# How to make grading fair

## Document 2 in a series of 3

Dennis Sherwood, July 2017

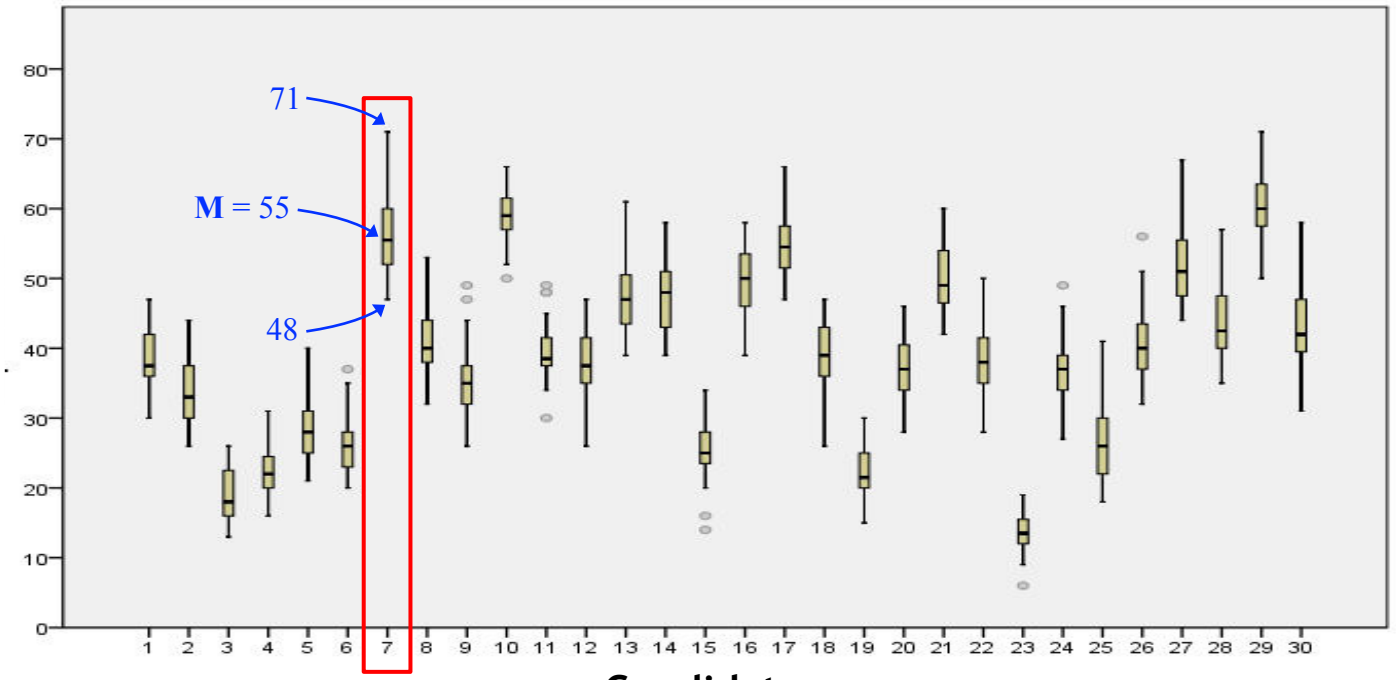
**Currently, at every grade boundary in every GCSE and A level examination, there is a population of candidates who are awarded a grade lower than they merit, so denying these 'disadvantaged' candidates valuable life chances.**
**This is fundamentally unfair.**

**This unfairness is not inevitable - it is a consequence of the current policy of assigning grades, a policy that ignores the inherent variability in marking.**

**As described in the accompanying document *The great grading scandal*, by adopting a grading policy which takes the variability in marking into account, the number of 'disadvantaged' candidates can be reduced to close to zero.**

