

Assessing the Quality of Peer Reviews in Public Health: Use of a Standardised Proforma among Online Course Participants

Abstract

This study evaluates the use of a newly developed proforma, designed to standardize peer reviews of public health research papers. The proforma was implemented as part of an online peer review training course for African health researchers. A total of 210 participants submitted completed proformas as their final assignment, reviewing a public health preprint. The proforma included structured multiple-choice questions and free-form responses assessing comprehension, critical appraisal, and editorial considerations. Participant responses were systematically compared to a model answer developed by course tutors. Overall, participants' evaluations were generally less critical of the paper under review; only 13% of the participants' recommendations aligned with the tutors' suggestion of "approve with major reservations". Scores were given to each response based on closeness to the model answer. Of a maximum score of 8 given to the four critical appraisal questions on the proforma, only 24 (11%) participants achieved a score of 6 or more. Free-form responses from these high scorers were further evaluated using a five-criteria rubric, with a mean score of 21 out of 25. Inter-rater reliability among assessors was high (ICC = .939, 95% CI [.876, .972], $p < .0001$). These findings support the proforma's feasibility as a scalable tool for training and evaluating peer review skills in public health, with potential applications in future educational programs.

Keywords: Peer review training, online peer review course, online learning, critical appraisal, standardized assessment, public health

Background

Peer review remains a cornerstone of scholarly communication and quality assurance in academic publishing (Hillard & Baber, 2021; Köhler et al., 2020). Rigorous peer review is essential to ensure that evidence informing policy and practice is methodologically sound, ethically conducted, and clearly communicated (Mavrogenis & Scarlat, 2023). The Coalition for Open Access Publishing of Public Health in Africa (COPPHA) is a collaborative initiative committed to advancing public health in Africa through the promotion of open science and equitable access to knowledge. COPPHA aims to democratize the research and publication landscape by fostering transparent, inclusive, and community-driven publication and peer review processes that empower African scholars and institutions. Further, it seeks to build on previous work which indicated both a need to assist journals and researchers to make their work more accessible through open access publications and an expressed need for online courses in open publishing including review skills (Agyei et al., 2023; Ruredzo et al., 2024). An initial initiative of the collaboration was the development and delivery of an online course to build capacity in peer review among public health professionals in Africa. The course and its outcomes are described in a companion paper (Arogundade et al., 2025).

Over the years, structured critical appraisal tools have been developed to assess the quality of research articles and the validity of the findings. These tools often integrate checklist-based approaches, mostly adapted from recognized critical appraisal frameworks, including those used in evidence-based medicine (Critical Appraisal Skills Programme, 2025; Darzins et al., 1992), public health evaluation (Heller et al., 2008), evaluations (Heller, 2024) and psychological research (Kerschbaumer et al., 2025). Although various standardized checklists exist to support the evaluation of scholarly manuscripts, the majority of scientific journals employ their own formal

procedures for eliciting reviewers' feedback on the scientific merit, rigor, and overall suitability of a submission for publication. PREReview, an open peer review platform, has developed a comprehensive Reviewer Guide to facilitate high-quality peer review. This guide was adapted from resources provided by the PLOS Peer Review Center, specifically drawing from the authoritative publication 'How to Write a Peer Review' (PLOS, 2025). These existing guides for peer review primarily focus on the structural aspects of the review, but neglect to provide instruction on the critical evaluation of research quality. To address this limitation, the Coalition for Open Access Publishing of Public Health in Africa (COPPHA) developed a novel proforma tool, based on a critical appraisal checklist. This instrument aims to enhance the peer review process, with particular relevance to open reviews of public health research. This study presents a case study on the implementation of this tool within an online training program, designed to augment capacity in peer review among health researchers in Africa.

Research questions

1. In an online peer review course, where students reviewed a public health research preprint using a proforma tool, to what extent do the quantitative scores derived from the proforma tool, differ between students and the course instructors' model answers?
2. What was the inter-rater reliability, as measured by the degree of agreement between independent observers, in the scores generated from a rubric designed to evaluate the validity of open-ended responses on the proforma, specifically assessing the quality of a research study?

Methods

The COPPHA online peer-review course served as the research activity through which this study was conducted. The course's primary objective was to train African researchers in all aspects of the peer review process, enabling them to conduct high-quality reviews. Key skills developed through the course included critically assessing research manuscripts, providing constructive and actionable feedback, and upholding ethical standards.

