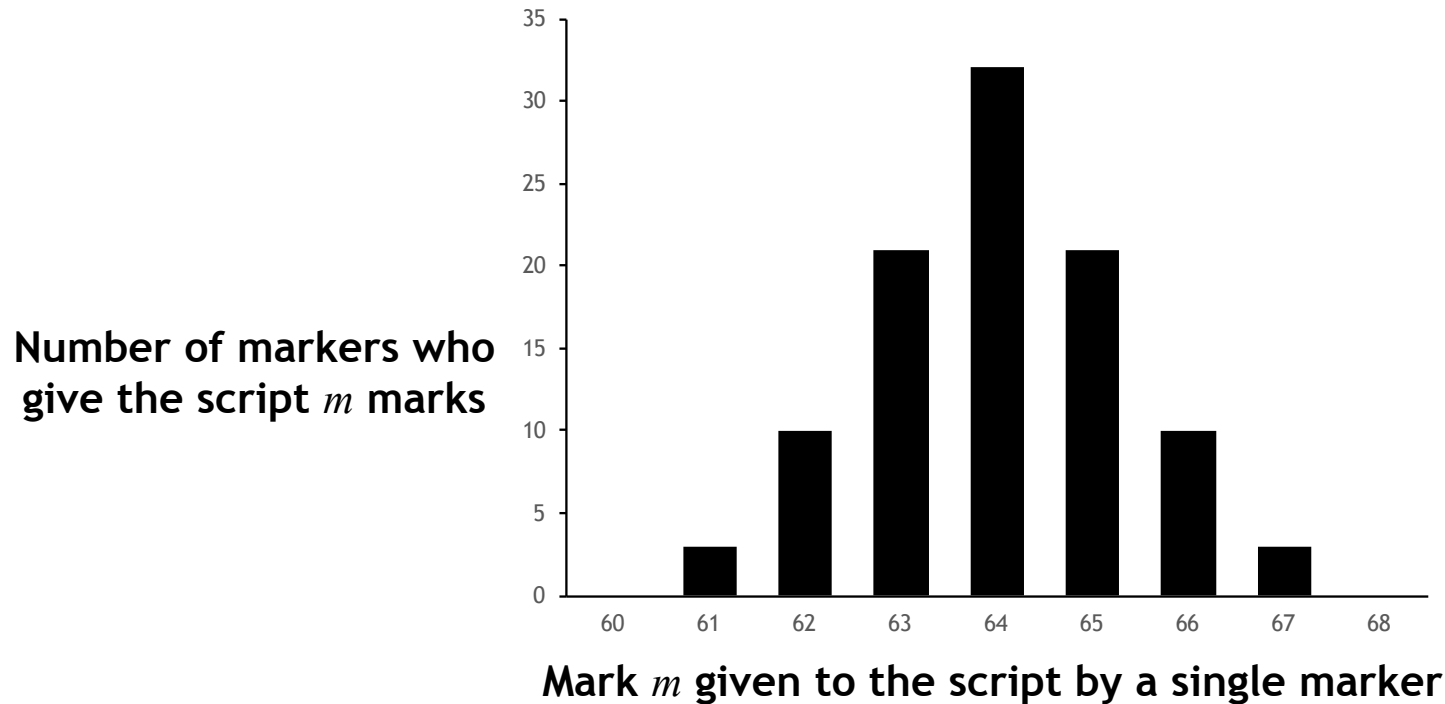# Identifying marking 'errors'

Dennis Sherwood, July 2017

The accompanying document *The great grading scandal* shows that a policy of grading examinations on the basis of a single mark m results, at every grade boundary in every subject, in the misallocation of grades. As a result, some 'lucky' candidates are awarded a grade higher than they merit, whilst others are awarded a grade lower than they merit and are therefore 'disadvantaged'. This unfairness is attributable to the failure to take the variability in marking into account when awarding grades.

The document *How to make grading fair* shows how this unfairness can be resolved by awarding grades based on the 'adjusted mark' $m + f$, where f is a measure of the variability in marking, as discussed in the document *How to measure f*.

This supplementary document explores the question "What happens if the originally given mark m was in error?"

How can a 'marking error' be identified?