**This document presents an overview of how this can actually be done.**

# Different markers can give the same script different marks

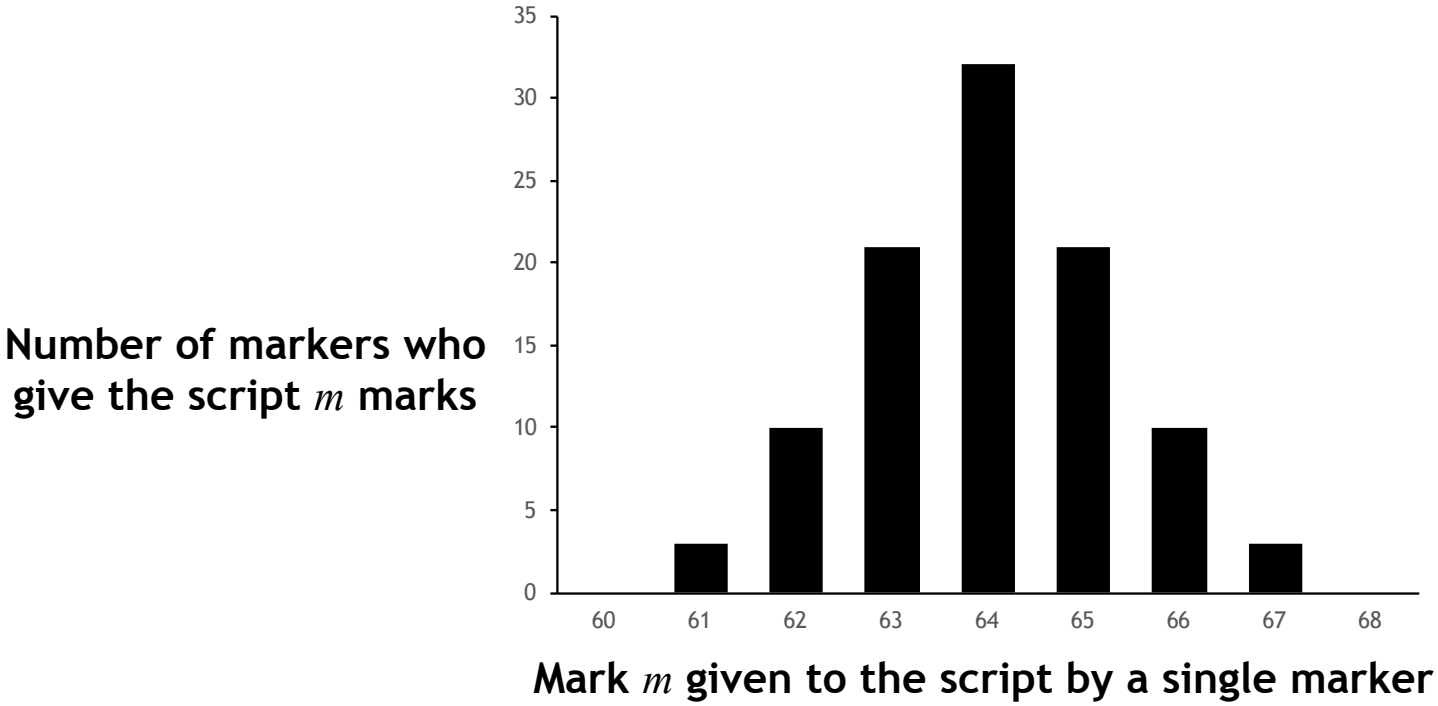



**This diagram* shows the marks given by 40 different markers to each of 30 GCSE History scripts. For each candidate, the heavy, central line, is the median mark, M; the 'box' shows the range of marks given by 50% of markers; the 'whiskers' encompass more than 95% of markers; the 'bubbles' show outliers. This variability is not a result of marking errors, nor of poor control over the quality of the marking process: it is a consequence of an examination system based not on right/wrong multiple choice, but on more open-ended questions.**

**Consider, for example, candidate 7, whose marks span the range from 48 to 71. Which of these marks should be used to determine the candidate's grade? If all the marks between 48 to 71 fall within the same grade width, it doesn't matter. But if this range straddles one or more grade boundaries, then the grade awarded depends not on the candidate's ability, but on which marker marks the script. And the awarded grade might therefore be wrong.**

*Figure 3.5 from *Component reliability in GCSE and GCE*, Sandra Johnson and Rod Johnson, Ofqual, November 2010, http://webarchive.nationalarchives.gov.uk/20140402200706/http://ofqual.gov.uk/documents/component-reliability-gcse-gce/all-versions/

