Reading Peer Review – What a dataset of peer review reports can teach us about changing research culture

Eve, Martin, Daniel Paul O'Donnell, Cameron Neylon, Samuel Moore, Robert Gadie, Victoria Odeniyi, and Shahina Parvin. 2021. "Reading Peer Review – What a Dataset of Peer Review Reports Can Teach Us about Changing Research Culture." *Impact of Social Sciences* (blog). March 31, 2021. https://blogs.lse.ac.uk/impactofsocialsciences/2021/03/31/reading-peer-review-what-a-dataset-of-peer-review-reports-can-teach-us-about-changing-research-culture/.

Most scholars are familiar with peer review; the sometimes brutal system by which academic work is judged. The COVID-19 pandemic has even mainstreamed this process, with media outlets commenting on unreviewed preprints with warnings that work has "not yet undergone peer review". Yet few venues, over the past two decades, have done more to challenge the way that peer review works than the Public Library of Science's PLOS ONE title.

PLOS ONE initiated a revolution in peer-review that was predicated on the principle of 'technical soundness'. The idea being that papers are judged according to whether or not their methods and procedures are thought to be solid, rather than whether their contents are judged to be important. A fairly radical idea from its inception, PLOS ONE was designed to foster Kuhnian "normal science", rather than shooting for high-profile 'discoveries'. Detractors worried PLOS ONE's review process would let through a raft of useless, insignificant (non-)results. Proponents believed it could usher in an era of better science.

As this change represented a fundamental shift in a core research practice, we wanted to understand how peer reviewers behave at PLOS ONE, and were generously given access to an anonymised set of peer review reports from the journal. What did we find? In purely quantitative terms, the mean length of reviews was 3,081 characters, while the longest report we found was just under 14,000 words(!)

Of more interest than this counting exercise was the qualitative taxonomy of peer review statements we developed. Roughly speaking, peer reviewers remark upon data, the field of knowledge, the expression of the paper under review, the methodology of the experiment, on what is missing from the manuscript, and in particular, expected aspects of the paper (e.g. the abstract). In addition to looking for categories, we also assessed sentiment; that is, how positively or negatively a reviewer felt about the aspect on which they were remarking. This provided a high-level picture of the structure of these reports.

In some ways, the peer review reports at PLOS ONE are highly standardised. 87% of the reports that we studied contained a summary statement within the first six sentences of their review. Such statements usually restate the aim of the paper under review and can serve a legitimation and verification function (the correct paper is being discussed and the reviewer demonstrates their

competence). People are extremely wedded to a particular structural style of peer review and this persists in PLOS ONE.

Yet the true area where PLOS ONE aimed to foster change was in terms of not appraising novelty but instead valuing elements such as reproducibility – the "technical soundness" criterion. We found, nonetheless, that reviewers for PLOS ONE frequently comment on the novelty of the papers that they are reviewing, but they do not usually remark upon reproducibility. Thus, although the review criteria of PLOS ONE are set at a strict boundary of "technical soundness" and we expected that reviewers would understand this, 77% of the reports that we examined commented either positively or negatively on (and, on some occasions, both positively and negatively upon different aspects of) the potential, novelty, and significance of the paper.

This is not a straightforward disregard for the rules. At least part of the PLOS ONE review process does request that reviewers comment on the significance, novelty, or importance of the work, albeit not as a criterion for acceptance, but to assist the editors in highlighting work – after publication – that reviewers believe merits especial attention. That said, most instances of commentary on novelty in our sample used this appraisal within the review as an evaluation criterion, rather than as guidance markers to editors.

More in line with the PLOS ONE's goals, however, we did encounter reports where reviewers chastised authors for making what the referees considered to be unnecessary or counterproductive claims of novelty*:

"the study is novel, but is that necessarily important?",

"the authors make much ado about the fact that the study is the first of its kind, but is it better to be first or to make an important contribution"

On occasion, reviewers praised authors' modesty, noting in one case that they "make reasonable and humble claims about the finding with a well-balanced explanation of the limitations of the study". Nonetheless, in general, despite PLOS's intention radically to modify peer review, reviewers' behaviours turn out to be far more resistant to change.

Given a rare chance to interrogate the normally hidden aspects of peer review, another question that we explored within the dataset was the extent to which the mythically harsh statements of Reviewer Two were a reality. Were people really as brutal in their peer review reports as are usually claimed?

The answer is "yes, sometimes". Statements often included quite negative and at times scathing and personal attacks directed towards the authors and, in some cases, the editors who agreed to send the article out for review:

"in summary, this manuscript as currently conceived and written should not be published in any reputable peer reviewed journal";

"these findings alone represent a minimal manuscript with effectively no major selling points to warrant publication";

"the manuscript reads very much like a portion of a thesis rather than a self-contained manuscript for public dissemination";

"the bulk of the writing is conjecture, speculation, unsupported theories, statements that lack data, etc."; "there is essentially no possibility of publication in the current format";

"these findings are not worth much of a publication, much less a publication in PLoS";

"the text is ridiculously unsupported, to the point where I would have been embarrassed to submit such a manuscript";

"I could complain more and more … but this manuscript is so poor that I am surprised it made it past editorial review";

And finally, "this is a poorly conceived paper with muddled logic which has no point".

We also wondered whether reviewers in PLOS ONE might adopt a strategy of "sugar coating" their reviews; a practice colloquially referred to as the "shit sandwich" where one begins with some good news, packs in the bad, and concludes with some words of encouragement. There was little evidence of this. However, among the papers that we assessed to have a high sentiment value, that is those that were strongly rejected, few received any positive feedback. Although the sample size here is small, at least some reviewers in PLOS ONE who deliver unfavourable verdicts are direct and unambiguous in their negative feedback. Reviewer two might be tough, but at least she tells you what she really thinks.

There is little real question as to whether PLOS ONE changed the scholarly publishing landscape. The concept of a "megajournal" reshaped the way publishing works and PLOS ONE was responsible for a significant and sustained increase in levels of open access. But its core goal was nothing less then reshaping peer review. It turns out, this was a much harder goal, and that process of change appears slower than the shift in the economics of the publishing landscape. The message here is not limited to the specifics of peer review. Even when you change an entire system around them, shifting the culture of academics remains a challenge. Our research highlights these "sticking points" in the process; internalised conventions that may be unhelpful for science but that remain resistant to change. It is only by understanding the way that people behave, at scale, behind the veil of peer-review anonymity that we can hope properly to modify reviewing conventions for the better.

* The quotations in this blog have been lightly paraphrased and anonymised in keeping with the conditions under which we were allowed to access the PLOS ONE data.