The great grading scandal

Document 1 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 this document shows, 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 policy choice is important, and the policy debate must be held openly, widely, and robustly.

The problem

# Different markers can give the same script different marks
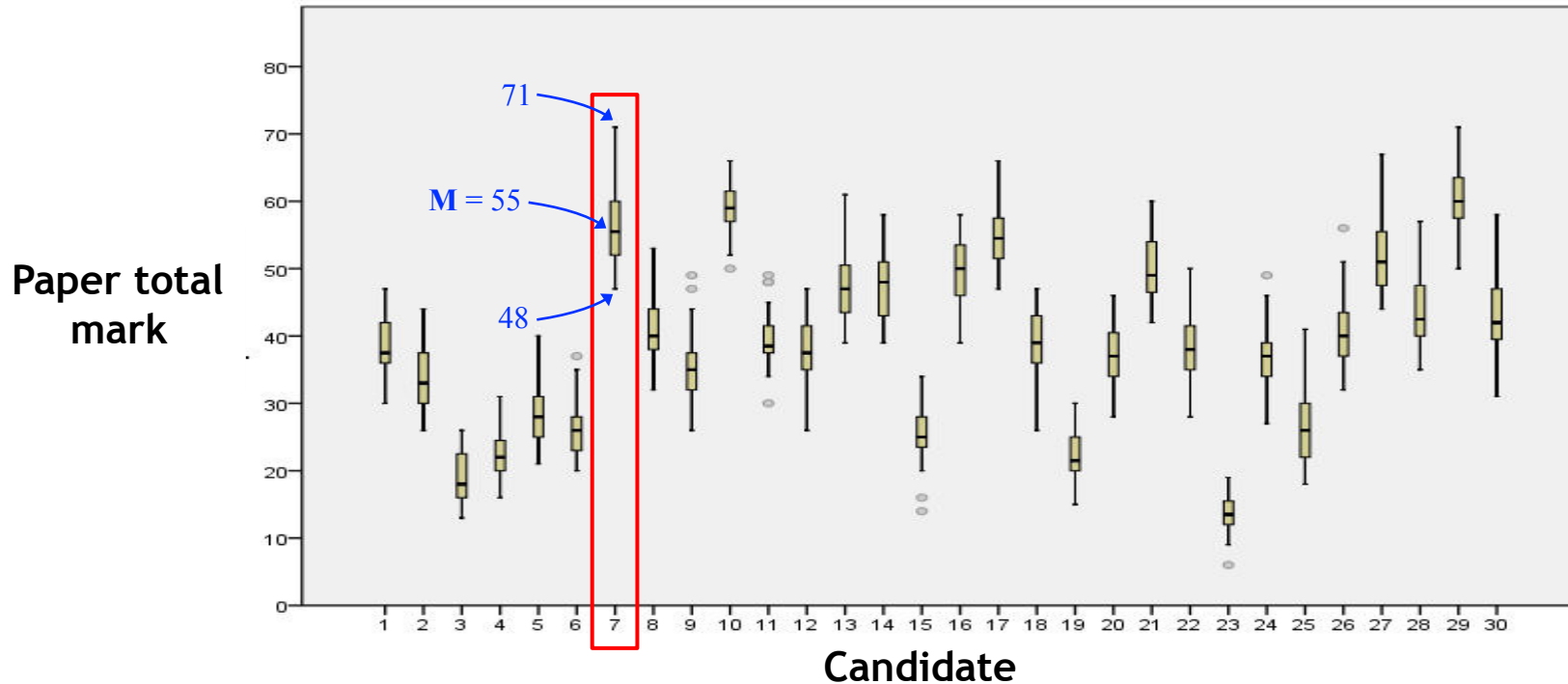
This diagram* shows the marks given by 40 different markers to each of 30 GCSE History scripts. For each candidate, the heavy, central line, is the median mark, M; the 'box' shows the range of marks given by 50% of markers; the 'whiskers' encompass more than 95% of markers; the 'bubbles' show outliers. 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.**

