

Annex 3 – Technical Paper on Grading



June 2013

Ofqual/13/5293

Contents

What are the purposes of grades?	2
Reliability and validity – Cresswell (1986)	3
Should we change the way we report GCSE qualifications?	4
Continue with the current system (A*–G).....	4
Develop a new reporting system.....	5
Argument for grades and/or scaled scores.....	6
How many grades should there be?	7
Should there be additional information?	7

What are the purposes of grades?

To make sensible decisions about what grading a qualification should have, it is important to consider why grades are necessary in the first place. What uses will the grades be put to? Are they designed in such a way that makes these uses valid?

To meet the primary purposes of the reformed GCSE defined by the government, we propose that the key purposes of grades are to:

- certificate students' achievement in the subject;
- differentiate between students¹;
- indicate that a particular threshold has been reached, for example:
 - readiness to progress to A level;
 - “good pass” – currently, people regard grade C as this indicator.

Another significant use of qualifications is for school accountability measures, and so we also need to consider the implications around how performance tables use grade information.

One of the purposes set out by the Secretary of State for the reformed GCSE is that in English and maths respectively a “good pass” will provide greater reassurance than currently, that the candidate is numerate or literate. In a compensatory model such as GCSE this is difficult to achieve. If this competency is required then it needs to be assessed by a specific exam (part of the GCSE or otherwise) and represent a “hurdle” or the majority of marks available. Both would change the core purpose of GCSEs.

Finally, Cresswell's paper in 1986² makes a valuable point about the way in which decisions will be made, primarily in the context of selecting a candidate for further study or a job:

In any case, the use of a small number of grades for reporting examination results causes greater emphasis to be placed on other selection criteria. Clearly if the other criteria are less reliable than the examinations, greater

¹ For example, selection for employment

² Cresswell, M. J. (1986) *Examination Grades: How Many Should There Be?* British Educational Research Journal, vol. 12, no. 1.

reliance on them will lead to less reliable selection decisions (Cresswell, 1986).

Reliability and validity – Cresswell (1986)

Validity, and its component reliability, is critical for any qualification. In simple terms the use a qualification is put to is valid if it is a reasonable conclusion based on the evidence. A use might not be valid because of its content (ability to communicate in France based on GCSE in French that does not test speaking), inappropriate conclusions (ability to run a power station based on GCSE in Physics) or because the grade awarded was not appropriate to the candidate's ability.

The biggest source of inappropriate awarding of marks is candidates performing better or worse than usual on the day of the exam, but the reliability of the assessment is also a factor. An assessment outcome is reliable if the same result would occur if two different people marked the same assessment. Such variations often occur due to differences in judgements rather than an error and so are more common in certain types of questions (such as essays) than others (short one-mark answers).

Cresswell's paper in 1986 included some technical analysis on the impact of reliability on different numbers of grades.

..., with two grades (pass/fail) the achievement of a candidate whose observed score is, erroneously, just below the borderline is classified as on a par with that of the poorest candidate. With twenty grades such a candidate's achievement might be erroneously graded as 11 when it should be 10, but would certainly not, thereby be indistinguishable from that of the poorest candidates. Thus using a small number of grades leads to a greater proportion of candidates getting their true grades but means that those whose grades are incorrect, suffer very large errors...
(Cresswell, 1986)

So in summary, the more grades there are, the greater the chance of misclassification but the smaller the impact of that misclassification (see also Bradshaw and Wheeler³, 2012). Two of the major factors that affect the number of grades are:

- the reliability of the results of the assessment;

³ Bradshaw, J. & Wheeler, R. (2012) *International Survey of Results Reporting*, Opposs, D. & He, Q. (eds.), Ofqual's Reliability Compendium, pp.557–603. Coventry, Ofqual.

- the interpretations attached to the grades that are meaningful and distinguishable between the grades.

The conclusions from these pieces of research provide the foundations for our substantive discussions on how we ensure that any way of producing GCSE outcomes (for example, grades) is appropriate.