# Representing the variability in marking for a single script

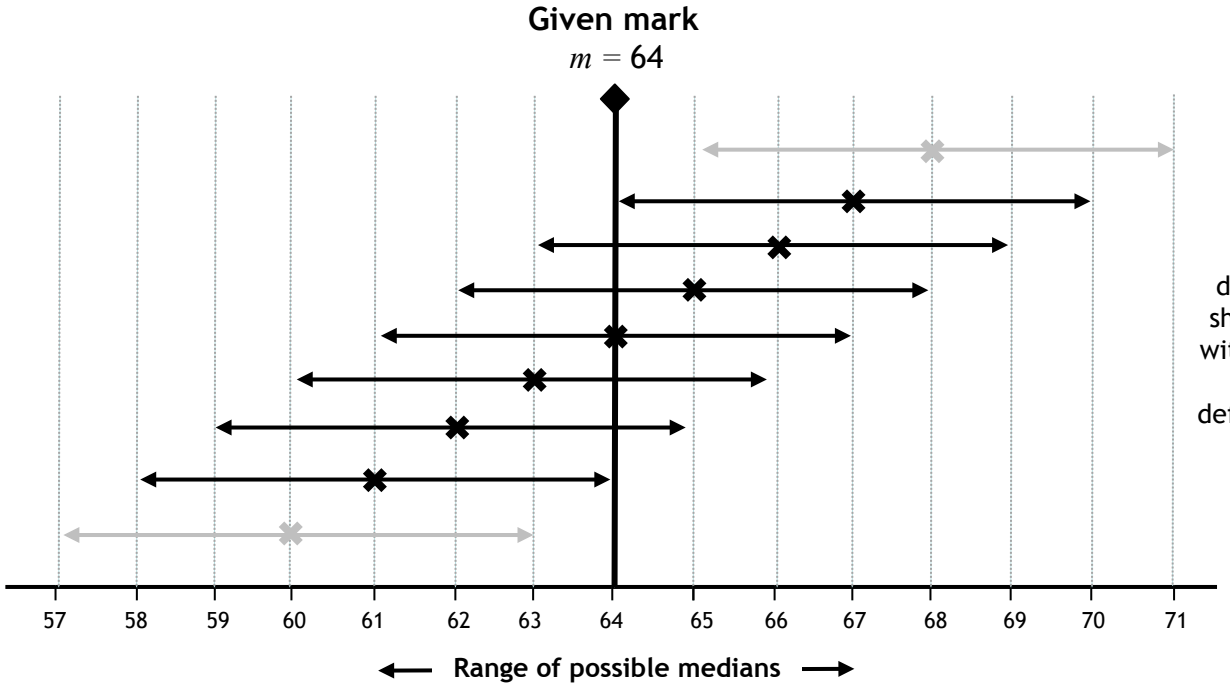



**The previous diagram, on page 2, shows the variability in marking for 30 different scripts, where, for each script, the different marks awarded by different markers are shown as a 'box' diagram. The diagram on this page presents the marks given by a panel of 100 different markers to a particular script in the form a histogram known as the 'panel distribution': 3 markers gave the script 61 marks; 10 markers gave 62 marks; 32 markers gave the median mark, 64; 21 markers gave 65 marks; 10 markers gave 66 marks; and 3 markers gave 67 marks.**

**Any mark $m$ given to this script by a single marker must be a sample from this panel distribution, which has a total width of 6 marks, from 61 to 67. To determine the candidate's grade, we must choose one of these marks as 'the' mark on which the grade is determined. In principle, it doesn't matter which mark is chosen, provided that the same 'rule' is used for all candidates - so suppose, for example, that the policy is that the grade is determined by the median mark, in this example, 64.**

The distribution shown here is hypothetical, and for illustrative purposes; it also happens to be symmetrical, as will be the case for some real distributions, whilst other real distributions will be skewed. The analysis presented in this document will use this symmetrical distribution; the analysis for a skewed distribution is very similar: full details are available from the author who may be contacted at dennis@silverbulletmachine.com.

# The fundamental problem




Each arrow represents a distribution of the shape as shown on the previous page, with the median shown as ✖, and it is this median that defines the mark on which the candidate's grade is determined.