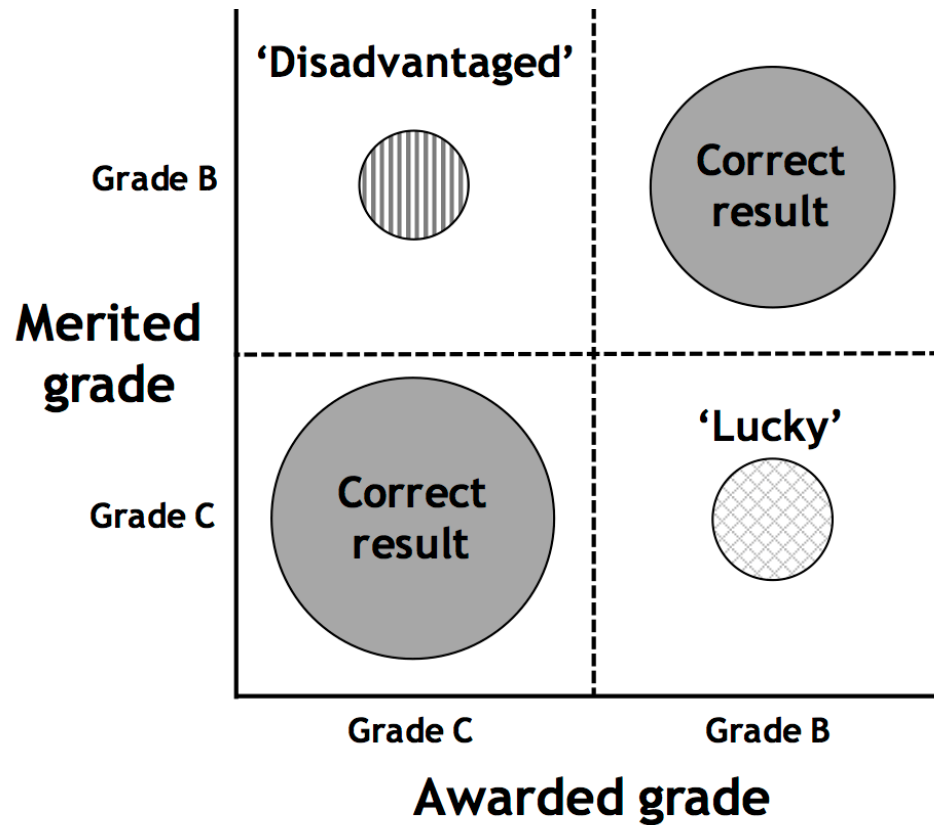# How many candidates are awarded the wrong grade?
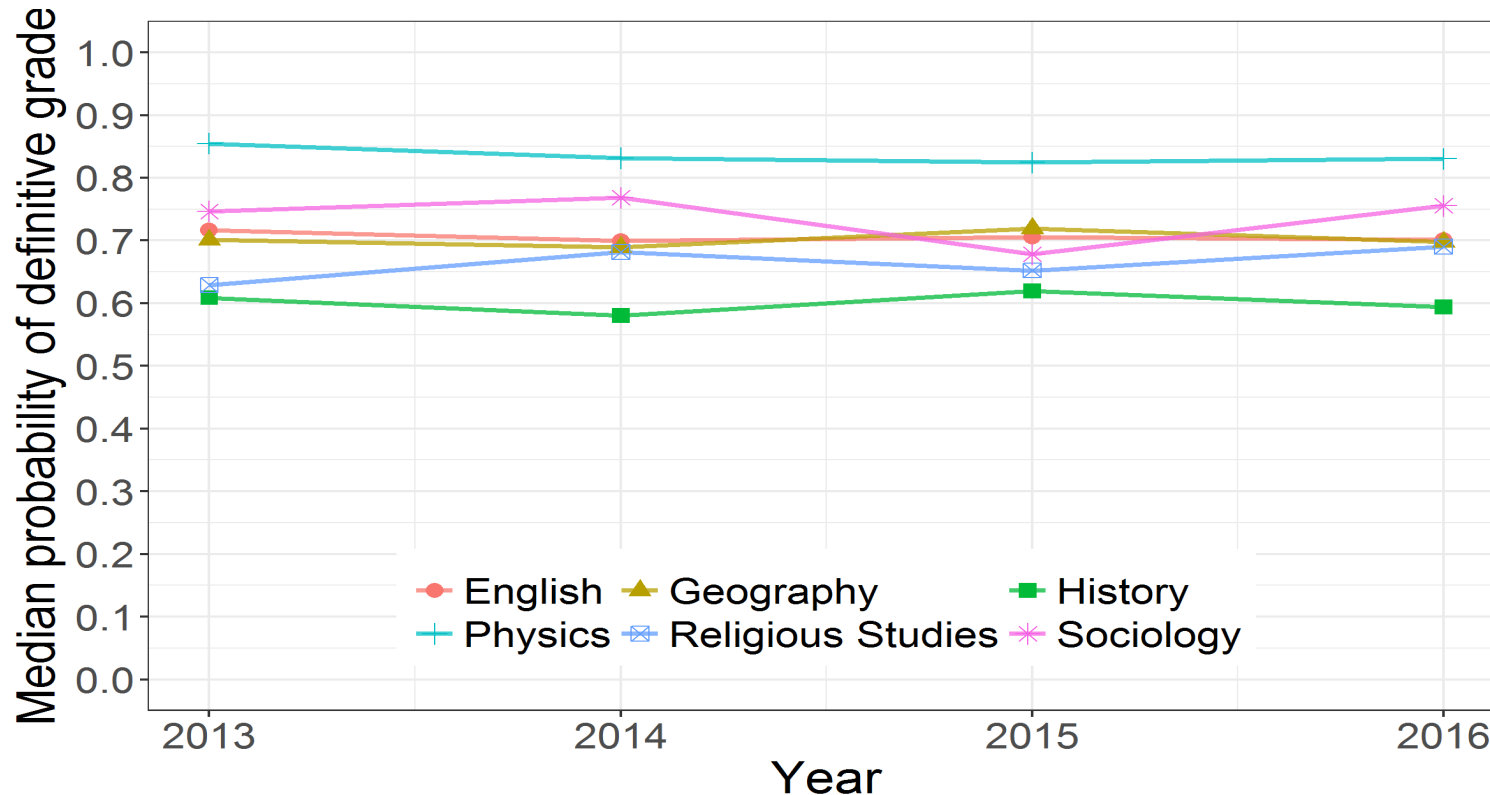
The diagram on the previous page illustrated the variability in marking, and discussed the possibility that the range of marks associated with a particular script might straddle a grade boundary. This diagram* shows an important consequence of this - the probability that the grade awarded to a GCSE or A level candidate in a given subject is the grade the candidate actually merits.