Should we change the way we report GCSE qualifications?

Currently, GCSEs are reported on an eight-point scale, which has been used for many years. The tables below outline the arguments in favour of and against changing this system as part of this reform.

Continue with the current system (A*–G)

Arguments in favour	Arguments against
<ul style="list-style-type: none"> ■ The existing system is familiar to users (especially employers who often take the longest to adapt to new systems). ■ No risks are associated with changing to the new system at the same time that we are introducing a range of other changes to GCSEs. ■ The existing grading system appears to broadly meet the needs for differentiation. ■ There is the option of amending the grading system slightly, by merging or separating specific grades. 	<ul style="list-style-type: none"> ■ The current system will not allow users to differentiate between students with reformed and existing GCSEs. If the reformed GCSEs have a higher status than the existing GCSEs, students will need to highlight which version they have done in a different way. ■ There is no indication that the standard of the new grade C will actually be the same as that of the existing grade C, given the other changes to the qualification. In particular, the existing grade descriptors may no longer fit new grades and/or assessment objectives. ■ There could be a public perception that there has been no change in standards, which would be inaccurate.

Develop a new reporting system

Arguments in favour	Arguments against
<ul style="list-style-type: none"> ■ A new system will provide a clear break from existing qualifications. It will be clear immediately which version of the GCSE the student has taken. This also limits any risks of confusion during the period of “mixed economy” when both existing and reformed GCSEs are in use, which will exist for many years in the workplace. ■ The new system can be designed to meet exactly the needs of the new qualifications. ■ If the new system were based on grades this would have the advantage of a concept that is familiar to users, although if a scaled score or a hybrid model were adopted this would not be the case. 	<ul style="list-style-type: none"> ■ A new system requires more work to define what is meant by different grades. ■ It will require careful management of the concept of “equivalence” between old and new grades.

In his 7th February 2013 letter to us on the reform of the GCSE, the Secretary of State said:

“I consider there to be a strong case for the reformed GCSEs to have a new grading scale, to reflect the step change in expectations for pupils, and would welcome your advice on this. Any changes should apply across all subjects, and should differentiate performance more clearly, particularly at the top end.”

We welcome this clarity and are required to give due regard to this steer when considering the different options.

Please refer to section 4 of the consultation document⁴ for details of our recommendation and the opportunity to comment on this proposal.

⁴ Ofqual (2013) *GCSE Reform Consultation – June 2013*

See www.ofqual.gov.uk/2013-06-11-gcse-reform-consultation-june-2013.pdf

Argument for grades and/or scaled scores

If we explore the possibility of changing the reporting system, the first question to address is whether to use grades at all or just give marks (or scaled scores, see Bramley⁵ (2013) for a discussion about scaled scores). There is a range of views on this, illustrated by the quotes below from a discussion we held with assessment experts:

“Using anything other than scaled scores is just throwing away information.”

“Grades are a useful shorthand.”

On considering the balance between these viewpoints the table below gives some of the arguments either way.

Advantages and disadvantages of grades and scaled scores

	Arguments in favour	Arguments against
Grades	<ul style="list-style-type: none"> ■ relatively easy to compare between students and subjects; ■ relatively straightforward to regulate; ■ hide a certain amount of unreliability (in a technical sense), as variability within a grade is not significant. 	<ul style="list-style-type: none"> ■ do not allow fine-grain distinctions to be made between students; ■ create a “cliff edge” at grade boundaries, where a small increase in marks makes a significant difference in outcome, and this is particularly significant in terms of school accountability measures.
Scaled Scores	<ul style="list-style-type: none"> ■ provide far more detail about the student; ■ provide a better level of information for school accountability measures. 	<ul style="list-style-type: none"> ■ there is still a need for some kind of boundary mark to indicate to the general public what a “pass” looks like; ■ encourage over-discrimination, for example 75% is better than 74% (even if scores are accurate within plus or minus 4%).