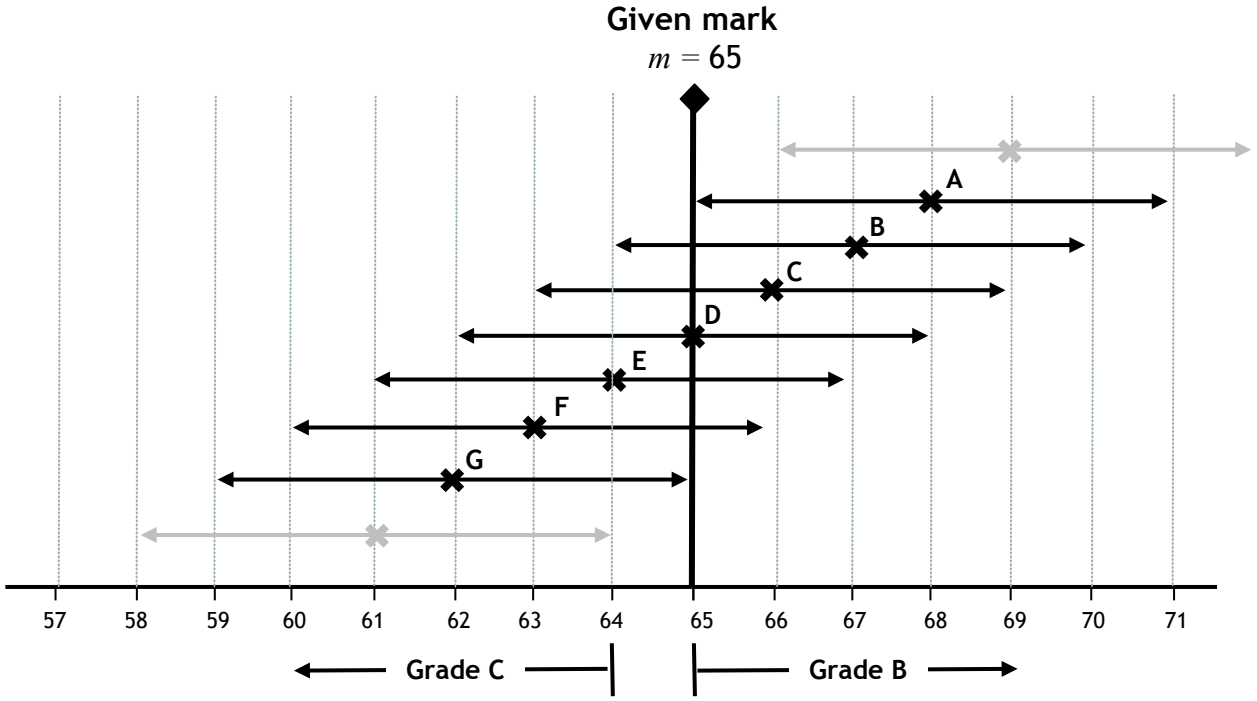
**Suppose that a single script is given a single mark $m$ by a single marker. To determine the candidate's grade, we need to determine the median of the corresponding panel distribution - of the type shown on the previous page - of which this mark is a member. The fundamental problem, however, is that *there is no knowledge as to which, specific, distribution this is*.**

**Suppose further that it is known that the distribution is of the shape as shown on the previous page. Accordingly, since this distribution has a total width of 6 marks, then the mark $m = 64$ must be a member of one of the seven distributions corresponding to medians from 61 to 67 marks, for only these distributions overlap $m = 64$.**

