Structured Peer Review: Pilot Results from 23 Elsevier Journals

Mario Malički (Stanford), Bahar Mehmani (Elsevier)

Editor-in-Chief of Research Integrity and Peer Review
https://researchintegrityjournal.biomedcentral.com/
Member of the EASE Peer Review Committee

History of Peer Review

1665. Journal des Sçavans (Paris)

1665. Philosophical Transactions (London)

1860. Slovinski prvenci o naravi i zdravlju (Croatia, Vienna)

1869. Nature (formal peer review 1967)

1879. Index Medicus

1890. Science

1994. World Wide Web

1997. PubMed (PubMed Central 2000)

2000. Many transition to online publishing

2003. PLOS Biology

History of Peer Review cont.

- Fewer submissions
- Editors were or societies – were the "peers" or "experts" or societies
- 1950s – 1970s most journals started using external peer review

On average **2 external** reviewers per article (invited by the editor)

And reviewers spend on average
3 to 8 hours for review

Overall acceptance rate is 35% to 40%

A Reliability-Generalization Study of Journal Peer Reviews: A Multilevel Meta-Analysis of Inter-Rater Reliability and Its Determinants (2010)

- The level of IRR (mean ICC/r^2=.34, mean Cohen's Kappa=.17) was low

- no indication of low validity or low legitimacy of the assessments… might indicate that the panel is highly competent because it represents a wide sample of the various views on what is good and valuable research

- Very few studies have investigated reviewer agreement with the purpose of identifying the actual reasons behind reviewer disagreement

7,220,243 manuscripts
2019 to 2021 across 2,416 journals
30% absolute agreement

Peer Review Workbench

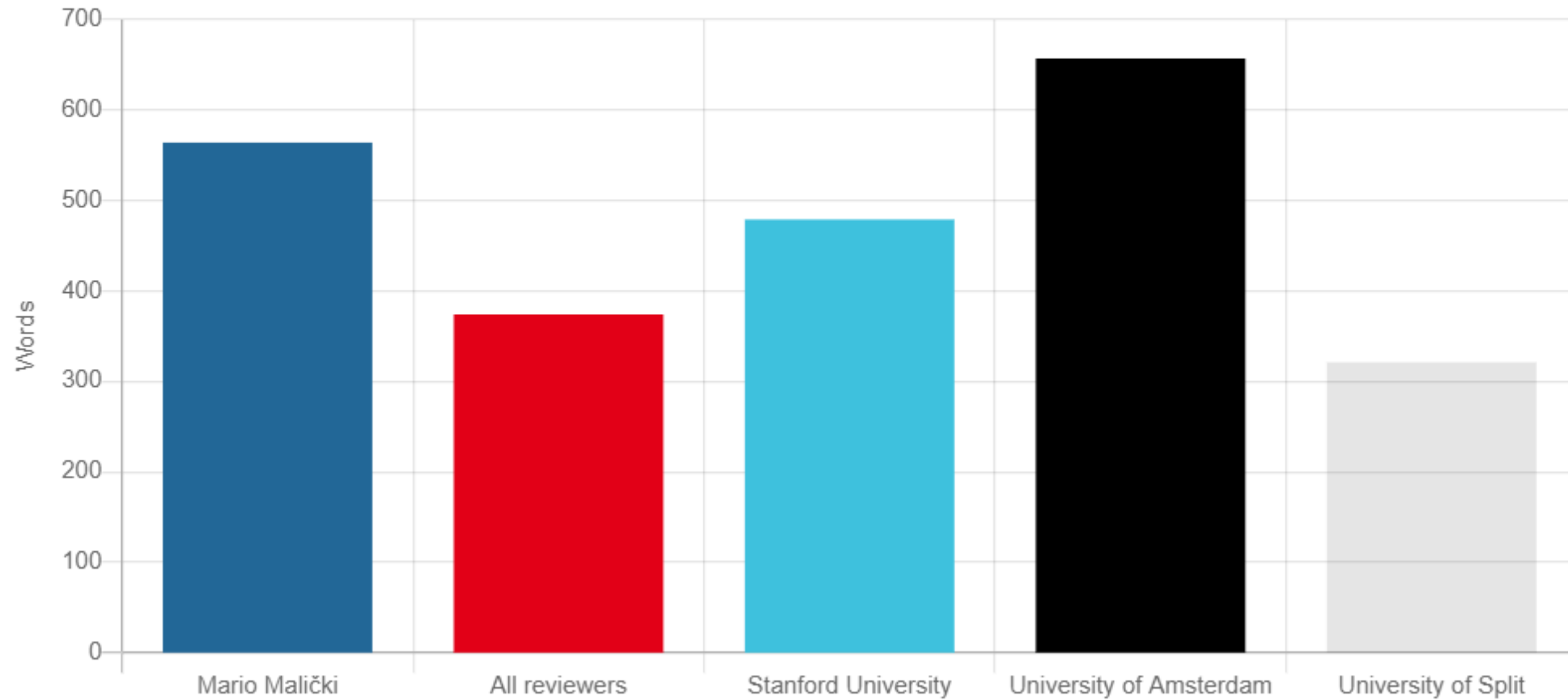New dataset offering researchers unique insights into journal peer review process

