

How many grades should a qualification have?

Summary

- Experts are divided on this question. Some experts argue for few grades on the basis that examinations are relatively crude measures whose reliability should not be overstated. With this approach, those seeking better distinction between candidates would need to look to other measures beyond academic achievement to build a fuller picture of applicants.
- Other experts argue that many grades should be available so that information from the examination is not lost. This approach would allow universities to take a much more fine-grained approach in their admissions processes.
- In the current A-level system, the vast majority of students are classified accurately, to within one grade – evidence that there are not too many grades.
- Any change to the number of grades available must balance the need for accurate classification (fewer grades) with the need to be able to distinguish adequately between students of different ability (more grades).

Introduction

The number of grades used to report examination results varies considerably. For example, in the UK, GCSE examinations use nine grades (A*, A, B, C, D, E, F, G, U); A-levels use seven (A*, A, B, C, D, E, U); and the Principal Learning and Project qualifications use seven at Level 3 (as in A-level), five at Level 2 (A*, A, B, C, U) and four at Level 1 (A*, A, B, U).

It is possible that, as part of the A-level reform process, the number of grades offered may be increased, with the intention of enhancing the ability to distinguish between the highest grades and so help universities with their selection process. However, there are many factors to consider when deciding the number of grades which should be used to report examination results; these are summarised briefly below (for an extended discussion see Meyer, 2011).

Grading

Assessment theory tells us that candidates of a given ability have a certain probability of achieving a certain grade. Once they have taken an exam, it is possible to determine the probability that they would have received a different grade. This probability is a measure of the grading accuracy of an examination. For individual candidates, the closer they are to a grade boundary the greater the chance that they could have achieved a different grade or, in other words, have been misclassified (Figure 1).

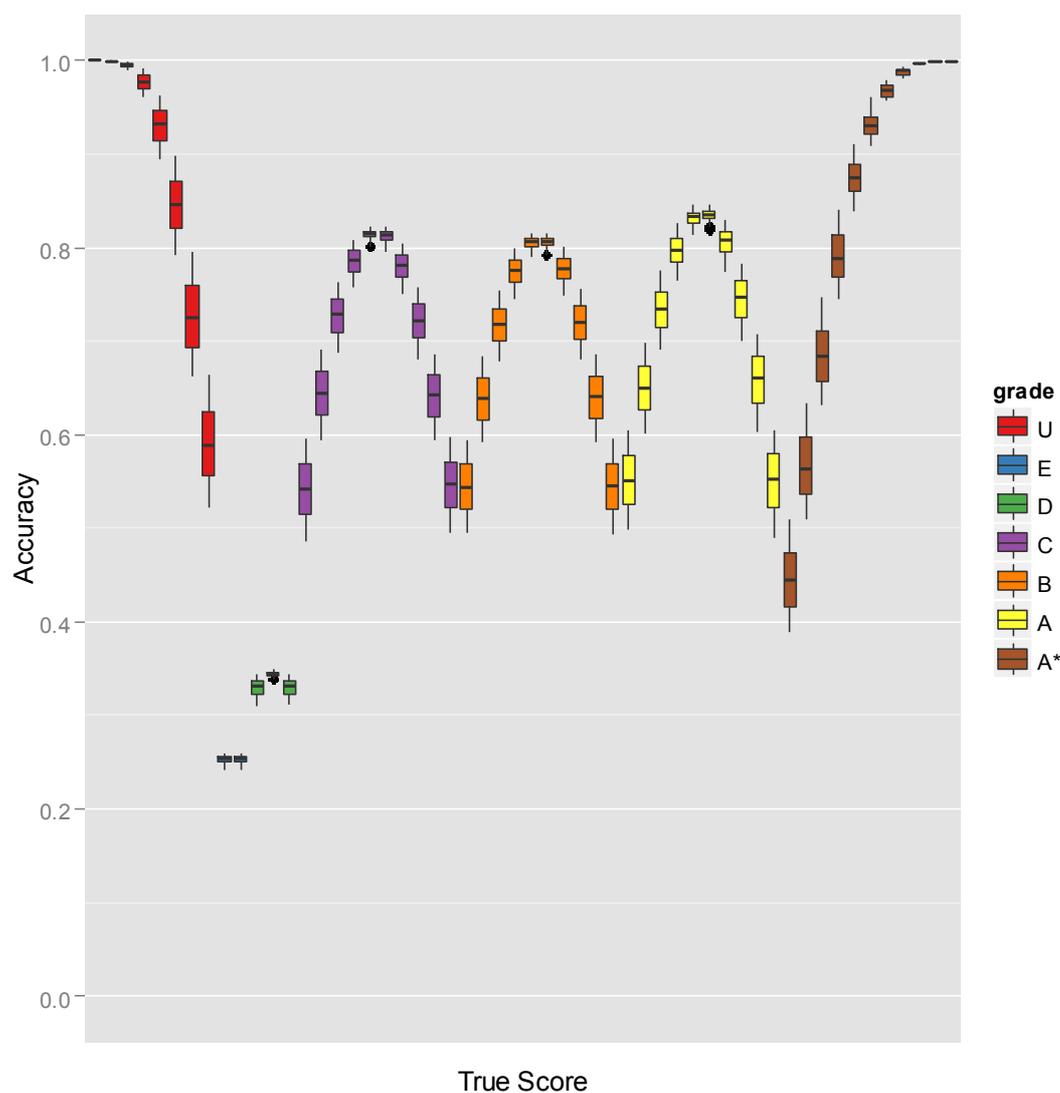


Figure 1 Chances of learners of different ability being correctly classified on an exam with multiple grade boundaries

As the number of grades on an assessment is increased the distance between each grade boundary gets smaller and the number of misclassifications increases. Decreasing the grade boundaries decreases the rate of misclassification, but increases the potential severity of the misclassification. At university, where there are few grades, the probability of a 2:1 candidate being misclassified as a 2:2 is small, but the consequences are quite restrictive in terms of future opportunities. At A-level, where there are more grades, the probability of an A* candidate receiving a grade A is probably higher than a misclassification at university. In an ideal world, the impact of such a misclassification should be less than at degree level, though there are some examples of university offers with very specific requirements being sent to A-level students. In such cases, being awarded an A instead of an A* can have a significant impact on the individual affected.