Delivered as a five-week open online course in October and November, 2024, the program blended asynchronous learning on the Peoples-Praxis platform with optional live virtual sessions using the Zoom platform (a copy of the course can be accessed here <https://courses.peoples-praxis.org/course/view.php?id=34>).

The course covered: 1) Fundamentals and challenges of peer review; 2) AI and the review process; 3) How to review a journal article; and 4) Performing a peer review assignment.

The course concluded with participants completing a peer review assignment, and those who successfully completed the course received a certificate of completion. A final assessment for the students to gain a certificate of completion required the completion and submission of a review of a public health preprint, using the proforma. Submission of the proforma was enough to gain a certificate of completion, and it also triggered access to a model answer prepared by three of the course tutors using the same proforma.

The proforma used in this study is a combination of three aspects of assessing a peer review of a research article, each comprising four questions. First, general questions to assess understanding of the paper. Second, questions based on critical appraisal of the background, methods, results and discussion/conclusions, with options of major, minor or no concerns. Third, questions related to potential editorial issues including language, transparency and recommendations. In addition to

multiple choice options, questions included a space for open-ended answers to amplify the reasons for making a choice of major or minor concerns about an aspect of the paper. The version of the proforma used in this study is provided as Supplementary File 1, and a later revised version is provided as Supplementary File 2.

Assessment of Peer-Review

The respondents' performance was assessed by reviewing their submitted responses on the proforma. The first stage was to compare the frequency of responses to each of the 12 questions with the agreed responses from the tutors as expressed on the model answer.

The second stage was restricted to the four questions that specifically focused on critical analysis of the literature (items analyzing the background, method, results and discussion/conclusion sections). Picking the same response as in the model answer gave a score of 2 (usually that there were major concerns), being close (usually that there were minor concerns) gave a score of 1, but either not answering or having no concerns gave a score of 0. There were four questions in this group, hence a maximum score of 8. A frequency distribution of scores was obtained (see Figure 1 for an outline of the study methods and sequence).

The third stage of the analysis was performed on the free-form answers provided by the participants in justification of their answers to the critical appraisal questions (questions 5, 6, 7 and 8). This stage of analysis was restricted to those who obtained a high aggregate score on the questions as above. In this case, the assessors had determined the high score to be an aggregate score of 6 or above. Students with a low score included many who did not complete the free-form questions, so there was limited value in further assessment of their responses. The assessors graded the free-form responses from 0 to 5 using a five-criteria rubric. These criteria were based on guides to a good peer review, as covered in the course and based largely on the PREReview and PloS guides to a good review – completeness, constructiveness, depth of critical analysis, tone and professionalism, and evidence and justification. A score of 25 represented the maximum score in this assessment stage. The assessments of the participants' proformas were independently done by the four research team members. An inter-rater reliability analysis was conducted to compare the level of agreement among the reviewers. Table 1 shows the way in which the scores were derived for the free-form critical appraisal questions.

Table 1. Components of the rubric to assess free-form responses to the critical appraisal questions

Assessment Component	Description
Completeness	The reviewer's written response to the review question is complete and coherent.
Constructive Feedback (Usefulness of Suggestions)	The reviewer's written response to the review question is constructive and useful. The comments provide highly actionable, specific, and insightful feedback, which offers concrete suggestions for improvement.
Depth of Analysis (Critical Evaluation)	The reviewer's written response demonstrates a deep and critical evaluation of the work.

Tone and Professionalism	The reviewer's written response to the review maintains a respectful, professional, and constructive tone throughout.
Evidence and Justification	The reviewer's written response to the review question provides well-supported critiques and recommendations, with clear references to specific examples from the work.

