

Quality and value: The true purpose of peer review

Nature (2006) | doi:10.1038/nature05032

What you can't measure, you can't manage: the need for quantitative indicators in peer review

Given its importance in steering the global research enterprise, peer review seems under-studied. There is a growing literature on the subject, some of which is highlighted at the quadrennial Peer Review Congress, but for the most part we are still only seeing snapshots ¹. A more systematic approach is needed if we are to understand peer review as it is currently practiced, or to evaluate the pros and cons of any alternative approaches.

Whether there is any such thing as a paper so bad that it cannot be published in any peer reviewed journal is debatable. Nevertheless, scientists understand that peer review per se provides only a minimal assurance of quality, and that the public conception of peer review as a stamp of authentication is far from the truth.

Given that many papers are never cited (and one suspects seldom read), it probably does not matter much to anyone except the author whether a weak paper is published in an obscure journal. Far more important is where a paper is published, and in fact this is the major function of peer review. It is generally understood among scientists that there is a hierarchy of journals ². At the apex of the (power law-shaped?) pyramid stand the most prestigious multidisciplinary journals; below them is a middle tier of good discipline-specific journals with varying degrees of selectivity and specialization; and propping up the base lies a large and heterogeneous collection of journals whose purviews are narrow, regional or merely unselective.

To succeed in science, one must climb this pyramid: in academia at least, publication in the more prestigious journals is the key to professional advancement. Some critics hold that journals should not fulfill this role, but this begs the question of what else might take their place. Competition is inherent to science, as to any activity where talented individuals strive for excellence. It is not just a matter of limited funding for grants and jobs. There is also an ever-increasing competition for 'mind space' among one's fellow scientists. With more than a million papers per year and rising, nobody has time to read every paper in any but the narrowest fields, so some selection is essential. Authors naturally want visibility for their own work, but time spent reading their papers will be time taken away from reading someone else's.

Scrutinizing peer review

Given the importance of peer review in determining the allocation of career rewards and public resources, it deserves close scrutiny. As a fan of 'Freakonomics', I suspect that an economic perspective might be enlightening, and in that spirit I have tried below to suggest some specific questions that, if answered, could illuminate both the costs and the benefits of the peer review system. In this age of digital publication, the relevant data are increasingly available, and so it should be possible - at least in principle - to supplement opinion and anecdote with quantitative evidence. (Many publishers have a vested interest in protecting the status quo and will be unwilling to open themselves

to critical scrutiny, but university libraries may be a more fruitful source of data.)

Not all journals are equal, and not all peer review is equal either. My own bias, as a former *Nature* editor, is that the answers to most of the questions below will differ widely between journals. In particular, I suspect the measurable benefits of peer review will be greater for the more prestigious journals. ('Prestige' is admittedly difficult to quantify, but relevant parameters include reputation among experts, acceptance rate, readership numbers and impact factor, relative to other journals in the same field.) Whether the costs of peer review are similarly skewed is an interesting question. My guess is that they are not, and that the cost/benefit ratio will be less favourable for the lower tiers of the scientific literature. If I am right, then this is where advocates for reform should focus their efforts.

Some measurable costs of peer review:

- How long is the interval from submission to publication, and how much of this is attributable to peer review? (This is difficult to quantify for papers that are rejected from one journal and eventually published elsewhere.) Publication delays are of course frustrating to individual authors competing for recognition, but in the race for priority one author's loss is another's gain. More important is the aggregate delay in the dissemination of new knowledge, which represents a cost to the scientific community and general public. It might be interesting to estimate the monetary value of this delay: if new knowledge represents the return on public investment in research, what is the cost of delaying the realization of this return?
- How much time do referees expend on peer review? Although referees may derive benefits from reviewing, it still represents time taken away from other activities (research, teaching and so forth) that they would have otherwise prioritized. Referees are normally unpaid but presumably their time has some monetary value, as reflected in their salaries.
- What is the actual cost of access to peer reviewed papers? Of course subscription costs pay for other things besides peer review (copy editing, print and online production, and so on), but operating the review system costs money, and most journal publishers will argue that peer review represents a substantial component of the value they add. The cost paid for *potential* access is easily calculated (site license fee divided by number of papers published), but a more interesting number would be the cost of *actual* access (the cost of a license divided by number of actual downloads). This would be a more meaningful measure of actual value generated (for a paper that is never read, zero) and a comparison across journals would probably be revealing

