RESEARCH · RESEARCH INTEGRITY

Peer Review: How It Works and How It Fails

How peer review works at journals and funders, its main models, what the evidence shows about its reliability, bias and ability to catch errors, and the reforms and alternatives being tried.

ModQu research article · 1,995 words · 20 references (APA 7th edition)

Peer review is often described as the gold standard of quality control in research. Before a study is published in a scholarly journal, or before a grant is awarded, other experts in the field assess it and advise whether it should be accepted. The phrase "peer-reviewed" signals to readers that a piece of work has passed this test. Yet peer review is a human process with well-documented weaknesses, and the evidence on how well it works is surprisingly thin. This article explains how peer review works, what it is good at, where it fails and how it is changing.

A short history

Peer review is younger than many people assume. Some learned societies sought expert opinions on submissions from the eighteenth century onward, but many journals relied mainly on their editors' judgement well into the twentieth century. Csiszar (2016) traced how the practice developed unevenly and was contested from the start. Baldwin (2018) showed that the term "peer review" itself became common only in the Cold War United States, when scientists used it to argue that public funding of research should be overseen by experts rather than by politicians. Systematic external review of manuscripts became standard at many leading journals only in the second half of the twentieth century.

How journal peer review works

The typical process is shown below.



Figure 1. The typical journal peer review process. Diagram by ModQu.

An author submits a manuscript to a journal. An editor first checks whether it fits the journal's scope and meets basic standards; many submissions are rejected at this stage without external review. If the editor sends it for review, usually two or three experts are asked to assess the work. Reviewers comment on the importance of the question, the soundness of the methods, whether the conclusions follow from the results and how the presentation could be improved. They typically recommend acceptance, revision or rejection. The editor weighs the reports and makes the decision, which is often a request for revision followed by further rounds of review.

Grant peer review follows similar steps, but reviewers judge proposed rather than completed work, and panels often rank many applications against each other for limited funds.

Models of peer review

Journals use several models, which differ in who knows whose identity.

Models of peer review

They differ mainly in who knows whose identity

Model	Who knows what	Advantage / drawback
Single-anonymous	Reviewers know authors	Common / prestige bias possible
Double-anonymous	Neither knows the other	Less bias / identities often guessed
Open	Identities and often reports public	Accountable / some decline to review
Post-publication	Anyone can comment after release	Wide scrutiny / uneven coverage

Drawing on Ross-Hellauer (2017) and Tomkins et al. (2017) ModQu.com

Figure 2. Four models of peer review. Table by ModQu, drawing on Ross-Hellauer (2017) and Tomkins et al. (2017).

In **single-anonymous** review, reviewers know the authors' identities but authors do not know the reviewers'. In **double-anonymous** review, neither side knows the other's identity, although reviewers can often guess. In **open** review, identities are disclosed, and in some forms the reviews themselves are published alongside the article. Ross-Hellauer (2017) found that "open peer review" was used to mean many different things and identified several distinct traits, including open identities, open reports and open participation by the wider community.

What peer review is meant to do

Peer review serves several purposes at once: filtering out work that is unsound or unoriginal, improving manuscripts through expert feedback, helping editors choose among submissions, and certifying work for readers who cannot judge it themselves. Researchers value it. Mulligan et al. (2013) surveyed several thousand researchers internationally and found strong support for peer review and a widespread belief that it improves the quality of published papers, alongside concerns about delays, bias and the burden on reviewers.

Pressures on the system

The volume of research has grown much faster than the pool of willing reviewers. Editors report that it is increasingly difficult to find qualified people to review, and that a relatively small share of researchers does a large share of the reviewing. Reviewing is usually unpaid and receives little formal credit in hiring and promotion, so it competes with activities that do advance careers. Delays of months between submission and decision are common in some fields. These pressures do not change the principle of peer review, but they affect its quality: tired, rushed or reluctant reviewers are less likely to read closely, check analyses or engage with unfamiliar methods.

What the evidence shows

Despite its central role, peer review has been studied surprisingly little, and the evidence that exists is sobering. Jefferson et al. (2002) systematically reviewed studies of editorial peer review and concluded that there was little empirical evidence to support its use as a mechanism for ensuring quality. Rennie (2016), who organised a long-running series of international congresses on peer review, called for the process itself to be studied scientifically.

Reliability

If peer review measures quality, different reviewers of the same manuscript should broadly agree. They often do not. Cicchetti (1991) reviewed studies across disciplines and found low agreement among reviewers of both manuscripts and grant applications. Bornmann et al. (2010) pooled studies of journal peer review and found that agreement between reviewers was, on average, low. Kravitz et al. (2010) found that reviewers' recommendations at a general medical journal agreed only slightly more than would be expected by chance. Pier et al. (2018) asked many reviewers to assess the same set of grant applications to a major United States funder and found almost no agreement in their scores.