These numbers are shocking. For the last four years, only about 60% of candidates in history have been awarded the grade they merit. **And 40% of candidates haven't.** Likewise for the other subjects - but even in a 'precise' subject such as physics, 15% of the candidates are awarded the wrong grade.

To make that real: each year, some 500,000 candidates take GCSE English Language. Approximately 30% of those candidates - that's about **150,000 candidates annually** - have been awarded **the wrong grade**.

Grade misallocation across any grade boundary

The diagram on the previous page showed that, in every subject, some candidates are awarded the grade they merit, whereas others are awarded the wrong grade - either a higher grade, or a lower grade. This diagram presents a different way of looking at 'grade misallocation' across a grade boundary, this example being the C/B boundary. As can be seen, some candidates who are awarded grade B merit grade B, and likewise for grade C. However, as a result of grade misallocation, some candidates who merit grade C are awarded grade B, and so are 'lucky', whilst there are others, who merit grade B but are awarded grade C, and so are 'disadvantaged'.

Currently, however, neither Ofqual nor the examination boards routinely publish data, or estimates, of the numbers of candidates who are either 'lucky' or 'disadvantaged' in each of the subjects for GCSE and A level - this is all hidden.

The 'appeal and re-mark' process
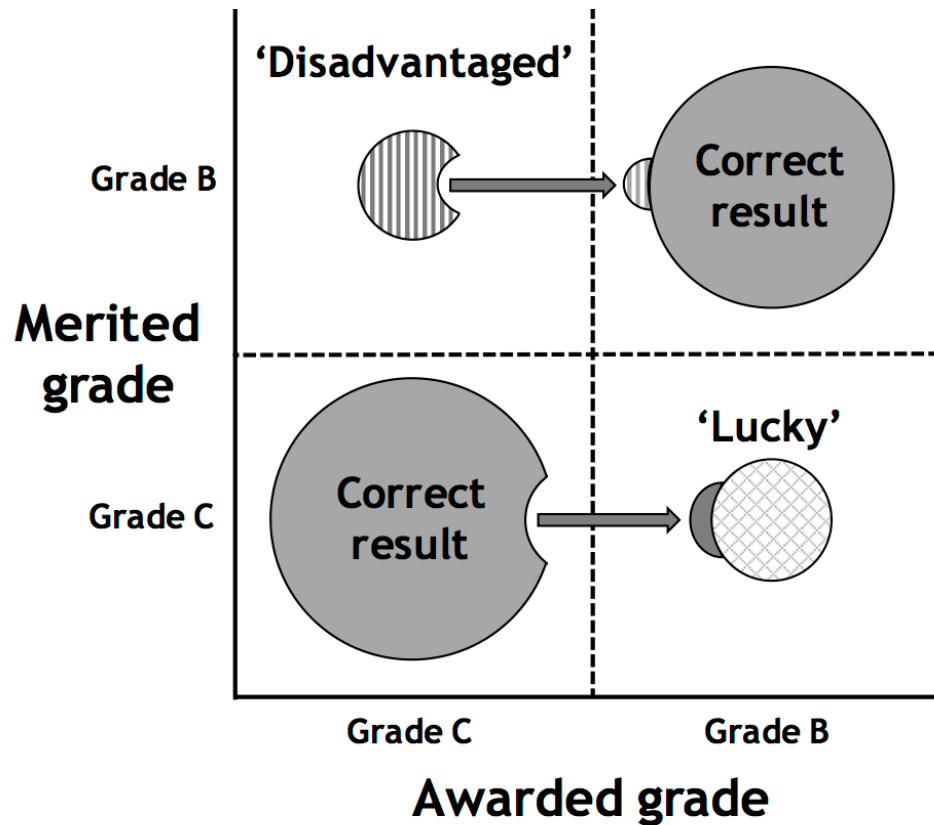
It might be thought that the process whereby a candidate can appeal the awarded grade, so resulting in a re-mark, would resolve the original grade misallocation. In fact, it doesn't - it simply makes the muddle worse.

Firstly, there is no reason why any 'lucky' candidate might appeal, so the population who were originally 'lucky' remains. Secondly, since candidates do not know they have been 'disadvantaged', many do not appeal: perhaps the appeal fee is a disincentive; perhaps they just trust 'the system', and accept the lower grade, thinking that they had not done as well as they had hoped. And thirdly, since appeals can be raised by any candidate, not just by those who were originally 'disadvantaged', appeals by candidates who were originally awarded the grade they merit can result in a re-mark that makes them 'lucky' - as indeed can, and does, happen.

The $m \pm f$ solution

A 'thought experiment'…

Grade misallocation is real, and exists at *every* grade boundary, for *every* subject at both GCSE and A level.