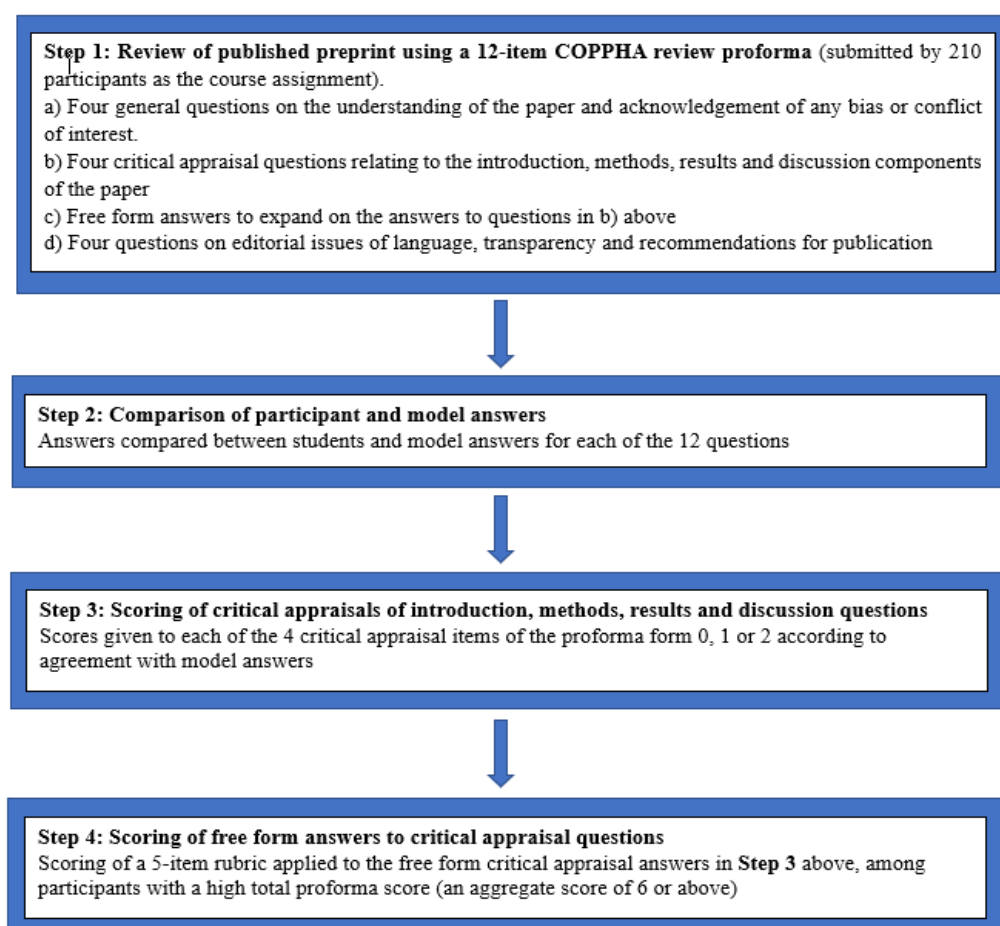


Figure 1: An outline of the study methods and sequence.

Results

Among the 651 participants who accessed the course, 210 (32%) completed the assignment. Among these, the proforma options were mostly completed, however, the responses provided by the participants were in general less critical of the paper than those expressed in the model answer (Table 2). Only 28 respondents (13%) agreed with the recommendation in the model answer to approve the paper but with major reservations, while 126 (60%) would require minor revisions and 39 (19%) no revisions.

Table 2: Frequency of answers by participants and completion status (filled, blank, or response checked) for each proforma question

Question	Answer	n = 210	%
1. What is the title of the manuscript you are reviewing?	Filled*	201	96
	Blank	9	4
2. Please specify if you have, or think you might have, any bias or conflict of interest in reviewing the manuscript	Filled*	200	95
	Blank	10	45
3. Please give a brief description of the topic, objective and main findings of the study. Briefly highlight the strength(s) of the study.	Filled*	197	94
	Blank	13	6
4. Does the abstract provide a concise summary of the study, including introduction, methods, results, and conclusions?	Yes	145	69
	No	9	4
	Partially*	45	21
	Blank	11	5
5. Do you have any concerns about the Introduction or Background to the study? These might include concerns about the literature review or the research question or hypothesis.	Major concerns	19	9
	Minor concerns*	81	39
	No concerns	97	46
	Blank	13	6
6. Do you have any concerns about the Methods section of the paper? These might include concerns about selection, measurement or confounding biases or the statistical analyses.	Major concerns*	33	16
	Minor concerns	87	41
	No concerns	77	37
	Blank	13	6
7. Do you have any concerns about the Results section of the paper? These might include concerns about selection, measurement or confounding biases or the statistical analyzes - or if each table, figure or graph is fully understandable as a stand-alone.	Major concerns*	22	10
	Minor concerns	81	39
	No concerns	95	45
	Blank	12	6