⁵ Bramley, T. (2013) *The Case for Scale Scores – Reporting Outcomes in the Reformed GCSE*.

Cambridge, Cambridge Assessment. Available online at:

www.cambridgeassessment.org.uk/ca/digitalAssets/207175_ScalingAndGradingReformedGCSE_TomBramley_250413_FINAL.pdf

How many grades should there be?

Cresswell's paper in 1986 looked at the approaches for determining the number of grades:

It is evident that there is no generally accepted rationale for deciding the number of scale points which should be used to report examination results... (Cresswell, 1986)

If we now extend our modelling of how the reporting system might change, to also assume that we retain the model of reporting GCSEs using grades, the next question to consider is the number of grades that should be included.

In answering this question it is important to go back to the purpose of the grades. In the first section of this paper we outlined three different purposes, and the main considerations behind the number of grades need to relate to them:

- What are the key meanings we want grades to have? With the current policy, for example, there is an expectation for a grade indicating a specific level of knowledge (good pass) and a readiness to progress to A levels⁶.
- How much differentiation do we want between students at different parts of the scale? The Secretary of State's steer is for more differentiation at the top grades.

Looking at international comparisons there is not a clear conclusion. A very wide range of models exists, with a range of four to six grades or 10 to 20 being most common, although there are examples of 100 or more. The consensus from discussions with subject and assessment experts is that the current number of eight is about right, but this does not mean that the current distribution of grades best meets the policy steer.

Should there be additional information?

Grades can be a blunt measure of attainment/performance, particularly with a compensatory model⁷, so it is worth considering whether additional information

⁶ In this paper we make no claims about the validity of using grades for these purposes. The interested reader is directed to the wide body of external research evidence that discusses these issues.

⁷ The compensatory approach to assessment is that strong performance (many marks) in one area can compensate for weakness (few marks) in another. All GCSEs and A levels are currently based on compensation. The alternative, a mastery model, requires that the candidate achieves a certain

should be supplied together with grades to help users differentiate between candidates.

Probably the simplest piece of additional information is the scaled mark. However, that would bring many of the disadvantages of using the scaled mark instead of the grade, while retaining the disadvantages of a grading approach.

Another approach is to use sub-grades within the main grade. For example, a candidate might obtain a grade i, and a sub-grade of 2 (which might be reported as i.2). This has the advantage of allowing greater differentiation between candidates while making it clear that the sub-grade is different to the main grade. By making a clear distinction between the two types of grading, it also presents the possibility of allowing very different approaches in determining these sub-grades.

These sub-grades would have all the issues around reliability of having more categories, but it is possible this may be more acceptable as they are less important/significant than the main grade, and awarding organisations could be clear about this. There is a strong likelihood that these subtle differences will not be appreciated by many users of the qualification and this represents a considerable risk to this approach.

The third option is the inclusion of a profile with the qualification. In addition to their grade, candidates could be provided with information on their achievement in specific topic areas/learning outcomes:

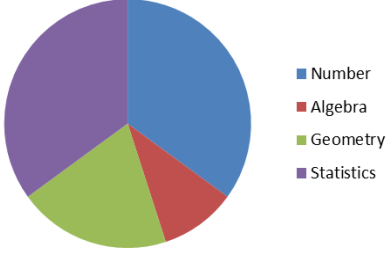
- Profiles do give the possibility for far more sophisticated discrimination between students, for example if a particular topic was essential for study in a different subject (for example, statistics for a student going on to study psychology).
- Almost inevitably, the profiles will need to look different for different subjects, for mathematics it might be by topic area, while for history it is likely to relate to different skills.
- One potential problem with profiles is that they could be seen as “endorsements”, so creating unintended consequences. The most obvious risk here is around introducing a “hurdle” for numeracy and/or literacy through the back door.
- Such profiles potentially have very low levels of reliability as they will be based on a small number of questions.

standard in all the required areas. An example of this might be a driving test: poor quality steering cannot be compensated for by excellent gear changing.