Number of markers who give the script m marks

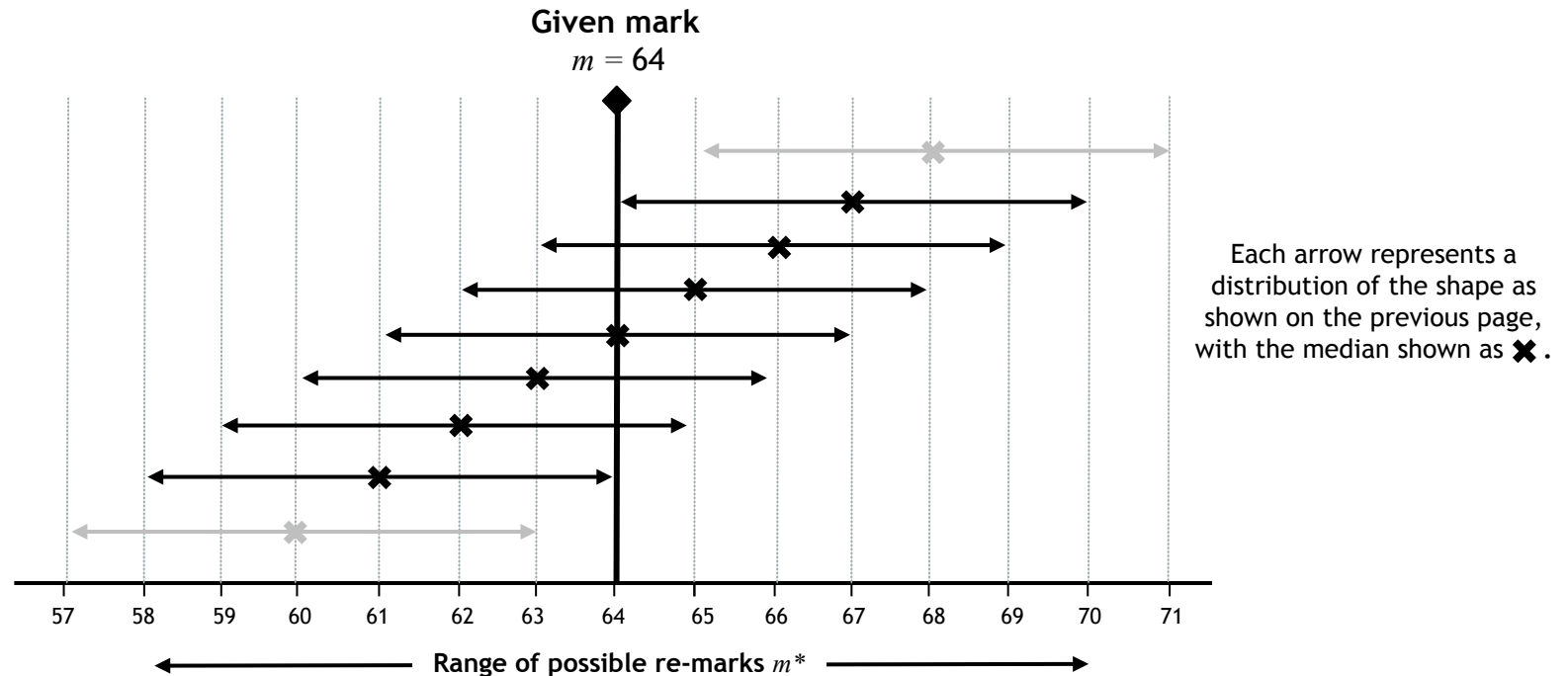**Mark m given to the script by a single marker**

Suppose that a submission is given a single mark $m = 64$ by a single marker. Suppose further that, perhaps as the result of an appeal, the same submission is given a re-mark $m* = 66$. Does this imply that the original mark $m = 64$ was a mistake?

Marking errors - as caused, for example, by the failure of a marker to follow the marking scheme, or a failure in internal quality control - can certainly occur, and it is possible that the original mark $m = 64$ was indeed a 'marking error'. Operationally, however, a re-mark $m* = 66$ is only very weak evidence that a mistake has been made. Why so?

