

Written evidence submitted by Dennis Sherwood

Context

I write this submission with reference to “*The effect of cancelling formal exams, including the fairness of qualifications awarded and pupils’ progression to the next stage of education*” in general, and “*the fairness of qualifications awarded*” in particular. I do so on the day following the [announcement](#) by Ofqual of the procedures to be followed this year.

This year’s procedures

My view is that Ofqual’s guidelines, as published yesterday, will deliver the most fair qualifications possible, and much fairer than the grades delivered by the exam system over recent years. Ofqual are therefore to be congratulated. As evidence to support this opinion, please refer to

<https://www.hepi.ac.uk/2020/04/04/weekend-reading-a-for-ofqual-and-the-sqa-this-years-school-exam-grades-could-well-be-the-fairest-ever/>

and

<https://www.hepi.ac.uk/2020/03/21/trusting-teachers-is-the-best-way-to-deliver-exam-results-this-summer-and-after/>

I suggest that there are two reasons why this year’s grades will be more fair than those awarded in recent years:

1. The intrinsic fairness of relying on teachers’ opinions, especially as regards determining fair rank orders (subject to the problem of ‘clusters’, as explained below).
2. The intrinsic unfairness of the current policy for awarding examination grades.

It is this second point that is the substantive content of this submission; let me firstly make one point in relation to the determination of rank orders by teachers.

In principle, teachers are in a strong position to do this fairly, provided that they act with honesty and integrity. In so doing, however, it is most likely that ‘clusters’ will emerge: the most able students will be readily identified, as will those least able; but for many ‘in the middle’, I suspect that some students will be ‘indistinguishable’, and would show as ‘joint equals’ in any rank order. This is, I think, a real possibility, and is not indicative of a teacher’s indecisiveness, but rather a reflection of a truth. If such ‘clusters’ are remote from the grade boundaries, that is unlikely to present a problem: I would therefore expect that the teacher-assessed grades would ensure that any such clusters avoid grade boundaries.

However, if the central ‘statistical standardisation process’ changes the location of those grade boundaries, it is possible that any ‘cluster’ might be cleaved in two, so that a community of ‘equal’ students is divided into two groups, one with a higher grade than the other. This is intrinsically unfair.

In seeking to avoid this, Ofqual are quite likely to require that each centre submit a ‘clean’ rank order, with each student distinguished from neighbouring students, and with no ‘joint equals’. This solves Ofqual’s problem of drawing lines, and passes the buck back to the school, who “own” the rank order. I suspect, however, that the reality is more nuanced. In forcing a ‘clean’ rank order, Ofqual are requiring many teachers to exercise the ‘wisdom of Solomon’ - which is unfair both to the student and the teacher too.

As already noted, I write this very soon after the guidelines have been published, so there is not yet any evidence of applying these guidelines in practice. I would not be surprised, however, if the ‘cluster’ problem were to emerge as very real, and it may be that Ofqual will need to address it in due course.

The unfairness of recent exam grades

My second point - the intrinsic unfairness of the current policy for awarding exam grades - forms the main content of my submission.

I appreciate that, strictly, the unfairness of grades as awarded by the exam system in the past might be considered to be irrelevant to the terms of reference of this Inquiry. I take the liberty of addressing it, however, for the Inquiry might be considering implications for the future, after the COVID-19 crisis has been overcome. In which case, it is very relevant indeed, for if any future assessment process involves examinations, it would be a public good for the corresponding grades to be reliable, which the recently-awarded grades are not.

My starting point is Section 22 (2) (a) of the Education Act 2011 which [states](#) that

*“The qualifications standards objective is to secure that
(a) regulated qualifications give a [reliable](#) indication of knowledge, skills and understanding...”*

where I have emphasised the word “reliable”.

On 11 August 2019, an [announcement](#) was posted to the Ofqual website that includes these words

“...more than one grade could well be a legitimate reflection of a student’s performance...”