But it is hidden. The candidate does not know whether he or she has been awarded the grade merited, or has been awarded the grade above, and so is 'lucky', or has been awarded the grade below, and so is 'disadvantaged'. Nor does the 'appeal and re-mark process' resolve matters. Furthermore, neither Ofqual, nor the Examination Boards, routinely publish any relevant statistics on grade misallocation, or grade reliability.

As illustrated on page 3, grade misallocation is fundamentally attributable to the fact that different equally qualified markers can legitimately award different marks to the same script: marking is subject to variability. The *BIG PROBLEM* is that, currently, this variability is not recognised when the mark given to any script is mapped onto the corresponding grade. But suppose that it is possible to measure this variability for any examination so that any script, originally given a single mark m by a single marker, could then be associated with a range of marks, represented as $m \pm f$. Accordingly, for an examination for which $f = 3$ marks, the variability in marking can be taken into account by associating a script originally marked $m = 64$ with a range of marks $m \pm f$ from 61 to 67, where $m - f = 64 - 3 = 61$, and $m + f = 64 + 3 = 67$.

Here is a 'thought experiment'…

Currently, grades are assigned according the mark m, ignoring the variability in marking. What might happen if this variability is explicitly recognised, so that the grade is assigned *not* according to the mark m, but according to the 'adjusted' mark $m - f$, or the 'adjusted' mark $m + f$?

