

Structured peer review: pilot results from 23 Elsevier journals

Mario Malički¹, Bahar Mehmani²

¹Stanford University, USA

²Elsevier

ABSTRACT

Background and objectives: Reviewers rarely comment on the same aspects of a manuscript, making it difficult to properly assess manuscripts' quality and the quality of the peer review process itself. With regards to reviewers' recommendations, a 2010 meta-analysis found a very low inter-reviewer agreement of 0.34(1), and the Elsevier data covering 7,220,243 manuscripts from 2019 to 2021 across 2,416 journals, found 30% absolute reviewer recommendation agreement after the first review round.(2) This study aimed to evaluate a pilot program of structured peer review by: 1) exploring if and how reviewers answered structured peer review questions, 2) analysing their agreement, 3) comparing that agreement to agreement rate before implementation of structured peer review, and 4) further enhancing the piloted set of structured peer review questions.

Design: In August 2022, we introduced structured peer review, consisting of nine questions, in 220 Elsevier journals.. For the pilot analysis we aimed for 10% of this sample. We applied a random selection of journals across all fields and IF quartiles, and then selected research manuscripts that received two reviewer reports in the first two months of the pilot, leaving us with 107 manuscripts belonging to 23 journals. We did not have access to further review rounds, or final editors' recommendations for these manuscripts. Review reports were qualitatively analysed, with (partial) agreement defined as reviewers answering a question with the same answer (e.g., yes, no, NA, etc.) or a similar answer (i.e., one reviewer answering yes, the other - yes, but I would suggest improving...). Eight questions had open ended fields, while the ninth question (on language editing) had only a yes/no option. After the nine questions, reviewers could leave Comments-to-Author, and Comments-to-Editor. All answers (for questions 1 to 8 and Comments-to-Author) were independently coded by MM and BM with an inter-rater agreement of 94%, who then met on regular intervals to reach a consensus that was used for results reporting.

Results: Almost all reviewers (n=196, 92%) provided answers to all questions, with 12 (6%) skipping one question, and 6 (3%) skipping two questions. Overall the length of reviewers' answers to the eight questions (9th question was a yes/no) was 164 words (IQR 73 to 357), with the longest answer (Md 27 words, IQR 11 to 68) provided for question 2 (reporting methods with sufficient details for replicability or reproducibility). Reviewers had the highest (partial) agreement (72%) for assessing the flow and structure of the manuscript, and lowest (53%) for assessing if the interpretation of results is supported by data, and for assessing if statistical analyses were appropriate and reported in sufficient detail (also 53%). Two thirds of reviewers (n=145, 68%) filled out the Comments-to-Author section, which resembled standard peer review reports compiled during the review process and then copied to the field. Those Comments-to-Author sections contained on average 4 out

of 9 topics (SD 2) covered by the structured questions. Absolute agreement regarding final recommendations (exact match of recommendation choice) was 41%, which was higher than what those journals had in the period of 2019 to 2021 (31% agreement, $P=0.0275$).

Conclusions: Our preliminary results indicate that the adoption of structured peer review leads to reviewers covering more topics than they usually do in their reports. Individual question analysis indicated the highest disagreement regarding interpretation of results and conducting and reporting of statistical analyses. While structured peer review did lead to improvement in reviewer final recommendation agreements, this was not a randomised trial, and further studies should be done to corroborate this. Further research is also needed to determine if structured peer review leads to greater knowledge transfer or improvement of the final version of manuscripts.

KEYWORDS

peer review; inter-rater agreement; scholarly publishing

REFERENCES

1. Bornmann, L., Mutz, R., & Daniel, H. (2010). A Reliability-Generalization Study of Journal Peer Reviews: A Multilevel Meta-Analysis of Inter-Rater Reliability and its Determinants. *PLOS ONE*, 5(12), e14331. <https://doi.org/10.1371/journal.pone.0014331>
2. Petchiappan, R., James, K., Plume, A., Tsakonas, E., Marušić, A., Malicki, M., Grimaldo, F., & Mehmani, B. (2022). Analysing Elsevier Journal Metadata with a New Specialized Workbench inside ICSR Lab. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4211833>