https://www.elsevier.com/icsr/icsrlab/how_to_apply

Chaired by Dr. Bahar Mehmani

Average review length

The average number of words per review (for which we have content), compared to the average of All fields reviewers and the average of reviewers at affiliated institutions.

Ratio 0.3 to 1

Methods

Structured peer review consisting of 9 questions was piloted in August 2022 in 220 Elsevier journals

Comments to Authors + Editor

10% random sample across all IF quartiles

Questions

Q1. Are the objectives and the rationale of the study clearly stated?

Q2. If applicable, is the application/theory/method/study reported in sufficient detail to allow for its replicability and/or reproducibility?

Q3. If applicable, are statistical analyses, controls, sampling mechanism, and statistical reporting (e.g., P-values, CIs, effect sizes) appropriate and well described?

Q4. Could the manuscript benefit from additional tables or figures, or from improving or removing (some of the) existing ones?

structured peer review

Questions

Q5. If applicable, are the interpretation of results and study conclusions supported by the data?

Q6. Have the authors clearly emphasized the strengths of their study/theory/methods/argument?

Q7. Have the authors clearly stated the limitations of their study/methods?

Q8. Does the manuscript structure, flow or writing need improving (e.g., the addition of subheadings, shortening of text, reorganization of sections, or moving details from one section to another)?

Q9. Could the manuscript benefit from language editing?

structured peer review

Objectives

1) explore if and how reviewers answered structured peer review questions,

2) analyze inter-rater agreement,

3) compare that agreement to agreement rate before implementation of structured peer review, and

4) further enhance the piloted set of structured peer review questions

1) If and how reviewers answered the q.

- 196 (92%) answered all questions
- 12 (6%) skipped one question
- 6 (3%) skipped two questions

- 15 (7%) of reviewers directed to attachments or answered the question and then directed to more details in the attachments
- 81 (38%) directed (or answered and directed to more details) to either the *Comments-to-Author* section or to their answers to other questions (mostly for single question on limitations or strengths.

1) If and how reviewers answered the q.

- 145 (68%) filled out the ***Comments-to-Author*** section

- 105 (72%) out of 145 (49% out of total sample of 214) provided answers that we classified as resembling full peer review reports (i.e. prepared review reports copied in the *Comments-to-Author* section)

- 74 (35%) of reviewers left *Comments-to-Editor (not analysed)*

Comments-to-Author

- Md word 323 for all, 482 full review reports

- **4 (SD of 2)** out of 9 questions - methods, results, least on limitations

2 and 3) Inter-rater agreement

- 41% absolute - 60% recoded

- journal before for all manuscripts (31%, P=0.0275)
- only manuscripts with 2 reviewers (36%, P=0.2891)

Q

- 72% was found for assessing the flow and structure of the manuscript
- 53% interpretation of results are supported by data, and for assessing if statistical analyses were appropriate and reported in sufficient detail

Conclusions

- Adoption leads to reviewers covering **more topics** than they usually do in their reports

- Individual question analysis indicated highest disagreement regarding interpretation of **results and conducting and reporting of statistical analyses**

- Further research is also needed to determine if structured peer review leads to greater knowledge transfer or **improvement** of final version of manuscripts and to determine reasons for disagreement

Thank you

mario.malicki@mefst.hr
mmalicki@stanford.edu