Peters and Ceci (1982) carried out one of the most famous tests. They resubmitted twelve articles that had already been published by prestigious psychology journals, changing the authors' names and institutions to fictitious, less prestigious ones. Most resubmissions were not recognised, and of those that went through review, nearly all were rejected, often for serious methodological flaws that had not prevented their original publication.

Detecting errors

Peer review is widely assumed to catch errors, but tests suggest it catches only some. Godlee et al. (1998) inserted deliberate errors into a manuscript and sent it to hundreds of reviewers; on average, reviewers identified only a small fraction of them. Schroter et al. (2008) repeated the exercise with several papers and found similar results, and also found that short training courses did little to improve error detection.

Bohannon (2013) sent a fabricated paper with obvious scientific flaws to hundreds of open access journals, and more than half accepted it. The exercise did not test traditional subscription journals and so does not show that open access journals are worse, but it exposed how weak review can be at some outlets.

Bias

Peer review can be biased in several ways. Mahoney (1977) sent reviewers versions of the same manuscript that differed only in whether the results supported or contradicted the reviewers' likely theoretical position, and found that reviewers rated the methods more favourably when the results agreed with their views. This **confirmation bias** works against unexpected or unwelcome findings.

Reviewers may also be influenced by authors' identities. Tomkins et al. (2017) compared single-anonymous and double-anonymous review at a computer science conference and found that reviewers who could see authors' names were more likely to favour papers by famous authors and by authors from top universities and companies. Lee et al. (2013), reviewing the literature, identified many potential sources of bias, including bias related to prestige, nationality, language, gender and the content of the work, while noting that evidence for some of them was mixed.

Effects of the review model

Experiments have tested whether changing the model improves quality. Godlee et al. (1998) found that asking reviewers to sign their reports, or concealing authors' names, did not significantly improve the detection of errors. Van Rooyen et al. (1999) found that revealing reviewers' identities to authors did not affect the quality of reviews or reviewers' recommendations, although it made some reviewers more likely to decline to review.

The case against and for

Smith (2006), a former editor of the *BMJ*, summarised the critics' view: peer review is slow, expensive, inconsistent, prone to bias, easily abused and poor at detecting errors and fraud, and there is little evidence that it achieves its aims. Yet he and most other critics stop short of calling for its abolition, because no clearly better alternative exists, and because review often does improve manuscripts in ways that are hard to measure.

A balanced view is that peer review is a useful but limited filter. It can improve clarity, catch some errors and prompt authors to address obvious weaknesses. It cannot guarantee that published work is correct, and it should not be treated as if it does.

Ethics of reviewing

Reviewers have responsibilities as well as power. The Committee on Publication Ethics (2017) sets out ethical guidelines for peer reviewers: to agree to review only work they are qualified to assess and can review in time, to declare conflicts of interest, to keep manuscripts confidential, not to use unpublished information for their own advantage, and to give constructive, objective feedback rather than personal criticism.

Writing a useful review

A good review helps both the editor and the authors. It usually begins with a short summary of what the paper claims, which shows the authors how their work was understood. It then sets out the main strengths and the most important problems, distinguishing issues that would change the conclusions from minor points of presentation. Useful comments are specific: rather than saying that the analysis is weak, a reviewer should explain which analysis, why it is a problem and what would address it. Reviewers should judge the paper the authors wrote, not the paper the reviewer would have written, and should be clear about which points are requirements and which are suggestions.

Responding to reviewers

For authors, the response to reviewers is part of the process. A clear response letter lists each comment and explains what was changed, with page or line references, or why a suggestion was not adopted. Disagreement is acceptable when it is reasoned and courteous. Editors generally look for evidence that authors have taken every serious point seriously, even when they do not accept it. Treating review as a dialogue rather than an obstacle usually improves both the paper and its chances of publication.

Reforms and alternatives

Several developments aim to address peer review's weaknesses:

- **Registered Reports**, in which review happens before results are known, so that decisions rest on the question and methods rather than the findings.
- **Open reports**, which make reviews public so that readers can see what was questioned and how authors responded.
- **Preprints with public comment**, which allow rapid sharing and wider scrutiny before and after formal review.
- **Post-publication review**, through comments, replication and reanalysis after an article appears.
- **Reviewer training and recognition**, including credit for reviewing in research evaluation.
- **Statistical and data checks**, by specialist reviewers or software, to catch errors that subject experts miss.
- **Partial randomisation in funding**, used by some funders to choose among proposals that reviewers judge to be of similar quality, acknowledging that fine distinctions between them are unreliable.