As shown in this diagram, there is an unavoidable variability in fair marking. If 100 markers mark the same submission, all markers can comply with the marking scheme, and quality control can be of the highest standard, yet different markers can still, legitimately, give different marks. As discussed in the accompanying documents *The great grading scandal* and *How to make grading fair*, and as illustrated here, it is quite possible for the same submission to be marked $m = 64$ by one marker, and $m* = 66$ by another. Accordingly, the fact that the re-mark $m* = 66$ is different from (and higher than) the original mark $m = 64$ does not 'prove' that the original mark was 'wrong'.

The range of possible, valid, re-marks is wide

Given mark
$m = 64$

Each arrow represents a distribution of the shape as shown on the previous page, with the median shown as ✖.

57 58 59 60 61 62 63 64 65 66 67 68 69 70 71

Range of possible re-marks $m*$

A further consideration is that an originally given mark, say $m = 64$, can be a member of any one of a number of distributions, each of the shape shown on the previous page and each of which contains the mark $m = 64$ (see page 4 of the document *How to make grading fair*). If a script is validly re-marked, and if the original mark m was not in error, then the re-mark $m*$ must be a member of the same, specific, distribution. Operationally, however, there is no knowledge as to which of these possible distributions this particular distribution might be. Accordingly, if a submission originally, and validly, marked $m = 64$ is given a valid re-mark $m*$ on appeal, and if the examination can be described by a distribution of the shape shown on page 2, then it is possible for that re-mark $m*$ to be any number in the range from $m* = 58$ to $m* = 70$. This confirms the conclusion drawn on the previous page: a re-mark $m* = 66$ *does not imply* that the original mark $m = 64$ was in error.

But it might be: the problem is that one consequence of the inherent variability in marking is that it 'masks' marking errors that are within the range of that variability. Only relatively 'large' errors can be identified unambiguously by a re-mark: in the case illustrated here, only if the re-mark $m*$ is either equal to, or greater than, 71, or equal to, or less than, 57 is this reliable evidence that the original mark $m = 64$ is an error, and that the submission should be re-graded on the basis of $m* + f$ accordingly.

3

The statistics of re-marking
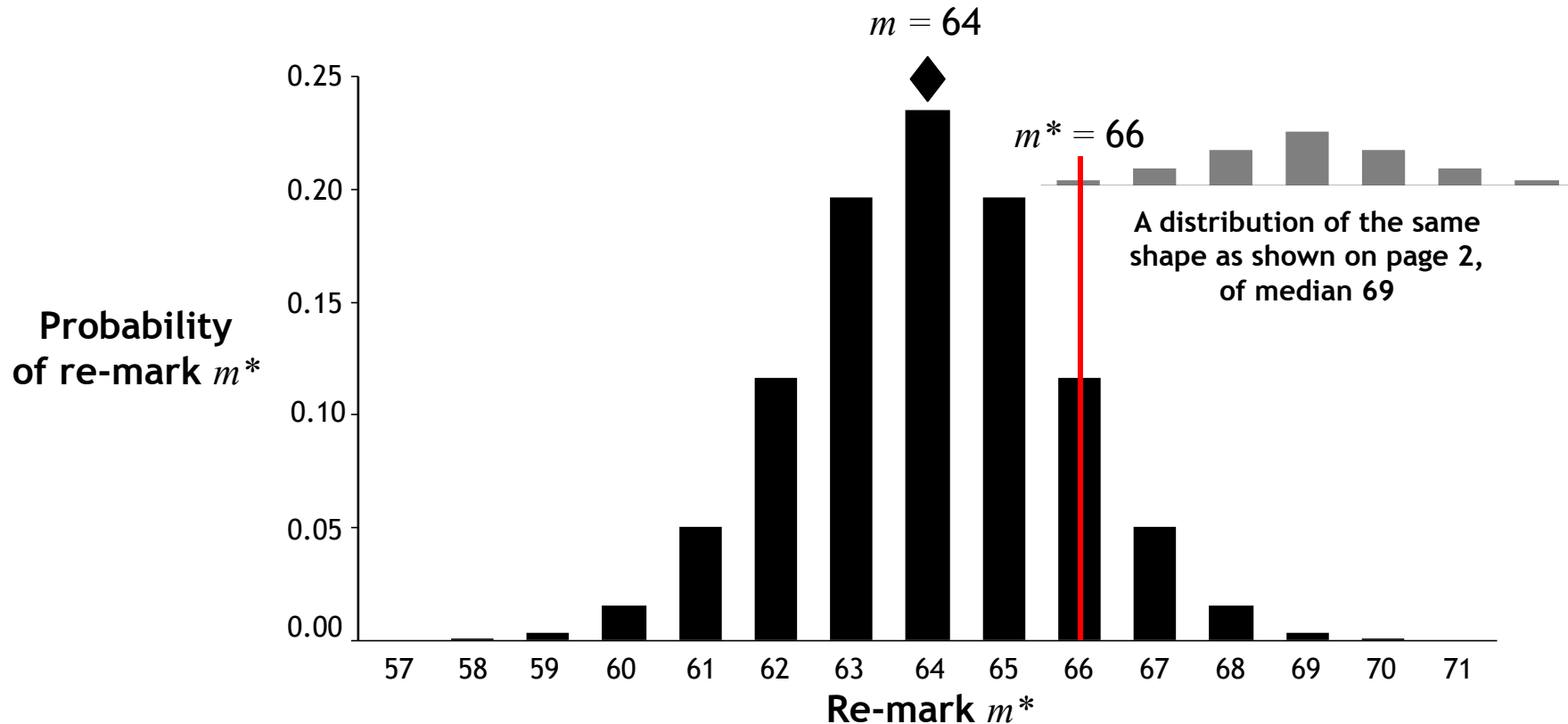
Consider a submission given a valid original mark $m = 64$. Although, as shown on the previous page, re-marks $m* = 58$ to $m* = 70$ are possible, they are not equally probable, as illustrated here (see also page 6 of *How to determine f*).