This statement is unqualified, and so - presumably - applies to all grades in all subjects. The statement is within an announcement about A levels, and so might, by inference, apply to A level grades, but not to GCSE or AS grades: I will refer shortly to some further evidence suggesting that the statement applies not only to all grades within all subjects at A level, but also in the same way to GCSE and AS grades and subjects too.

Ofqual's statement merits some reflection. What, precisely, does it mean? If "more than one grade could well be a legitimate reflection...", how many other "legitimate" grades might there be? Might any of these 'other' grades be higher than the single grade actually awarded, as declared on the candidate's certificate? Or perhaps lower? What are the implications as regards how awarded grades are used by, for example, prospective employers, or admissions officers in colleges or universities? And for those awarded grade 3 in the key subjects of Mathematics and English and therefore obliged to re-sit, might another "legitimate reflection" be grade 4 - in which case there is no need to re-sit at all?

Most importantly, what are the implications of this statement as regards Ofqual's statutory obligation under the Education Act 2011?

Further evidence of the (un)reliability of GCSE, AS and A level grades will be found in two Ofqual reports - [Marking Consistency Metrics](#), published in November 2016 (see in particular Figures 13 and 14), and [Marking Consistency Metrics - An update](#), published in November 2018 (Figures 9, 10 and 12).

For some interpretations of these reports, may I refer you to these blogs, published on the website of HEPI, the Higher Education Policy Institute:

<https://www.hepi.ac.uk/2019/01/15/1-school-exam-grade-in-4-is-wrong-does-this-matter/>

<https://www.hepi.ac.uk/2019/02/25/1-school-exam-grade-in-4-is-wrong-thats-the-good-news/>

As these documents show, across all grades and across all subjects, the average reliability of school exam grades is about 75%; a further inference is that for any subject, and for any grade, the reliability of the grades awarded to scripts marked at, or very close to, any grade boundary, is 50% at best. Tossing a coin would be more fair. Nor does the current appeal process [right these wrongs](#).

Let me stress that the underlying cause of this unreliability is NOT poor marking or ambiguous mark schemes. It is directly attributable to the weakness of the current policy for determining a script's grade from that script's (single) mark - a policy that fails to take into account the fact, [acknowledged by Ofqual](#), that "it is possible for two examiners to give different but appropriate marks to the same answer". If those "different but appropriate" marks are within the same grade width, there is no problem. But if they straddle a grade boundary, it is a [very big problem indeed](#). And grade boundaries are straddled for very many scripts: it happens to about 1 script in every 4.

Which brings me back to this question: what are the implications of this real data on grade (un)reliability as regards Ofqual's delivery of its statutory obligation?

Looking ahead

As I have already mentioned, the fact that the grades awarded in summer 2019, and for several earlier years too, were, on average, only 75% reliable may not be directly relevant to the fairness of the very different process, based not on exam results but on teacher assessments, proposed for this summer - save, perhaps, for noting that the benchmark against which the proposed process might be compared is woefully low.

But it is hugely relevant to what might happen in the future.

It is most likely that examinations will continue to play a role determining candidates' assessment in the future - perhaps as the sole determinant, as in the recent past; perhaps in combination with teachers' judgements if the assessment processes and policies are to be reformed. But however examinations might be used, I believe it to be imperative that the grades to be awarded are - as the Education Act 2011 requires - reliable. And therefore trusted.

To continue, in the future, to use the policies for grading that have been used in the past will perpetuate the historic unfairness. This Inquiry therefore offers a unique opportunity for this problem to be solved - which is not difficult, as outlined in this HEPI blog:

<https://www.hepi.ac.uk/2019/07/16/students-will-be-given-more-than-1-5-million-wrong-gcse-as-and-a-level-grades-this-summer-here-are-some-potential-solutions-which-do-you-prefer/>

Conclusion

May I thank the Committee for allowing me to make this submission.

I have much more information, and, if this might be of interest, I would be happy to provide whatever might be helpful, either electronically, or, if appropriate, in person.

April 2020