pretty unforgiving, given that initial announcements suggested students are being promised grades in July rather than August.