So how many grades should there be?

Experts have remained divided on this question since it first arose (Cresswell, 1986). In one camp are those that relate the number of grades to the reliability of the underlying mark scale (see Skurnik and Nuttall (1986); Pilliner (1969); Please (1971); Mitchelmore (1981)). In its most extreme form this camp believes that no candidate should ever be misclassified by more than one grade (Pilliner 1969) or that no candidate should have a greater than 50 per cent chance of being misclassified (Cresswell, 1986). Use of more grades, they argue, overestimates the

reliability of an examination. Exponents of this view tend to prefer very few grades. Indeed simulations can be run to show exactly how many grades a qualification can support so that this minimum quality is maintained (Pilliner, 1969). The relatively rare occurrence of misclassification, it is argued, maintains high public confidence in the examination system. The drop in ability to distinguish between candidates based on grades should be compensated for by use of other contextual information giving a more rounded picture of applicants rather than more and more detail regarding academic attainment.

Those who disagree with this view argue that reducing a lengthy examination to relatively few grades causes an unacceptable loss of information (see, for example, Ebel (1965) and Taylor (1977)). Under a coarse grading system candidates with very different marks on an examination will receive the same grade. When selection decisions are made on combinations of coarse grades, such as the A-level system where offers are made dependent on a grade profile of ABB, for example, it is easy to show how inequities can arise. A candidate with three high grade Bs could lose out to a candidate with an ABB profile even though the three B candidate scored higher overall marks (table 1). Exponents of this view tend to prefer a large number of grades. A review in 1983 argued for 20 grades at A-level (JMB, 1983). More recently there has been a call for ranking of candidates, which would essentially remove grades completely, but there are substantial issues with this approach (see [Is ranking A-level students useful?](#) in the CERP A-level reform series).

Table 1: Candidate B has more marks than candidate A, yet worse grades.

Paper	Candidate A		Candidate B	
	Grade	Mark	Grade	Mark
1	A	80	B	79
2	B	70	B	78
3	B	70	B	78
Total		220		235

There is a third camp, who argue that grades should be qualitatively different, but as Stringer (2011) cogently argues, grades shade into one another. The notion that there are clear distinctions between grades, although intuitive, is undermined by the empirical evidence.

Considering the current number of grades it is possible to make two observations. Firstly, that the vast majority of candidates are classified correctly to within a grade (Wheadon & Stockford, 2011). This finding would suggest that there are not currently too many grades.

So are there too few grades? Before the addition of an A* grade that may have been the case, as universities were finding A-level grades too coarse a scale on which to make decisions (see [The A* at A-level](#) in the CERP A-level reform series). If universities are now satisfied, there may however still be a reason for action.

This would be if research suggested that, as illustrated by table 1, unfair selection decisions were being made with composite grade offers, and the situation could be improved by the provision of more grades. If this were the case, it would seem sensible to balance the need for reasonably accurate classification with the need for reasonable distinction between candidates.

Bibliography

- Cresswell, M. J. (1986). [Examination Grades: how many should there be?](#) *British Educational Research Journal*, 12(1), 37–54. doi:10.1080/0141192860120104.
- Ebel, R. (1965). *Measuring Educational Achievement*. Englewood Cliffs, N.J.: Prentice-Hall.
- JMB (1983). *Problems of the GCE Advanced Level Grading Scheme*. Manchester: Joint Matriculation Board.
- Meyer, L. (2011). [Extending the number of grades on a qualification: Issues to consider](#). Manchester: AQA Centre for Education Research and Policy.
- Mitchelmore, M. (1981). [Reporting Student Achievement: How many Grades?](#) *British Journal of Educational Psychology*, 51(2), 218-227.
- Pilliner, A. (1969). *Estimation of number of grades to be awarded in an examination by consideration of its reliability coefficient*. Edinburgh: The Godfrey Thomson Unit for Educational Research.
- Please, N. (1971). [Estimation of the proportion of examination candidates who are wrongly graded](#). *British Journal of Mathematical and Statistical Psychology*, 24(2), 230-238.
- Skurnik, L.S. and Nuttall, D.L. (1968). [Describing the Reliability of Examinations](#), *The Statistician*, 18(2) pp. 119-128.
- Stringer, N. S. (2011). [Setting and maintaining GCSE and GCE grading standards: the case for contextualised cohort-referencing](#). *Research Papers in Education*, 1–20. doi:10.1080/02671522.2011.580364.
- Taylor, H. (1977). [Differences in Grading Systems among Canadian Universities](#). *The Canadian Journal of Higher Education*, 7(1) pp 47-54.
- Wheadon, C. and Stockford, I. (2010). [Classification accuracy and consistency in GCSE and A Level examinations offered by the Assessment and Qualifications Alliance \(AQA\) November 2008 to June 2009](#), report for Ofqual by AQA Centre for Education Research and Policy.