The 'cliff-edge' problem – some thoughts

Dennis Sherwood, 8th July 2016

The current 'cliff-edge' problem

Grade boundaries are 'hard-edged': if the boundary between grade C and grade B is say, 59/60, then a mark of 59 is awarded a C, and a mark of 60 is awarded a B. Under the current appeals system, if a candidate with 59 marks appeals, it is possible that the challenge will result in an up-grade to a B. This has led to the charge that some schools are deliberately 'gaming' the current system, challenging all marks just below a grade boundary, regardless of the calibre of the student, on the off-chance that some candidates will be 'lucky' enough to be ungraded.

The 'cliff-edge' problem in the context of the $m + f$ idea

As described in detail elsewhere (see for example *The Great Grading Scandal*), the essence of the $m + f$ idea is that a mark of m is awarded a grade based on $m + f$, where f is a statistically valid measure of the variability of marking, attributable to the fact that different equally qualified markers might legitimately award (slightly) different marks to the same script. Although the spread of marks is, in general, small – say ± 2 marks – this small difference makes all the difference when 'fuzzy' marks (say, 59 ± 2) are mapped onto a hard-edged grade boundary, 59/60.

Under the $m + f$ idea, a mark of 59 is graded as 59 + 2 = 61, and so is awarded a B. Likewise, a mark of 58 is graded as 58 + 2 = 60, also a B, but a mark of 57 is graded as 57 + 2 = 59, a C.

Suppose a candidate does indeed receive 57 marks, and is awarded a C. Under the $m + f$ principle, had the candidate received 58 marks, the grade would have been a B. This is true. It therefore seems that we now have a situation in which all candidates with 57 marks might now raise a challenge, just as under the current rules, all candidates with 59 marks might raise a challenge. Apparently, nothing has changed. So why bother?

The paradox resolved

In fact, something very important has changed. Under the current system, a challenge to a mark of 59 is quite likely to result in a mark within the range 59 ± 2, and if the candidate is fortunate enough for the re-mark to be 60 or 61, an upgrade is indeed awarded – as is our current experience.

Suppose that, under the $m + f$ idea, a candidate with 57 marks (which was graded on 57 + 2 = 59) raises a challenge. The script is conscientiously remarked, with a high probability that the re-mark is within the range 57 ± 2 – say, 59. *This remark is still within the range* 57 ± 2, and so still merits a C. There is no upgrade.

Some further thoughts

It is possible that the candidate is still aggrieved, and continues to think that, had the script been awarded 58 or 59 first time around, then a B would have been awarded. This is true, but sincerely lucky. The evidence, after the challenge, is that *two marks* have been awarded in the range 57 ± 2, suggesting that this range is fair, and that an original mark of 58 or 59 would have represented a mark to the far right-hand (high) side of the fair distribution of which this candidate's is a member.

It is also possible that the original mark of 57 happens to be at the far left-hand (low) side of a fair distribution centred on, say 60 or 61. This could happen, but it will be statistically rare. Rare, yes; but still possible.

So the candidate raises a challenge, and the script is fairly re-marked. If the original mark was truly a low-end outlier (57 = 60 – 3 is rare, but possible; 57 = 61 – 4 even rarer), then there would be a high probability that the re-mark is 60, 61, 62, 63 or 64. Since all of these marks are beyond the original $m + f$ range 57 ± 2, if the script, is remarked as, say 61, then the script is re-graded as 61 + 2 = 63, resulting in an upgrade to B.

The $m + f$ idea therefore does allow for upgrades, but only for those scripts for which there is genuine evidence that the first-given mark is at the far low-end tail of an appropriate fair distribution – evidence based on the fact that the remark m' is higher than the upper bound $m + f$ of the original mark. If f is determined statistically soundly, then the probability of a re-grade is low.

One further point in relation to the example of the original mark of 57. It is possible that the re-mark is 59, on the low-end side of the fair distribution centred on 60 or 61. The remark of 59, however, is still within the original range 57 ± 2, and so the original grade C is confirmed. There is indeed an argument that this candidate has been unlucky – not once, but twice. In principle, this is possible, but it will be rare – just how rare can be determined by the study of existing data sets, and by the wise setting of f. In an absolutist world, this potential unfairness is to be lamented; in a pragmatic world, this is a consequence of the sensible application of statistics.

Most importantly, the $m + f$ idea is manifestly and demonstrably *much more fair* than the current system, and therefore is a major and practical step in the right direction, even if it does not achieve perfection.