

Royal Statistical Society response to Ofqual’s consultation on setting the grade standards of new GCSEs in England

Response prepared by RSS Policy and Research Manager on behalf of the Royal Statistical Society.

- 1) *We have considered three possible ways by which the standard for new GCSEs could be set in the first year:*
 - (a) *An approach that uses statistical information to link the award of the new grades to current grades*
 - (b) *An approach in which awarders judge students’ work against descriptions of expected performance – criterion-referencing*
 - (c) *A norm referenced approach in which the proportion of each grade available to the cohort is pre-determined.*

Please rank these possible approaches, using 1 for your preferred approach and 3 for your least preferred approach:

Option (a).....1.....

Option (b).....3.....

Option (c).....2.....

Please give reasons for your answer

An approach that uses statistical information to link the award of new grades to current grades is most preferable, in part because those using the new grades need to have confidence as to their meaning. The new system will need to be compared to the old, and statistical information is helpful for this.

In addition to this though, we would like to warn against too extensive a reliance on statistical information to compare grade standards, in particular when comparing between different subjects. There is a perception of some subjects as ‘easier’ or ‘harder’. Measuring this statistically has its greatest validity in subjects that the vast majority of students pursue, such as GCSE English and Maths. Even for these subjects, it can only help to some extent. Grade standards for subjects that are not pursued by the whole cohort cannot be statistically compared.

It is important also to remember that pupils themselves influence what is ‘easier’ and ‘harder’ due to differences in their own interests and their perception of their strengths. For example, pupils in some GCSE subjects that are optional achieve higher average grades in comparison to what is achieved in maths. The disparity in results is useful to be aware of, but it does not necessarily



indicate that assessments in optional subjects are not challenging enough, or that maths assessments are too difficult. Maths is compulsory, whereas optional subjects are not, so there is likely to be an element of self-selection for success. Disparities in achievement between subject could indicate multiple issues, but the standard of assessment should not in large part be judged by a technical comparison of the grades achieved.

2) *We have proposed that in the first year the standard for a grade 4 should be set so that the proportion of students who would previously have been expected to be awarded at least a grade C in a subject will be awarded at least a grade 4 in the subject. To what extent do you agree or disagree with this proposition?* [No comment]

3) *Would you find it helpful if other points of reference between current and new grades were set and communicated before the first awards are made?*

Yes

4) *We have proposed that the standard of performance for a grade 5 should align to the expected standard for similar qualifications or exams taken in high performing countries. To what extent do you agree or disagree with this proposition?*

Strongly agree

Agree

Disagree

Strongly disagree

Don't know/no opinion

Please give reasons for your answer

International comparisons are fraught with difficulty. It would be necessary to know far more about how such comparisons will be made, and about the reliability of the underlying measures, before it would be possible to support this. Ofqual should clarify how it intends to use PISA or other international data to set standards in GCSEs.

If comparisons are to be made between England and other countries, the aim should be to understand underlying differences between countries, and to explore the data to reveal these. The existence of multiple ways in which countries differ should be reflected in comparisons.

5. *We are considering whether and, if so, at which points we should make a link between the new and the current grades. We would welcome your views on the appropriateness and the usefulness of the following possibilities: ...*

a) *setting the grade boundary for grade 7 so that, all things being equal, the same proportion of students who would previously have been awarded a grade A or above are awarded a grade 7 or above in the first year?*

Would you consider this link to be:

Appropriate **Yes**

Useful **Yes**



The new grade system should be comparable longitudinally in quantitative terms within subjects, and should be easily comprehensible to those using the grade system. Matching the new grade 7 to the old A grade would seem to help toward this.

6. *Would you find it helpful to have any additional or alternative points of reference between the current and the new grades?*

No.

7. *Is the current boundary between a grade G and an Unclassified outcome meaningful?*

Yes.

Please give reasons for your answer

It is important to measure educational progress at the bottom of the grade range, and in our view the difference between a current 'G' grade and unclassified is meaningful to recognise, particularly among the most disadvantaged students. Therefore the Grade 1 boundary should at a lower limit be set at G.

8. *In your view, should the grade 1 boundary be set to align with the current grade F or grade G?*

Grade G.

9. *Do you have any other views on the distribution of the new grades?*

In earlier implementations of new grading structures, faulty statistical procedures were used. For example, A* at A level was defined arithmetically as being as many raw marks above grade A as grade B was below. This resulted in subjects with the same proportions of grade A candidates having different proportions of Grade A* because the upper tails of the distribution curves had different shapes. Consulting with a statistician should help to avoid mistakes like this.

10. *We have proposed that the national reference test should be designed so that exam boards can use its outcomes to identify changes in the performance of the national cohort that could be reflected in the grades of new GCSEs awarded. To what extent do you agree or disagree with this proposition?*

Strongly agree

Agree

Disagree

Strongly disagree

Don't know/no opinion

GCSE syllabuses differ by design across exam boards, so the idea that there is a common 'standard' against which all can be calibrated does not hold true. It also can't be assumed that national cohort ability is the main source of changes in performance over time. It may be that pupils' abilities change from year to year, or it may be that entry policies and other related factors change the nature of those entering for particular exams.



A common reference test might provide weak evidence of serious problems in how grades are arrived at, but due to its weakness as a sole measure it could only be used as potential evidence of issues that can (and should) be picked up other means. We therefore see it as largely superfluous to requirements. Outside of the reference test, statistical analysis of exam results can identify potential problems in exam grading, which can be substantiated by checking how grades are arrived at and how examiners work.

If Ofqual were to press ahead in developing a test, we would need to know more in order to advise further. What would the test contain? How would the sample of students be chosen? How would the representativeness of the sample be determined?

Measurement of the motivation of each cohort would also be necessary, as the extent to which a cohort performs at their best is likely to differ in GCSE exams compared to the reference test.

The RSS thanks the following Fellows for their advice on this response:

- Prof Neville Davies
- Prof Harvey Goldstein
- Roger Porkess
- Neil Sheldon, RSS Vice President for Education and Statistical Literacy

Response submitted 30 June 2014.