What peer review means for readers

For readers, including business readers using academic research, "peer-reviewed" is a useful signal but not a guarantee. It indicates that a few experts read the work and did not find fatal flaws. It does not mean that the results are correct, that the data were checked or that the findings will replicate. Readers should still ask the usual questions about any study: how large and representative was the sample, how were outcomes measured, were the analyses planned, and do other studies agree?

Conclusion

Peer review is a central institution of research, valued by researchers and relied on by readers. The evidence shows that it is less reliable, less consistent and less effective at detecting errors than its reputation suggests, and that it is vulnerable to bias. It remains useful, chiefly as a way of improving manuscripts and filtering out the weakest work. Reforms such as Registered Reports, open review and post-publication scrutiny aim to make it more transparent and more effective. The lesson for readers is to treat peer review as one sign of quality among several, not as proof that a finding is true.

References

Baldwin, M. (2018). Scientific autonomy, public accountability, and the rise of "peer review" in the Cold War United States. *Isis, 109*(3), 538–558. https://doi.org/10.1086/700070

Bohannon, J. (2013). Who's afraid of peer review? *Science, 342*(6154), 60–65. https://doi.org/10.1126/science.342.6154.60

Bornmann, L., Mutz, R., & Daniel, H.-D. (2010). A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants. *PLoS ONE, 5*(12), Article e14331. https://doi.org/10.1371/journal.pone.0014331

Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences, 14*(1), 119–135. https://doi.org/10.1017/S0140525X00065675

Committee on Publication Ethics. (2017). *Ethical guidelines for peer reviewers*. https://doi.org/10.24318/cope.2019.1.9

Csiszar, A. (2016). Peer review: Troubled from the start. *Nature, 532*(7599), 306–308. https://doi.org/10.1038/532306a

Godlee, F., Gale, C. R., & Martyn, C. N. (1998). Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: A randomized controlled trial. *JAMA, 280*(3), 237–240. https://doi.org/10.1001/jama.280.3.237

Jefferson, T., Alderson, P., Wager, E., & Davidoff, F. (2002). Effects of editorial peer review: A systematic review. *JAMA, 287*(21), 2784–2786. https://doi.org/10.1001/jama.287.21.2784

Kravitz, R. L., Franks, P., Feldman, M. D., Gerrity, M., Byrne, C., & Tierney, W. M. (2010). Editorial peer reviewers' recommendations at a general medical journal: Are they reliable and do editors care? *PLoS ONE, 5*(4), Article e10072. https://doi.org/10.1371/journal.pone.0010072

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology, 64*(1), 2–17. https://doi.org/10.1002/asi.22784

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research, 1*(2), 161–175. https://doi.org/10.1007/BF01173636

Mulligan, A., Hall, L., & Raphael, E. (2013). Peer review in a changing world: An international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology, 64*(1), 132–161. https://doi.org/10.1002/asi.22798

Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences, 5*(2), 187–195. https://doi.org/10.1017/S0140525X00011183

Pier, E. L., Brauer, M., Filut, A., Kaatz, A., Raclaw, J., Nathan, M. J., Ford, C. E., & Carnes, M. (2018). Low agreement among reviewers evaluating the same NIH grant applications. *Proceedings of the National Academy of Sciences, 115*(12), 2952–2957. https://doi.org/10.1073/pnas.1714379115

Rennie, D. (2016). Let's make peer review scientific. *Nature, 535*(7610), 31–33. https://doi.org/10.1038/535031a

Ross-Hellauer, T. (2017). What is open peer review? A systematic review. *F1000Research, 6*, Article 588. https://doi.org/10.12688/f1000research.11369.2

Schroter, S., Black, N., Evans, S., Godlee, F., Osorio, L., & Smith, R. (2008). What errors do peer reviewers detect, and does training improve their ability to detect them? *Journal of the Royal Society of Medicine, 101*(10), 507–514. https://doi.org/10.1258/jrsm.2008.080062

Smith, R. (2006). Peer review: A flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine, 99*(4), 178–182. https://doi.org/10.1177/014107680609900414

Tomkins, A., Zhang, M., & Heavlin, W. D. (2017). Reviewer bias in single- versus double-blind peer review. *Proceedings of the National Academy of Sciences, 114*(48), 12708–12713. https://doi.org/10.1073/pnas.1707323114

van Rooyen, S., Godlee, F., Evans, S., Black, N., & Smith, R. (1999). Effect of open peer review on quality of reviews and on reviewers' recommendations: A randomised trial. *BMJ, 318*(7175), 23–27. https://doi.org/10.1136/bmj.318.7175.23