In fact, it *is* possible to estimate the value of f, which measures the variability of marking. To do so requires some statistical analysis, the details of which are not presented here, but are available on request*. These details are important, but if f can be reliably estimated, then even more important is the outcome - what actually happens when grades are based on $m - f$ or $m + f$ - an outcome that addresses the *BIG PROBLEM* by explicitly recognising the variability of marking when assigning grades, as described on the following pages…
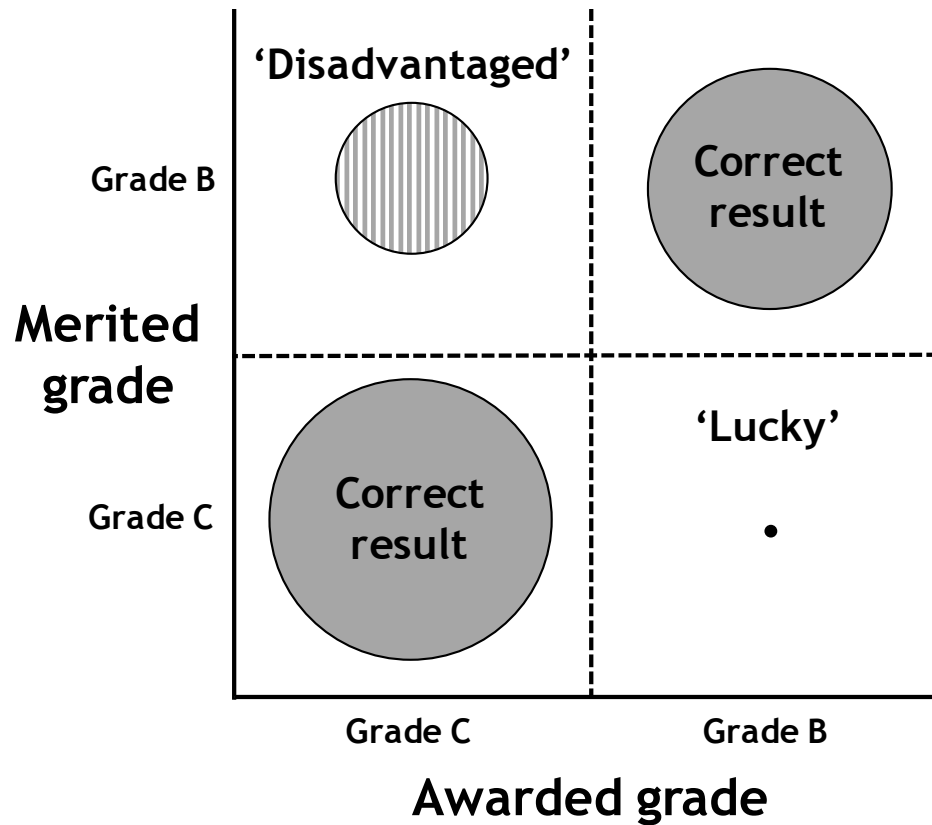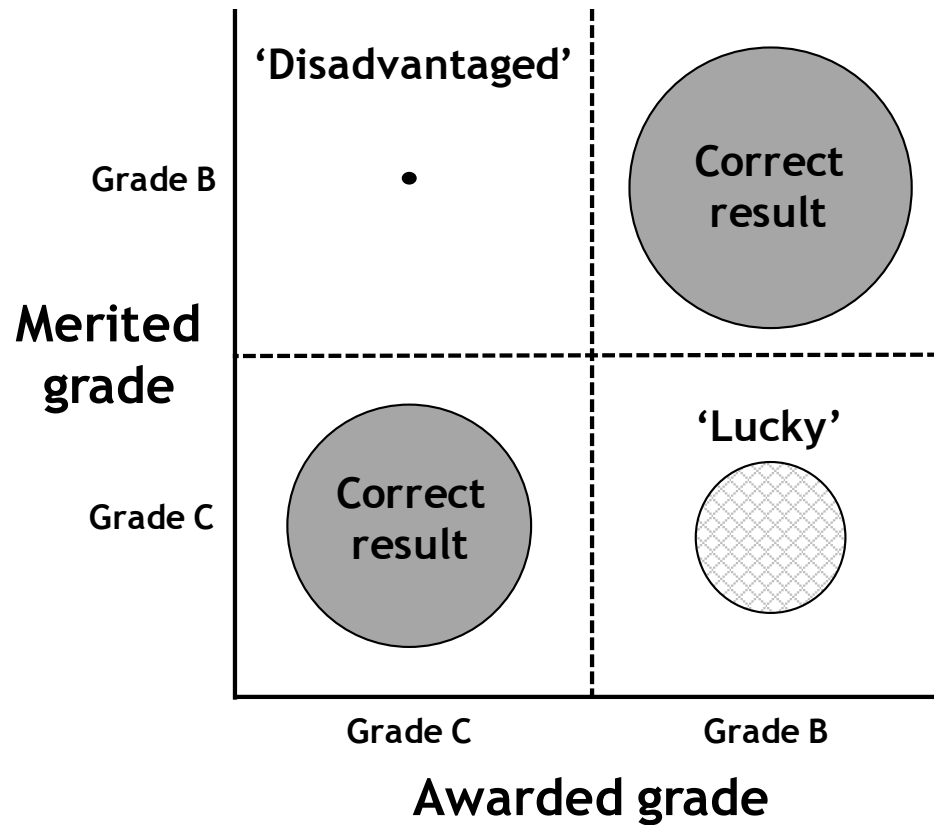
Grades based on $m - f$

If grades are based on $m - f$, grade misallocation has not been eliminated, but **the number of 'lucky' candidates is reduced to close to zero**, whilst the number of 'disadvantaged' candidates increases, as compared to grades based on m (see page 5).

This is very stringent, but might be appropriate for examinations associated with skills or qualifications which must comply with very high standards. Few of us, for example, would wish to undergo an operation performed by a 'lucky' brain surgeon; rather less dramatically, awarding grades based on $m - f$ might be sensible for vocational qualifications for, say, gas fitters or electricians - and perhaps also for the driving test.

Grades based on $m + f$

If grades are based on $m + f$, grade misallocation is, once again, not eliminated, but - as shown by comparison with page 5 - **the number of 'disadvantaged' candidates is reduced to close to zero,** whilst the number of 'lucky' candidates is increased.