**ANY ONE OF THE MARKS FROM 61 TO 67 MIGHT BE THE MARK ON WHICH THE CANDIDATE IS FAIRLY GRADED.**

**BUT WHICH ONE IS IT? NOBODY KNOWS.**

# Some candidates are ‘lucky’...

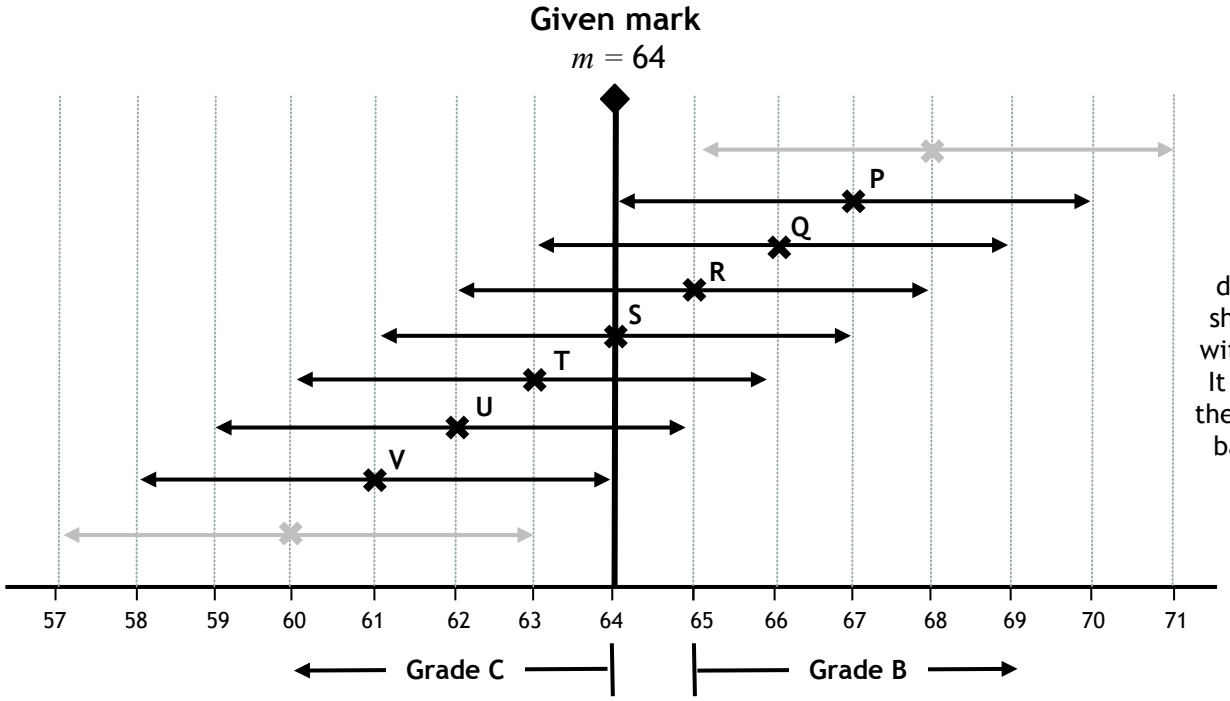



**Seven candidates submit scripts, all marked $m = 65$, and the C/B grade boundary is 64/65. If - as is the current policy - grades are awarded on the basis of the mark $m = 65$, all candidates are awarded grade B.**

**But suppose that the scripts submitted by these candidates are members of the distributions shown by the arrows, each associated with the median shown by ✖. For candidates A, B, C and D, the corresponding medians are within grade B, and so these four candidates are awarded the grade they merit.**

**But for candidates E, F and G, the corresponding medians are within grade C - yet they are awarded grade B. These candidates are ‘lucky’.**