Suppose that a submission originally marked $m = 64$ is validly re-marked $m* = 66$. The re-mark $m* = 66$ is within the range from $m* = 58$ to $m* = 70$, and so a valid inference is that the original mark $m = 64$ is not a marking error. There is also a possibility, however, that the original mark $m = 64$ is an error, and that the submission is in fact a member of the distribution of median 69 - a distribution that does not overlap $m = 64$.

Which of these two alternatives is more likely? Are both $m = 64$ and $m* = 66$ valid marks, as illustrated by the distribution of re-marks $m*$ as shown in the main diagram on this page? Or is $m = 64$ an error, and $m* = 66$ a member of the distribution of median 69? Might this be resolved by a detailed statistical analysis of probabilities? Or might it be resolved by a further re-mark $m**$?   6

- **Marking errors** - as caused, for example, by the failure of a marker to comply with the marking scheme, or the failure of an internal quality control process - **can occur**. Accordingly, the originally given mark m might be an 'marking error'.

- On appeal, the submission is re-marked $m*$. **If the re-mark $m*$ is different from the original mark m, then it might be inferred that the original mark m was in error. Such an inference, however, is not necessarily correct**, for, given the inherent variability in marking, it is quite possible for two different markers to give the same submission different marks.

- This creates an ambiguity: if the re-mark $m*$ is different from the original mark m, then this could be a result of the variability in marking - but it could also be the case that the original mark m is in fact in error, but that the error has been 'masked' by the variability.

- This ambiguity is underpinned by two factors

  ➢ the magnitude of the inherent variability in marking, as measured by the parameter f (see the document *How to determine f*), and

  ➢ the difference $m*-m$ between the re-mark m and the original mark m.

- If the difference $m*-m$ is small, then it can be masked by the uncertainty f; for any given value of f, as the difference $m*-m$ increases, so does the probability that the original mark m is in error.

- This implies that **relatively small errors in marking cannot be reliably identified by re-marking**, and therefore that other methods - such as improved internal quality control procedures - should be used.

- Furthermore, **this 'masking' of small errors provides an additional argument in favour of awarding grades on the basis of $m+f$,** for this builds in a contingency.

Idea generation, evaluation and development

Making innovation happen

Silver Bullet

The Silver Bullet Machine Manufacturing Company Limited

Strategy development and scenario planning

Building ultimate competitive advantage

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

Building high-performing teams

Training and knowledge transfer

Conferences

Business and market modelling