This is generous, 'giving the benefit of the doubt', and ensuring that no candidate is 'disadvantaged'. For examinations such as GCSE and A level, which can influence a candidate's life chances, perhaps grading according to $m + f$ merits serious consideration.

A matter of policy

$$m - f$$

$$m$$

$$m + f$$
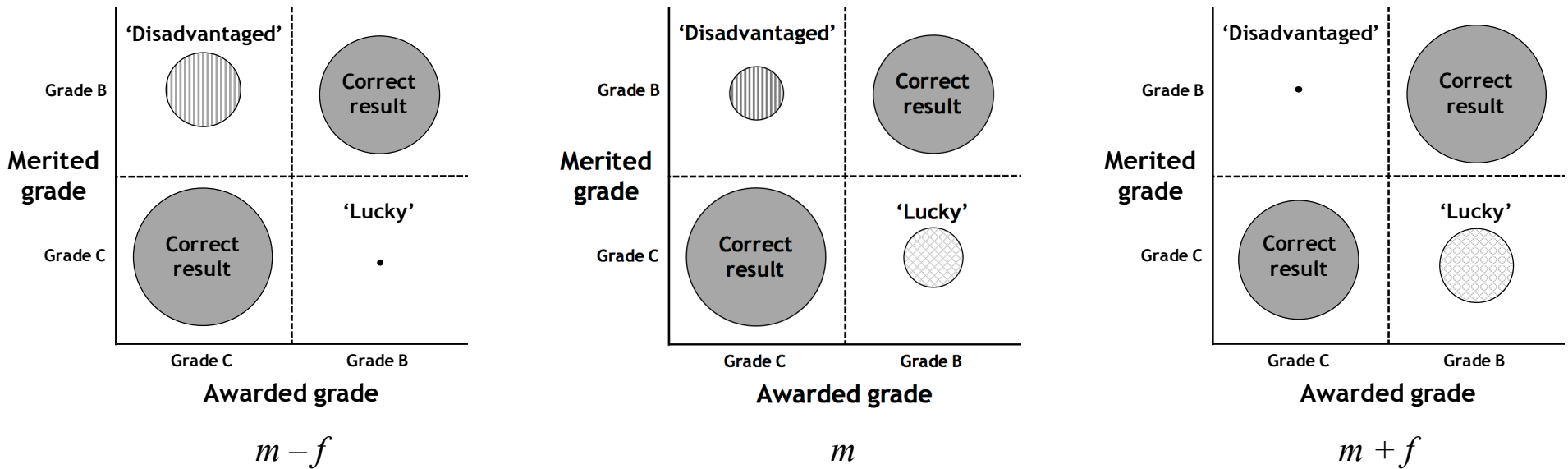
Grade allocation cannot be eliminated: it is attributable to the 'open' style of our examinations, and our recognition that equally qualified markers can legitimately award the same script somewhat different marks. But the nature of this misallocation - and, in particular, the relative sizes of the 'lucky' and 'disadvantaged' populations - can be controlled according to whether we award grades based on m, as at present, or based on either $m - f$ or $m + f$.

We therefore have a choice - a policy choice. Do we wish to perpetuate the current muddle, with grades based on m, and an unknown number of 'disadvantaged' and 'lucky' candidates? Do we wish to be stringent, grading on the basis of $m - f$? Or do we wish to award grades based on $m + f$, giving all candidates, fairly, the 'benefit of the doubt', and ensuring that no candidates are denied potentially vital life chances as a result of being 'disadvantaged'?

This policy choice is important, and it must be debated widely, and openly.

- 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 (page 3), which can be represented as $m \pm f$, where, for any examination, f is a statistically valid 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 on $m - f$, or 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 'lucky' candidates is reduced to close to zero, whilst the number of 'disadvantaged' candidates is increased.

- 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.

- We have a policy choice. Currently, grades are based on m:

Would it be more fair for GCSE and A level grades to be based on $m + f$?

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