Giving partial credit during a multiple-choice question assessment reappraisal does not make the assessment process fairer

L'attribution d'un crédit partiel lors de la réévaluation d'un questionnaire à choix multiple ne rend pas le processus d'évaluation plus équitable

Janeve Desy,[1] Adrian Harvey,[1] Kerri Martin,[1] Christopher Naugler,[1] Kevin McLaughlin[1]

[1]Office of Undergraduate Medical Education, Cumming School of Medicine, University of Calgary, Alberta, Canada
Correspondence to: Dr. Kevin McLaughlin, Office of Undergraduate Medical Education, Cumming School of Medicine, University of Calgary, Health Sciences Centre, 3330 Hospital Drive NW, Calgary, Alberta, Canada T2N 4N1; phone: (403) 220-4252; fax: (403)-270-2681; email: kmclaugh@ucalgary.ca

Having narrowly missed the minimum performance level (MPL) on their final clerkship multiple choice question (MCQ) exam, a student requests reappraisal of two questions where they felt their answer was correct or equally correct. They suggest that it would be "fairer" to give them at least partial credit on these questions, in which case they would pass the exam. A group of content experts reviewed the questions and felt that, although the student's answers were not the single best answers, they were reasonable answers/equally correct. So, would it actually be fairer to grant full or partial credit for these questions?

When Benjamin Wood developed the Type A MCQ examination format over 100 years ago, his reported motivation was both efficiency of scoring and fairness to students.[1] There are many reasons why Type A MCQ questions might be considered "unfair," but from the student perspective it is likely that this is dominated by how these questions are scored.

Choices for scoring Type A MCQ exams
The three most frequently used scoring methods are:

1. "Single answer" (SA) or "number correct"
Here there is full reward for one correct answer and neither reward nor punishment for all other choices. This offers simplicity in setting the MPL and scoring exams, and is designed to reward only *complete* knowledge. Of concern,

however, the lack of penalty for an incorrect answer may encourage guessing, which is problematic since rewarding a successful guess reduces reliability and validity of assessment.[2] There is also unease that dichotomizing the outcome of knowledge is an oversimplification and that grouping students with partial knowledge alongside students who lack knowledge is unfair to the former.

2. "Negative marking"
In this approach, a penalty for a wrong answer (typically a score of $-1/(n-1)$, where n = number of options) is incorporated into SA scoring to reduce the likelihood of guessing and improve reliability. While this approach might improve the psychometric properties of assessment, the worry here is that this type of scoring may disadvantage risk-averse student, including females who are typically more risk-averse during assessment than males.[3]

3. Elimination Testing (ET)
This approach is designed to reward partial knowledge.[4] Students consider each of the options and eliminate those they consider incorrect. Rewarding the elimination of incorrect options and penalizing the elimination of the correct option creates a score gradient from *full misinformation* (elimination of the correct answer only) to *partial misinformation* (elimination of the correct answer and some incorrect options), *absence of knowledge* (no options eliminated), *partial information* (elimination of

some incorrect option), and *full information* (elimination of all incorrect options).[4] Students typically express a preference for being rewarded for partial knowledge and, not surprisingly, when rewarding both complete and partial knowledge, students' scores are usually higher when using the ET approach to scoring type A questions.[4] A concern with ET is that higher scores leads to "grade inflation", which would be systematically unfair to students who completed their assessment with SA scoring.[5] So, one of the unresolved challenges with ET scoring is deciding whether to increase the MPL to mitigate against grade inflation and, if so, how best to do this.

Why fairness ≠ leniency

We have two concerns with granting full or partial credit for reasonable/equally correct answers in an assessment where the original intention was to use SA scoring. First, giving any degree of credit when a student's answer is not the single best option is in effect a post-hoc transition from SA scoring to an impromptu version of ET scoring for a small number of questions. As far as we are aware, no one has offered a validity argument for this modification or a description of how to revise the MPL after this adjustment. Second, granting full or partial credit can only produce a revised score that is the same or higher than the original score (*leniency bias* or *grade inflation*),[5] and intentionally introducing a second bias (*leniency*) to offset the first bias (*severity*) does not improve validity.[6]

For any assessment, validity is strengthened by judicious selection and consistent application of the scoring scheme, including during reappraisal. Our preferred approach to reappraisal of potentially biased questions is to simply remove these questions and then recalculate both the MPL and the student's score using the original SA method. Since the revised MPL can be the same, higher, or lower than on the original assessment, this avoids the addition of systematic bias, such as grade inflation.[5] This approach might not revise the student's score in their desired direction, but in the end we feel that maintaining or improving validity during assessment reappraisal is fairer to all concerned.

References

1. Tang SF CS. *Redesigning learning for greater social impact.* Taylor's 9th Teaching and Learning Conference.: Springer Nature; 2016.
2. Bereby-Meyer Y, Meyer J, Flascher OM. Prospect theory analysis of guessing in multiple choice tests. *J Behav Decis Mak* 2002(15):313-327. https://doi.org/10.1002/bdm.417
3. Kelly S, Dennick R. Evidence of gender bias in true-false-abstain medical examinations. *BMC Med Educ* 2009;9:32. https://doi.org/10.1186/1472-6920-9-32
4. Bond AE, Bodger O, Skibinski DO, et al. Negatively-marked MCQ assessments that reward partial knowledge do not introduce gender bias yet increase student performance and satisfaction and reduce anxiety. *PLoS One* 2013;8(2):e55956. https://doi.org/10.1371/journal.pone.0055956
5. Goldman L. The betrayal of the gatekeepers: grade inflation. *J Gen Ed* 1985;37:97-121.
6. Desy J, Bendiak G, McLaughlin K. Why we shouldn't grant partial credit when reappraising Type A MCQ questions. *Med Teach* 2023:1. https://doi.org/10.1080/0142159X.2023.2262108