As an example, the following table indicates a range of ways these profiles, and other information, could work alongside grades.

Reporting profiles alongside grades

<p>1. Status quo</p> <p>A single grade awarded based on the total score.</p>	<p>Jenny Smith achieved a grade B in her Mathematics exam.</p>
<p>2. Status quo with error margin</p> <p>A single grade awarded with an indication of the reliability.</p>	<p>Jenny Smith achieved a grade B in her Mathematics exam. There is a 96% certainty of this result.</p>
<p>3. Grade and mark with error margin</p> <p>A single grade with a mark, which also includes an indication of the error margin with this mark (but not grade).</p>	<p>Jenny Smith achieved a grade B in her Mathematics exam. She obtained 144 marks with a 95% certainty of being within plus or minus five marks.</p>
<p>4. Place within grade</p> <p>A single grade awarded with information on where in the grade the student was (bottom 25%, and so on). All based on an overall score.</p>	<p>Jenny Smith achieved a grade B in her Mathematics exam. Her mark placed her between 25% and 50% through the grade.</p>
<p>5. Rank within grade</p> <p>A single grade awarded with information on where the student is in the rank order of students for that grade.</p>	<p>Jenny Smith achieved a grade B in her Mathematics exam. She was in the 50% to 75% most successful students in this grade.</p>
<p>6. Where marks came from</p> <p>A single grade awarded with information on where the marks were scored.</p>	<p>Jenny Smith achieved a grade B in her Mathematics exam. Her marks came from the following areas:</p>

	
<p>7. Include sub-grades</p> <p>A grade based on overall marks, but also with grades in individual subject areas.</p>	<p>Jenny Smith achieved a grade B in her Mathematics exam. This overall grade is based on:</p> <ul style="list-style-type: none"> ■ Number – grade A ■ Algebra – grade D ■ Geometry – grade C ■ Statistics – grade A.
<p>8. Threshold grade</p> <p>A grade based on passing a threshold level in each subject area.</p>	<p>Jenny Smith achieved a grade C in her Mathematics exam (because she only reached the threshold in two of the four subject areas).</p>
<p>9. Threshold grade with detail and overall marks</p> <p>A grade based on passing a threshold level in each subject area but also with an overall mark.</p>	<p>Jenny Smith achieved a grade C in her Mathematics exam.</p> <p>She reached the threshold in Number and Statistics.</p> <p>Overall, she achieved 72%.</p>

The use of results in performance tables is a particular case where there may be considerable benefit from reporting additional or different results for this purpose: scores for performance tables; and grades for students.

Finally, awarding organisations already provide further information to schools on individual candidate’s performance. This approach could be extended to be available direct to the candidate if it is requested, but it would not form part of the formal reporting of the result.

Please refer to section 4 of the consultation document⁸ for details of our recommendation and the opportunity to comment on this proposal.

⁸ Ofqual (2013) *GCSE Reform Consultation – June 2013*

See www.ofqual.gov.uk/2013-06-11-gcse-reform-consultation-june-2013.pdf (accessed 11th June 2013).

We wish to make our publications widely accessible. Please contact us if you have any specific accessibility requirements.

First published by the Office of Qualifications and Examinations Regulation in 2013

© Crown copyright 2013

You may re-use this publication (not including logos) free of charge in any format or medium, under the terms of the [Open Government Licence](#). To view this licence, visit [The National Archives](#); or write to the Information Policy Team, The National Archives, Kew, Richmond, Surrey, TW9 4DU; or email: psi@nationalarchives.gsi.gov.uk

This publication is also available on our website at www.ofqual.gov.uk

Any enquiries regarding this publication should be sent to us at:

Office of Qualifications and Examinations Regulation	
Spring Place	2nd Floor
Coventry Business Park	Glendinning House
Herald Avenue	6 Murray Street
Coventry CV5 6UB	Belfast BT1 6DN

Telephone 0300 303 3344

Textphone 0300 303 3345

Helpline 0300 303 3346