8. Do you have any concerns about the Discussion or Conclusions section of the paper? These might include ethical concerns, usefulness or generalisability of the findings and if the study limitations have been adequately identified.	Major concerns*	20	10
	Minor concerns	74	35
	No concerns	102	49
	Blank	14	7
9. Transparency and Reproducibility: Is the research process transparent, and are data or code available for others to reproduce your findings?	Yes	131	62
	No	11	5
	Partial*	52	25
	Not relevant	4	2
	Blank	12	6
10. Language: Do you have any concerns about the writing style or use of language?	Yes	40	19
	No*	155	74
	Blank	15	7
11. What is your recommendation?	Approve without revision	39	19
	Approve, require minor revision	126	60
	Approve, require major revision*	28	13
	Reject	5	2
	Blank	12	6
12. Final comments: please add any final comments that you think would be helpful to the authors.	Filled*	187	89
	Blank	23	11

*Option chosen by the model answer

Scores for the critical appraisal response questions

Using the scoring system as described above, with a maximum possible score of 8, the median score of the 210 participants who completed the assignment was 3.0, with an interquartile range of 0.0 – 5.0. The distribution of scores is shown in Figure 2. Of the 210 participants, 53 scored

zero, while twenty-four participants (11%) were considered as having a high score (aggregate score of 6 or above).

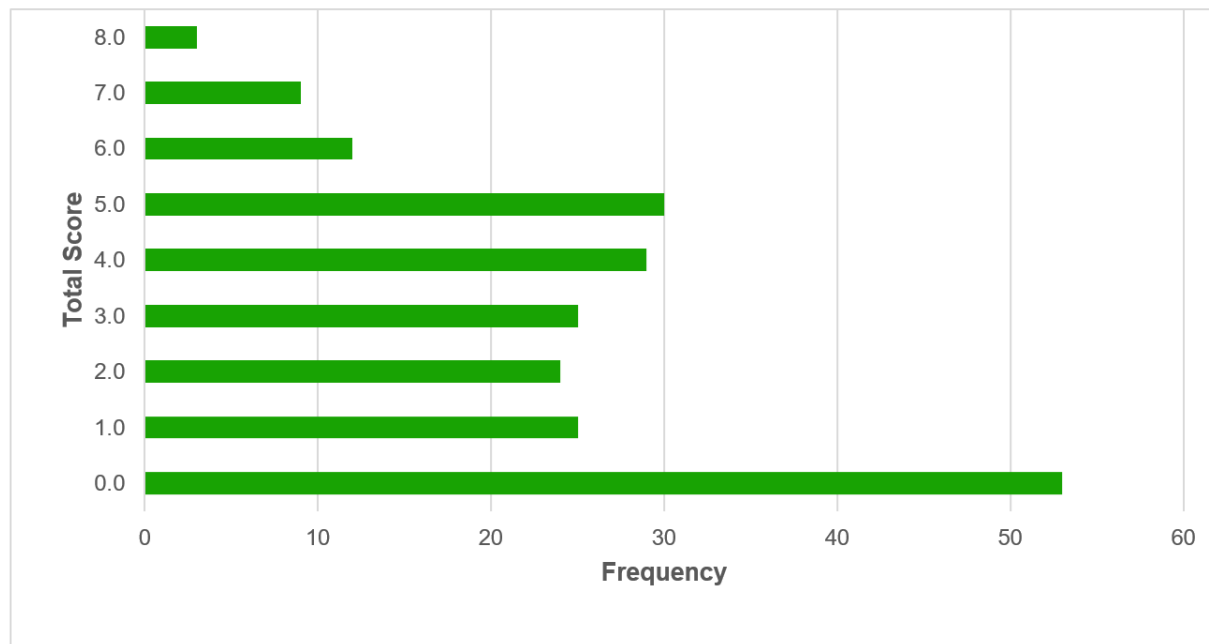


Figure 2: Distribution of scores for critical appraisal questions

Scores for the free-form critical appraisal questions

Based on the distribution shown in Figure 3, an exploration of the responses given by the 24 participants with high aggregate scores to the free-form questions was conducted. The mean score for the 24 participants was 21, with a range from 0 (only one participant) to the top score of 25. Among the four examiners, mean scores ranged from 20 to 23, and again the rankings were similar between them (Table 3).

