

RSS welcomes the publication of OSR's review of the use of models for awarding grades in the UK in 2020

The Royal Statistical Society welcomes today's publication of the Office for Statistics Regulation's (OSR) review of the way that statistical models were used for awarding grades in 2020 – [Ensuring statistical models command public confidence](#).

The RSS [requested that the statistics watchdog carry out the review](#) (PDF) after the exams regulator Ofqual and other qualification regulators had to abandon the algorithms being used to estimate students' grades, due to 'significant inconsistencies' in the results. In calling for the review, we highlighted the importance of learning lessons from this process – both in case the cancellation of exams was not a one-off occurrence, and to set a benchmark to help prevent these issues recurring in future in other areas. We are pleased that the OSR has looked thoroughly at these issues and continue to welcome their engagement with the larger issues of how government uses statistical models. And we particularly welcome the OSR's plans to issue more detailed guidance on the use of statistical models by government.

The OSR report offers a general account of the difficulties posed by last year's decisions about how to award grades, noting the challenges it posed, both for governments and qualification regulators. The RSS has noted the unprecedented circumstances throughout. This report was never meant to have been a technical review of the qualification regulators in the UK, nor of the process by which clearer political accountability should be sought for last year's decisions. The former may be enabled by the release of data, and the latter would seem to be more properly a task for the Education Select Committee.

The report makes a number of useful observations about links between public trust, trustworthiness and quality, and use of statistics. We have never accused anyone involved of acting in bad faith. But we believe the OSR report vindicates our views that even last year, with its unique difficulties and extraordinary time pressures, alternative approaches could have been considered, and that external and constructive challenge, including from a wider range of statistical experts, might have helped with this. In particular, we note that even in the discussion of the importance qualification regulators gave to communications and public engagement, there was no early, formative stage at which views on different approaches could be considered in the light of their implications.

For instance, at no point was there an open public discussion about the choice between using statistical modelling to adjust teacher grades or rankings for individuals (which would have the benefit of controlling 'grade inflation', but at the cost of greater departure from teacher-assessed grades for individuals) and using modelling to look at schools whose grade awarding processes could be queried (which would result in more 'grade inflation' than governments seemed to have hoped, but with less adjustment of individual marks and more attention to consistency between schools). This latter method seems to have been dismissed due to the extraordinary and real time pressures, but we note that Cambridge Assessment also raised this point, and we think it could have been done in a way that was both legal and fair.

For now, one immediate use of the lessons in this review is to make sure that the process of awarding grades for this year improves on last year's. Our concern is that last year, too much weight was placed on a statistical model – with a grade inflation ceiling that was too low – to make estimates of individual grades. The plans for this year risk a laissez-faire attitude in school processes, with only vague references to how statistical data could be used to inform monitoring and quality control. Neither approach is right.

As the OSR so clearly points out, there are competing principles, going far beyond the statistical, that are stake in these decisions. All that we can ask for, in an imperfect world, is that statistics are used to

inform decisions in the most helpful way possible, and in a way that makes it clear which issues are technical and which require wider public engagement, including – but not restricted to – more use of external experts.

The recommendations and lessons in the OSR's review provide a platform to improve the process for awarding grades this year. Given where we are now, the use of teacher assessed grades may well be inevitable, but quality assurance, making use of historical statistical data, is the only way to achieve some degree of consistency between schools and public confidence in the results compared to previous and future cohorts.

We call on the qualification regulators and the exam boards – where responsibility for this now lies -- to make it clear how they plan to take a robust approach to using school historical statistical data to quality-assure school grading processes before results are announced – drawing on the lessons set out in the OSR's review.

Sharon Witherspoon, Royal Statistical Society Vice-President for Education and Statistical Literacy, said:

“Gavin Williamson has proclaimed that this year's adjustment process will not use any algorithms. While it is good that some lessons have been learnt from last year, this risks missing the point.

“As the Office for Statistics Regulation review suggests, the problem was not that there is no role for algorithms – but if they are used, they must be fit for purpose. The review acknowledges that consistency between schools could be addressed by looking at statistical patterns to suggest when school grading estimates should be queried. This does not seem to have happened last year. This year, we believe there is still time to use historical statistical data both as a guide for schools in setting grades and for the external quality checks by the exam boards.

