

Commentary

Examination Script Marking System: The Issue of Scaling

Eluehike SU,¹ Awe OO²

1. Department of Radiology, Faculty of Clinical Sciences, Ambrose Alli University, Ekpoma
2. Department of Surgery, Faculty of Clinical Sciences, Ambrose Alli University, Ekpoma

Introduction

The performance of candidates who sit an examination - at any level - is usually determined by a defined scoring system. By such a predetermined system, marks are allotted by the examiners to each part of the test as well as each question.

This mark is the maximum attainable on each specific question. Therefore, following the assessment of a candidate's performance which is guided by a certain set of rules (marking scheme), the candidate is awarded a score ranging between the minimum and the maximum achievable; when all marks for a given question or section of the examination have been totaled, the final performance is expressed as a percentage of the maximum possible.

Marking Systems

There exist two systems of marking: the *open* and the *closed* system. Unfortunately, the authors cannot find any clear definitions of these terms in published texts; instead, as defined by AI, *open marking* involves keeping the marking criteria and marking scheme visible to students while they answer or while the teacher is marking, all designed to promote transparency and understanding of the assessment process. It is also used where candidates have the latitude to present their answers in their own words and style, as with essay questions.

In *closed marking*, on the other hand, questions have a predetermined set of correct or acceptable responses. It is often used for objective assessment such as multiple choice or short answer questions where a single right

answer is expected. Also, the marking scheme is not made accessible to the students at any time.

These terms are in common use in our centre, where they are applied – rather erroneously – to the manner of scoring or allocation of attainable mark per question. To expatiate, in the *open marking* system, all the attainable marks would be given to a candidate if they provide all the right answers to the examiner's satisfaction. For example, if the maximum score possible for a given question is 20, then a candidate can score 20/20 for a very excellent performance, or 0/20 for the worst performance possible.

On the other hand, closed marking means that the scores are restricted in such a manner that a new minimum and a new maximum are created whereby a candidate cannot be allotted the full score possible regardless of the brilliance of their performance, nor can they be scored “zero” no matter how bad their performance may be. There is a new “zero” (lower limit), and a new upper limit.

For instance, for a question to which is allocated 20 marks, the new lowest and highest attainable marks by any candidate could be 5 and 15 respectively (or 2.5 and 7.5 for a question whose full mark is 10).

This “contraction” or “compression” of marks has been named closed marking, but that is really a misnomer for *scaling*.

Scaling of marks is a known practice employed in the assessment of examination candidates at various levels. Examining boards of institutions employ it for various reasons; it has its advantages and disadvantages. As published by John C. McLachlan and Susan C. Whiten¹, some of the advantages include adjusting for examination difficulties, mitigating variability across cohorts,

Corresponding author Dr. Eluehike SU
Department of Radiology, Faculty of Clinical Sciences,
Ambrose Alli University, Ekpoma
Email - suehike@aauekpoma.edu.ng; ootchay@yahoo.com

adjusting for outliers, while the disadvantages include misrepresentation of performance, penalizing high performers/rewarding low performers, and certain ethical concerns. The Senate Standing Committee of the University of Wollongong, Australia² listed conditions which may necessitate the scaling of marks, viz: . if the average mark for the cohort is considerably higher or lower than performance demonstrated by the cohort in other assessments for that subject or compared to cohorts in previous years; if external forces unrelated to student performance have caused inappropriate variations between cohorts undertaking the same assessment; when marks are highly concentrated in a narrow band around the median; if the shape of the distribution is unusual; if a single assessment task proves to be problematic.

Notwithstanding, it has become the standard system for the marking of examination scripts in this centre (though not in all departments).

However, the focus of this commentary is not to extol the virtues of scaling or to expose its negative side. Since the relevant examining boards have decided in their wisdom to apply it, then that had better be done correctly and with consistency.

All along, there has not been a consensus as to how marks may be scaled. It is agreed among most – if not all - of the examiners that a script is to be marked “free-hand”, i.e., the appropriate raw marks awarded, then this mark is then adjusted by an appropriate function to obtain the final score. Verbal discussion with several colleagues and co-examiners has repeatedly revealed a frightening non-uniformity in the manner in which marks are scaled (still being called “closed marking”). The implications are serious!

What is important now, however, is to propose a workable formula which will apply to every “raw” or initial mark within any range of scores prior to the adjustment.

Let us therefore consider some hypothetical - albeit unstated – question to which a full or maximum score of 20 marks has been allotted. And by the practice of scaling, the highest score which may be awarded is 14, while the lowest is 6. The range 0 - 20 is now replaced by (“scaled to”) the range 6 - 14. (By this rule, of course, a score of 0 as a final take-home is no longer possible; 6, as it were, becomes the new 0).

By this system, 20 “steps” are reduced to 8, and any mark between 0 and 20 has its equivalent within the range 6 - 14.

This involves a conversion or “minification” factor of 0.4 (derived by making 20 a divisor of 8). This applies for any respective ranges of marks (*raw versus scaled*); the conversion factor is simply: the difference between the maximum and the minimum scaled marks, divided by the difference between the maximum and minimum raw marks.

It follows, then, that the score after scaling will be equal to the product of this conversion factor and the raw, free-hand score all added to the lowest possible score on the final scale.

This is given by the formula: $y = (c \cdot x) + \text{min}_s$.

where y = final score after scaling, c = the conversion factor, x = the initial score by free-hand assessment, and min_s is the minimum score on the scaled range of marks.

Hence, a candidate scoring 0-mark free hand – for the worst performance possible - is awarded:

$$y = (0.4 \times 0) + 6.$$

Therefore $y = 6$; and a candidate who scores 20 initially is awarded:

$$y = (0.4 \times 20) + 6.$$

Therefore $y = 14$.

Similarly, if a candidate scores 13 initially, his scaled score will be:

$$y = (0.4 \times 13) + 6. \text{ Therefore, } y = 11.2$$

Conclusion

This formula is reliable and reproducible and will apply effectively to all ranges of marks. Its use should put paid to all controversies attending the “closing”, nay, scaling of marks in our examinations.

More importantly still, it is time that we developed or at least agreed on a uniform marking and scoring system across all departments in the faculty or college in order to eliminate bias.

References

1. John C. McLachlan, Susan C. Whiten. “Marks, scores and grades: scaling and aggregating student assessment outcomes”. *Medical Education*, 34: 781 - 881.
2. Senate Standing Committee on examinations, University of Wollongong, Australia, 19th November, 2003.