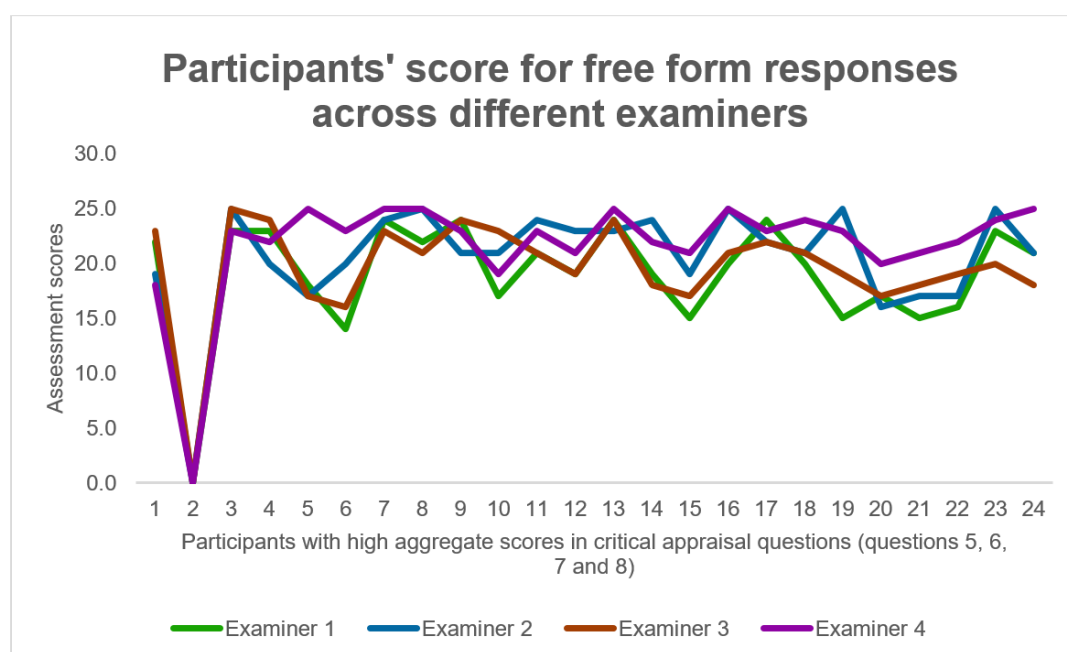


Figure 3: Participants' score of free-form responses across different examiners

Table 3: Level of agreement among the four examiners

Examiners	Mean $\pm$ SD	ICC	95% CI	p-value
1	19.0 $\pm$ 5.2	0.939	0.876 – 0.972	<0.0001
2	20.6 $\pm$ 5.3			
3	19.6 $\pm$ 4.9			
4	21.8 $\pm$ 5.0			

Inter-rater reliability analysis was conducted using intra-class correlation coefficient (ICC) which showed a near perfect agreement among the four examiners with an ICC=.939, 95% CI [.876, .972], $p<.0001$.

Discussion

A newly developed proforma to standardize peer reviews of research papers in the field of public health was completed as an assignment in an online course. The responses were less critical of the paper under review than were the tutors on the course, with 28 participants (13%) agreeing with the recommendation in the model answer to approve the paper but with major reservations, while 60% would require minor revisions and 19% no revisions. When the free-form responses to the questions involving a critical appraisal of the detailed content of the paper were evaluated using a specially designed rubric, there was good agreement between the four observers. These findings suggest that proformas are useful, assessable and standardized tools for conducting peer reviews in public health, and holds potential value for future educational programmes.

The use of a proforma provides a structured and systematic approach for the review process, ensuring that key aspects are addressed and also offers each student an opportunity to provide comprehensive feedback (Shah et al., 2019). However, this process may have some potential drawbacks such as stifled in-depth analysis and superficial engagement by the students (Thomas & Oliver, 2017). Proformas by their nature provide a checklist of items which can lead reviewers to focus solely on these predefined criteria, potentially overlooking novel aspects, subtle nuances, or interrelationships within the research that might not fit neatly into the proforma's boxes. Although, the developed proforma attempted to remedy this challenge by providing free text boxes to enable the students engage more with the process, the findings revealed that majority of them simply "ticked the boxes" or provided a brief, formulaic answers without engaging deeply with the intricacies of the article. This resulted in less insightful feedback for the authors. Nevertheless, comparing the performance of the students relative to the model answer allowed for evaluation of the quality of their performance of the review process.

The free text boxes allowed the students to expand on their answers as well as provide the reasons behind their choice. The responses provided were independently scored by the examiners and an analysis of the scoring revealed an excellent agreement between the examiners which suggests that the scoring system employed in this study may be a reliable method for assessing students' performance in an article review process. Inter-rater agreement as used in this study have been

identified as a measure of reliability in research where a fineness of discrimination must be employed during scoring (McHugh, 2012).