# ...and some candidates are 'disadvantaged'



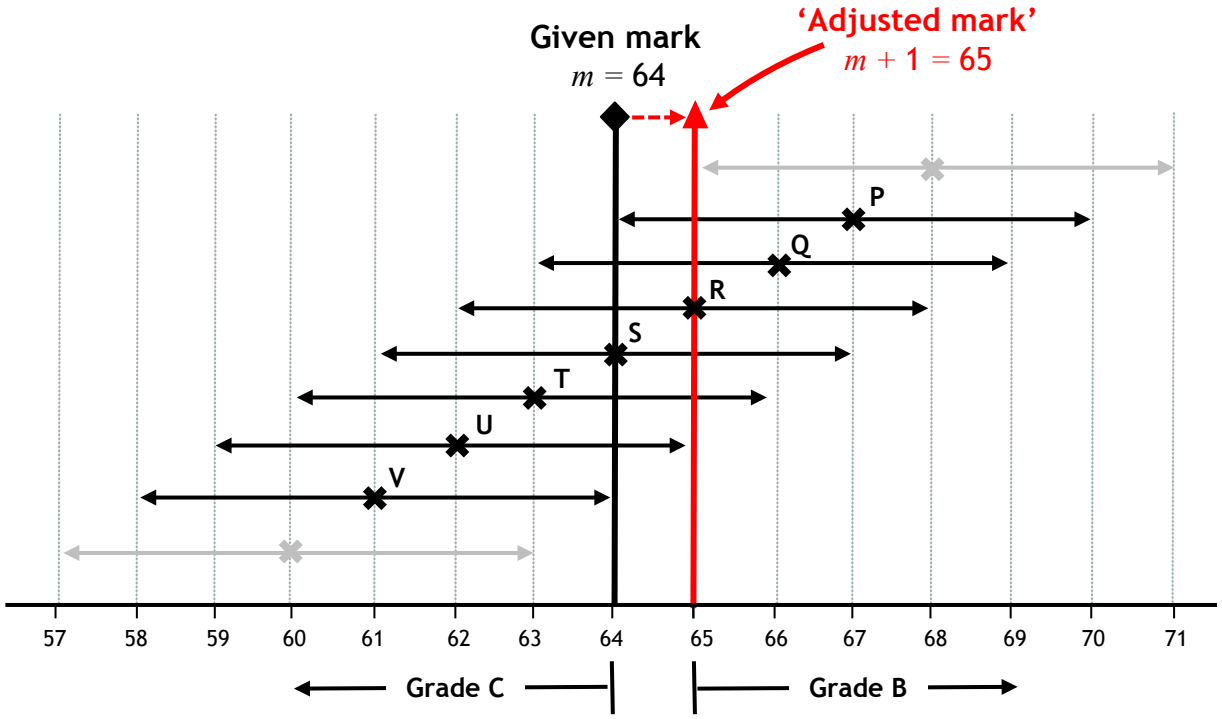

Each arrow represents a distribution of the shape as shown on the previous page, with the median shown as ✖. It is this median that defines the most fair mark on which to base the candidate's grade.

**Seven candidates submit scripts, are all marked $m = 64$, and the C/B grade boundary is 64/65. If - as is the current policy - grades are awarded on the basis of the mark $m = 64$, all candidates are awarded grade C.**

**But suppose that the scripts submitted by these candidates are members of the distributions shown by the arrows, each associated with the median shown by ✖. For candidates S, T, U, and V, the corresponding medians are within grade C, and so these four candidates are awarded the grade they merit.**

**But for candidates P, Q and R, the corresponding medians are within grade B - yet they are awarded grade C.**
**These candidates are 'disadvantaged', and could, as a result, be denied important life chances.**

