

Why we must incorporate *primum non nocere* into assessment reappraisal

Pourquoi nous devons intégrer le principe de *primum non nocere* dans la réévaluation des évaluations

Janeve Desy,¹ Glenda Bendiak,¹ Kevin McLaughlin¹

¹Office of Undergraduate Medical Education, Cumming School of Medicine, University of Calgary, Alberta, Canada

Correspondence to: Correspondence. 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: 403 944 5883; fax: 403 944 5083;

email: kmclaugh@ucalgary.ca

Published ahead of issue: Jun 17, 2024; published: Aug 30, 2024. CMEJ 2024, 15(4) Available at <https://doi.org/10.36834/cmej.79224>

© 2024 Desy, Bendiak, McLaughlin; licensee Synergies Partners. This is an Open Journal Systems article distributed under the terms of the Creative Commons Attribution License. (<https://creativecommons.org/licenses/by-nc-nd/4.0>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited.

Allowing students to request a reappraisal to test the hypothesis of bias during assessment is intended to enhance assessment validity. Yet, academic assessment reappraisal inevitably comes with associated costs, including opportunity cost (the reappraisal process consumes time and effort that cannot then be spent elsewhere) and emotional costs, which can range from disappointment and frustration at the outcome of reappraisal to psychological trauma from the experience itself.¹

Individuals differ in their susceptibility to experiencing psychological trauma and, specific to the reappraisal process, the risk of psychological trauma is also likely to vary by assessment format. We feel that the multiple-choice question (MCQ) format carries the lowest risk since the student's performance (their choice of best answer) can be observed directly and compared to a clearly defined standard (the best answer choice of content experts)—so the reappraisal generally considers data challenging the choice of best answer rather than the rating of the student's performance. By contrast, reappraisal of formats that assess competence or clinical performance are more prone to inducing psychological trauma since 1) the reappraisal committee typically cannot observe a student's performance directly; 2) the expected level of performance is more subjective than a designated single best answer; and related to these, 3) the grounds for reappraisal are usually alleged rater bias. When reappraising an objective

structured clinical evaluation (OSCE) station, we at least have the benefit of focusing on the standardized rating of a specific task that was also performed by other students and rated by other raters—thus allowing us to screen for systematic rater bias.² By comparison, reappraisal of longitudinal work place assessment, such as in-training evaluation report (ITER), represents the reappraisal challenge with the highest risk for psychological trauma because of the potential for selective sampling from a large amount of data by a single rater who is concurrently performing other clinical and teaching duties. And, since there are no data on other students performing the same clinical tasks and/or other raters assessing these students, ITER reappraisal data often comprises conflicting opinions with or without accusations on the performance of the student and rater.

Having encountered situations where the student, faculty member, or both reported psychological trauma induced by the ITER reappraisal process, we feel that avoidance of psychological trauma should be a priority of reappraisal. We can promote this by providing support and guidance to both student and faculty on how to prepare their reappraisal documents and then limiting the reappraisal data to these documents rather than inviting verbal presentations with rebuttals. And, if the reappraisal committee feels that there is a high risk of the reappraisal process inducing psychological trauma then the original assessment decisions can be voided—in which case the

assessment is neither satisfactory nor unsatisfactory and the student must repeat the assessment. The goal of these recommendations is not to steer the reappraisal process and decision in favour of the student, since this would introduce a leniency bias that would then diminish assessment validity.³ Instead, we believe that principle of *primum non nocere* should be incorporated into the reappraisal process as a means of *enhancing* assessment validity. Contemporary validity frameworks, such as those proposed by Kane or Messick, require us to consider the *implications or consequences* in the validity argument for our assessment process, which also includes assessment reappraisal. Thus, irrespective of whether the assessment decision is changed or not, a reappraisal process that avoids the consequence of psychological trauma to either party strengthens our assessment validity argument.^{4,5}

Conflicts of Interest: None of the authors have a conflict of interest to declare.

Funding: No funding was received for this work.

Edited by: Marcel D'Eon (editor-in-chief)

References

1. Brown T, Berman S, McDaniel K, et al. Trauma-Informed Medical Education (TIME): advancing curricular content and educational context. *Acad Med* 2021;96(5):661-667. <https://doi.org/10.1097/ACM.0000000000003587>.
2. McManus IC, Thompson M, Mollon J. Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling. *BMC Med Educ* 2006;6:42. <https://doi.org/10.1186/1472-6920-6-42>.
3. Hoyt WT. Rater bias in psychological research: when is it a problem and what can we do about it? *Psychol Methods* 2000;5(1):64-86. <https://doi.org/10.1037/1082-989x.5.1.64>.
4. Kane MT. Validation. In: Brennan RL, ed. *Educational measurement*. 4th ed. Westport.: Praeger; 2006:17-64.
5. Messick S. Test validity and the ethics of assessment. *Amer Psychol*. 1980;35:1012-1027.