The PRereview guide for evaluating peer reviews encompasses six distinct constructs: completeness, constructiveness, tone, professionalism, critical evaluation, and evidence-based justification of comments. While the COPPHA proforma incorporates these core dimensions, it further extends the assessment framework by integrating a comprehensive critical appraisal of the research study's background, methodology, results, and discussion/conclusions, thereby providing a more nuanced and multidimensional evaluation tool. The division into three definite areas - general questions to assess understanding of the paper, questions based on critical appraisal of the research itself, and finally questions relating to potential editorial issues – emphasizes the importance of each of these aspects as part of a useful peer review.

The development of a standardized rubric for evaluating free-form comments in peer reviews is anticipated to have broader applicability, extending beyond the present context. The rubric's constituent features – completeness, constructive feedback (assessed via usefulness of suggestions), depth of analysis (critical evaluation), tone and professionalism, and evidence-based justification – align with established guidelines from PRereview and PLOS. Notably, the scoring system derived from these features demonstrated high inter-rater reliability, as evidenced by strong between-observer examiner agreement. This finding suggests that the rubric may be adaptable for use in various peer review settings, enhancing the consistency and rigor of evaluation processes.

The evaluation of peer review reports is crucial for upholding the integrity, transparency, and quality of scientific communication and decision-making in public health, with implications extending to other research disciplines (Ferreira et al., 2016; Tennant et al., 2017). Peer reviews plays a critical role in validating research findings, shaping policy recommendations, and guiding interventions. However, inconsistencies in review quality, ranging from superficial feedback to biased or poorly structured evaluations, can undermine the credibility of the evidence base and impede effective responses. The development of a standardized tool for assessing the quality of peer review reports in public health has significant implications. Firstly, it provides a framework for objective evaluation of peer feedback, enhancing editorial decision-making and author revisions. Secondly, the tool can facilitate effective peer review training, serving as a unified resource for preprint reviewers. Thirdly, by increasing transparency and reproducibility, the tool supports open science principles, promoting a more robust and reliable scientific record.

Study limitations

The fact that students could gain a certificate of completion just by submitting the review proforma, with access to a model answer later rather than having a marked assignment may have influenced their level of engagement and impacted the depth and comprehensiveness of the information provided by the students.

Conclusions

This study demonstrates the effectiveness of a newly developed proforma in standardizing peer reviews of public health research papers. The results show that the proforma can be used to assess comprehension, critical appraisal, and editorial considerations among health researchers. While participants' evaluations tended to be more favourable towards the preprint, the proforma identified areas for improvement in critical appraisal skills. The high inter-rater reliability among assessors suggests that the proforma can be used reliably to evaluate peer review skills. Overall, this study supports the feasibility of the proforma as a scalable tool for training and evaluating peer review

skills in public health, with potential applications in future educational programs aimed at enhancing research quality and integrity.

Data Availability Statement:

The dataset obtained and used in this study has been deposited on Zenodo and is accessible at: <https://doi.org/10.5281/zenodo.15868466>