# Suppose that the grade is based not on the mark $m$, but on $m + 1$...

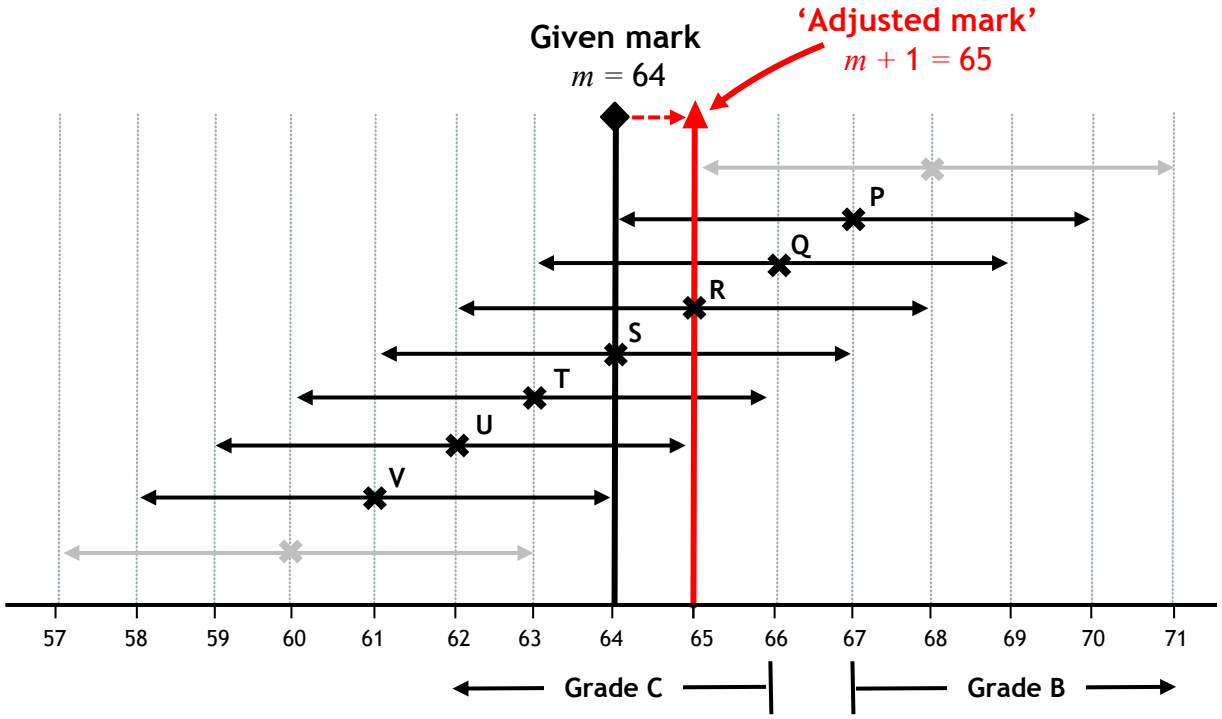



This is (almost) the same diagram as on the previous page: seven candidates submit scripts, all marked $m = 64$, and the C/B grade boundary is 64/65. But there is one, important, difference: the policy for the determination of grades has changed. On the previous page, the policy was 'the grade is determined according to the given mark $m$', and all candidates are awarded grade C. As was shown on the previous page, candidates S, T, U and V are awarded the merited grade, but candidates P, Q, and R are disadvantaged.

Suppose, however, that the policy is changed to 'the grade is determined according to the 'adjusted mark' $m + 1 = 65$'. All candidates are now awarded grade B. Importantly, candidates P, Q and R are now no longer 'disadvantaged', but are awarded the grade they merit, whilst candidates S, T, U and V have become 'lucky'.

This is much more fair to those candidates who are given marks one mark below a grade boundary: the next two pages describe what happens for candidates whose marks are further from the grade boundary...

# Another example of unfairness...

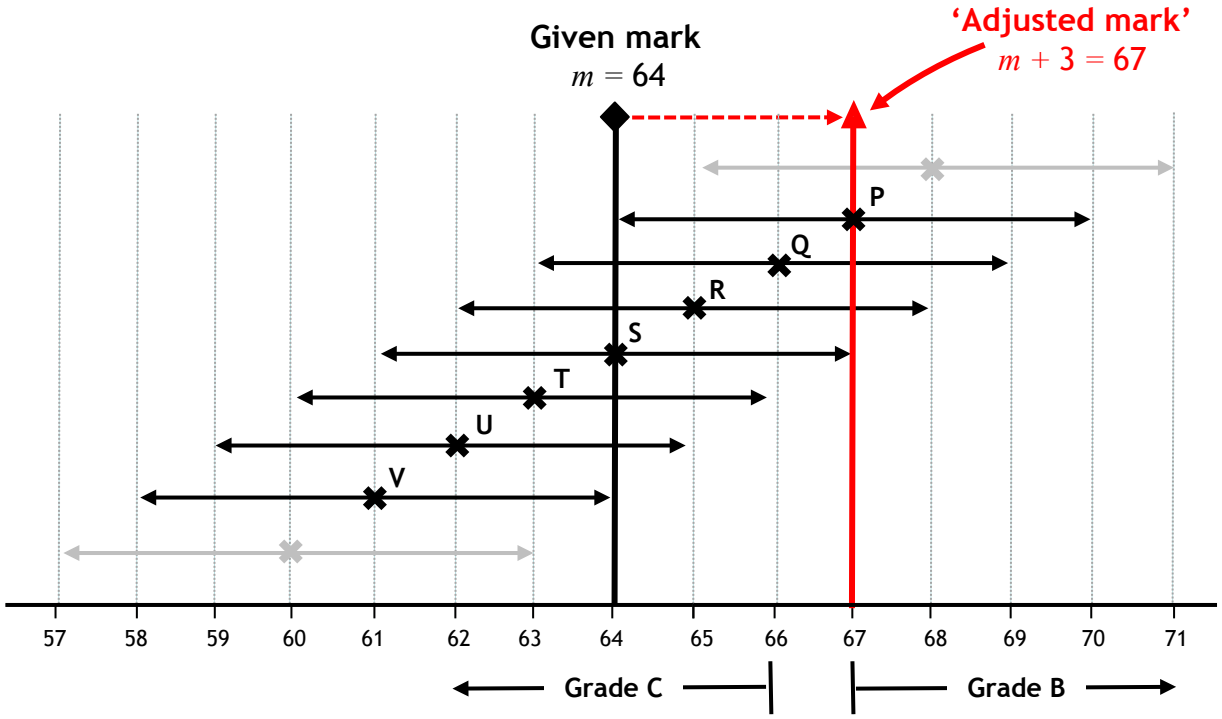



**This is another example of unfairness. As previously, the scripts of seven candidates are all marked $m = 64$, but in this example - in contrast to those shown on pages 6 and 7 - the C/B grade boundary is 66/67.**