References

- Agyei, D. D., Anyiam, F. E., Ruredzo, P. I. M., Sangare, M., Warnasekara, J., & Heller, R. F. (2023). Open access publication of public health research in African journals. *Insights: The UKSG Journal*, 36, 1–8. <https://doi.org/10.1629/uksg.605>
- Arogundade, F. Q., Suanon, T. B. A., & Heller, R. F. (2025). Predicting adherence to an online peer review course: A logistic regression analysis [Preprint]. BAOBAB. <https://pidlink.ren.africa/ark:/50962/bb67854ds5dn09ds>
- Critical Appraisal Skills Programme. (2025). *CASP Tools & Checklists*. CASP. <https://casp-uk.net/casp-tools-checklists/>
- Darzins, P. J., Smith, B. J., & Heller, R. F. (1992). How to read a journal article. *The Medical Journal of Australia*, 157(6), 389–394.
- Ferreira, C., Bastille-Rousseau, G., Bennett, A. M., Ellington, E. H., Terwissen, C., Austin, C., Borlestean, A., Boudreau, M. R., Chan, K., Forsythe, A., Hossie, T. J., Landolt, K., Longhi, J., Otis, J., Peers, M. J. L., Rae, J., Seguin, J., Watt, C., Wehtje, M., & Murray, D. L. (2016). The evolution of peer review as a basis for scientific publication: directional selection towards a robust discipline? *Biological Reviews*, 91(3), 597–610. <https://doi.org/10.1111/brv.12185>
- Heller, R. F. (2024, October 23). *A critical appraisal tool for evaluations: learning from Evidence Based Medicine*. <https://doi.org/10.31219/osf.io/s9zej>
- Heller, R. F., Verma, A., Gemmell, I., Harrison, R., Hart, J., & Edwards, R. (2008). Critical appraisal for public health: a new checklist. *Public Health*, 122(1), 92–98. <https://doi.org/10.1016/j.puhe.2007.04.012>
- Hillard, T., & Baber, R. (2021). Peer review: the cornerstone of scientific publishing integrity. In *Climacteric : the journal of the International Menopause Society* (Vol. 24, Issue 2, pp. 107–108). <https://doi.org/10.1080/13697137.2021.1882140>
- Kerschbaumer, S., Voracek, M., Aczél, B., Anderson, S. F., Booth, B. M., Buchanan, E. M., Carlsson, R., Heck, D. W., Hiekkaranta, A. P., Hoekstra, R., Karch, J. D., Lafit, G., Lin, Z., Liu, S., MacKinnon, D. P., McGorray, E. L., Moreau, D., Papadatou-Pastou, M., Paterson, H., ... Tran, U. S. (2025). VALID: A Checklist-Based Approach for Improving Validity in Psychological Research. *Advances in Methods and Practices in Psychological Science*, 8(1). <https://doi.org/10.1177/25152459241306432>
- Köhler, T., González-Morales, M. G., Banks, G. C., O’Boyle, E. H., Allen, J. A., Sinha, R., Woo, S. E., & Gulick, L. M. V. (2020). Supporting robust, rigorous, and reliable reviewing as the cornerstone of our profession: Introducing a competency framework for peer review.

Industrial and Organizational Psychology, 13(1), 1–27. <https://doi.org/DOI:10.1017/iop.2019.121>

Mavrogenis, A. F., & Scarlat, M. M. (2023). Quality peer review is mandatory for scientific journals: ethical constraints, computers, and progress of communication with the reviewers of International Orthopaedics. *International Orthopaedics*, 47(3), 605–609. <https://doi.org/10.1007/s00264-023-05715-y>

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276–282.

Okorie, C. G., Agyei, D. D., Izuchukwu, O., Qudus Arogundade, F., & Heller, R. F. (2025). Assessing the quality of peer reviews in Public Health: use of a standardised proforma among online course participants - dataset [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.15868466>

People-Praxis. (2024). *Peer reviewing – theory and practice*. <https://courses.peoples-praxis.org/course/view.php?id=34>

PLOS. (2025). *How to Write a Peer Review*. <https://plos.org/resource/how-to-write-a-peer-review/>

Ruredzo, P. I. M., Agyei, D. D., Sangare, M., & Heller, R. F. (2024). Open publishing of public health research in Africa: an exploratory investigation of the barriers and solutions. *Insights: The UKSG Journal*, 37. <https://doi.org/10.1629/uksg.635>

Shah, F. A., Ali, M. A., Nazar, Z., & Rasheed, H. U. (2019). Pattern of peer review proforma of medical journals of Pakistan. *Pakistan Journal of Medical Sciences*, 35(4), 1013–1017. <https://doi.org/10.12669/pjms.35.4.713>

Tennant, J. P., Dugan, J. M., Graziotin, D., Jacques, D. C., Waldner, F., Mietchen, D., Elkhatib, Y., B. Collister, L., Pikas, C. K., Crick, T., Masuzzo, P., Caravaggi, A., Berg, D. R., Niemeyer, K. E., Ross-Hellauer, T., Mannheimer, S., Rigling, L., Katz, D. S., Greshake Tzovaras, B., ... Colomb, J. (2017). A multi-disciplinary perspective on emergent and future innovations in peer review. *F1000Research*, 6, 1151. <https://doi.org/10.12688/f1000research.12037.3>

Thomas, L., & Oliver, E. (2017). Application of feedback principles to marking proformas increases student efficacy, perceived utility of feedback and likelihood of use. *Sport & Exercise Psychology Review*, 13, 39–47. <https://doi.org/10.53841/bpssepr.2017.13.2.39>