**Under the policy of awarding grades based on the original mark $m = 64$, all candidates are awarded grade C. As a consequence, six of the candidates - candidates Q, R, S, T, U and V - are awarded they grade they merit, but candidate P is disadvantaged.**

**Under the policy of awarding grades based on the 'adjusted mark' $m + 1 = 65$, all candidates are also awarded grade C, and it is still the case that six of the candidates - candidates Q, R, S, T, U and V - are awarded they grade they merit. Candidate P, however, remains disadvantaged.**

# ...and another possible solution




**But suppose that the policy for determining grades is 'the grade is determined according to the 'adjusted mark' $m + 3 = 67$'. All candidates are awarded grade B, candidate P is no longer disadvantaged, and candidates P, Q, R, S, T, U and V are all 'lucky'.**

**For an examination for which the variability in marking can be represented by the histogram shown on page 3 - which has a total end-to-end width of 6 marks - no candidate is given a mark $m$ greater than 3 marks away from the histogram's median.**

**The diagram shown on this page is a particular case of a more general rule: for an examination for which the variability in marking can be represented by a symmetrical histogram of width $2f$ marks, a policy of awarding grades based on the 'adjusted mark' $m + f$ takes the variability of marking fairly into account, and so ensures that *no candidate is ever disadvantaged*.**

**Ensuring that no candidate is ever disadvantaged is an important, but not the only possible, policy objective. Different policy objectives - such as ensuring both that no candidates are disadvantaged, and also that grading is robust under appeal - are associated with different ways of estimating the parameter $f$, as discussed in the accompanying document, *How to determine $f$*.**

- At *every* grade boundary, and for *every* subject at GCSE and A level, a (currently unknown) number of candidates are awarded the wrong grade ...
- ...with some 'lucky' candidates being awarded a grade higher than they merit...
- ...whilst others - 'disadvantaged' candidates - are awarded a grade lower than they merit.
- This is fundamentally unfair, and can deny 'disadvantaged' candidates life chances. Nor is this unfairness satisfactorily resolved by the appeals system.
- This grade misallocation is not attributable to erroneous marking, or failures in quality control; rather, it is a consequence of an examination system based on rather 'open' questions, which allows different markers, legitimately, to award different marks to the same script.
- This implies that a script given a mark $m$ is in fact more fairly associated with a range of marks, which can be represented as $m \pm f$, where, for any examination, $f$ is a statistical measure of the variability in marking.
- Currently, grades are awarded on the basis of the mark $m$, and there is no recognition of the variability in marking. As a consequence, grade misallocation occurs at every grade boundary, with the creation of populations of 'lucky' and 'disadvantaged' candidates.
- It is, however, possible to take the variability in marking into account when determining grades: rather than basing grades on the mark $m$, grades could be based, for example, on $m + f$. This does not eliminate grade misallocation - but it does enable the relative numbers of 'lucky' and 'disadvantaged' candidates to be controlled.
- If grades are based on $m + f$, the number of 'disadvantaged' candidates is reduced to close to zero, whilst the number of 'lucky' candidates is increased.
- The choice of grading based on $m$ or on $m + f$ (or indeed on $m - f$) is a policy choice, and this policy debate must be held.


Idea generation, evaluation and development
Making innovation happen
Silver Bullet
Strategy development and scenario planning
The Silver Bullet Machine Manufacturing Company Limited
Building ultimate competitive advantage
Building high-performing teams
Barnsdale Grange, The Avenue, Exton, Rutland LE15 8AH
E-mail: dennis@silverbulletmachine.com
Website: www.silverbulletmachine.com
Mobile and messages: 07715-047947
Telephone: 01572-813690
Training and knowledge transfer
Conferences
Business and market modelling