Some measurable benefits of peer review:

- How much do published papers differ from the initial submissions? In particular, how often do authors perform new experiments or analyses in response to referees' comments? At its best, the peer review system provides not only expert advice, but also a strong incentive for authors to heed the advice and to improve the paper. My experience as a *Nature* editor was that most papers went through considerable change (not always voluntary on the authors' part!) between submission and acceptance, but the *Nature* journals may be atypical.
- What is the acceptance rate after peer review? It might be interesting to address this question from the perspective of information theory. In principle, peer review acts as a filter, but if the acceptance rate is very high (or very low), then the amount of information added by the filter is low. (In practice this is difficult to address, because the answer will be confounded by the fact that the perceived stringency of the review process also affects decisions about submission. Experienced authors tend to target their submissions, sending only their more important papers to the most prestigious journals.)
- What value do readers derive from the current ranking system, as measured by their reading habits? To what extent do they make use of journal identity in deciding what to read? This question should be more easily answerable in the age of site licenses, now that 'high visibility'

and 'lower visibility' journals are equally accessible to institutional readers. Are the tables of contents pages browsed more frequently for the more prestigious journals, and are the corresponding articles more likely to be downloaded? It would be surprising if this were not the case, but how strong is the effect? Reading habits vary widely, and some scientists read much more than others, but we might expect journal identity to become more important as readers move further away from their core expertise (because outside one's area of specialization, there is less need to read comprehensively and thus more need for selection).

- How does publication in a prestigious journal affect career rewards such as recruitment and promotion, grant funding, invitations to speak at conferences, establishment of collaborations, student/postdoc applications, media coverage, and so on? Of course these represent benefits to individual authors (at the expense of other authors), but to the extent that information about journal identity is being used to allocate resources, there is also a presumed benefit to the users - that is, to hiring and funding committees, conference organizers, collaborators, students, science journalists, and so on. It is common to bemoan the over-reliance on quantitative markers such as impact factors for assessing scientists' abilities (and indeed there is much to bemoan), but until committee members have time to read every paper on every applicant's CV, they will have to rely at least in part on proxy indicators.
- Finally, the most important question is how accurately the peer review system predicts the longer-term judgments of the scientific community. One way to address this would be through citation data; articles that stand the test of time should be highly cited relative to others in the same field, even several years after their publication. Are such articles disproportionately represented in the most prestigious journals? Of course there are 'hidden gems' - papers that turn out to be important despite appearing in obscure journals - but are they the exception or the norm? Citation patterns vary by discipline, so any comparison would need to compare like with like, but in the age of semantic matching, this should be feasible. (Journal prestige is often equated - rightly or wrongly - with impact factor, but the impact factor calculation does not capture citations that happen more than 2 years post-publication.)

A tentative answer to this last question is suggested by a pilot study carried out by my former colleagues at *Nature Neuroscience*, who examined the assessments produced by *Faculty of 1000* (F1000), a website that seeks to identify and rank interesting papers based on the votes of handpicked expert 'faculty members'³. For a sample of 2,500 neuroscience papers listed on F1000, there was a strong correlation between the paper's F1000 factor and the impact factor of the journal in which it appeared. This finding, albeit preliminary, should give pause to anyone who believes that the current peer review system is fundamentally flawed or that a more distributed method of assessment would give different results.

Is there a better way ?

Peer review is not the one true solution for all time, and given the ever-increasing digitization of scientific communication, it would be foolish to think that no better solution can ever emerge to the problem of filtering scientific information. Many interesting alternatives have been suggested (to which I'll add one of my own: overhead cameras at the poster sessions of major meetings, tracking the flow of visitors to each poster). But I believe that the impressive rate of scientific progress over the past few decades is in part a tribute to the effectiveness of the current peer review system. The bar for any new alternative should thus be set fairly high. I suggest that any new system should meet the following criteria:

- It must be reliable - it must predict the significance of a paper with a level of accuracy comparable to or better than the current journal system.
- It must produce a recommendation that is easily digestible, allowing busy scientists to make quick decisions about what to read. A nuanced commentary on the merits and demerits of each

paper may be valuable to experts who have already read the paper (see contribution to this debate by [Koonin et al.](#)), but it will not help much with the initial screening.

- It must be economical, not only in terms of direct costs such as web operations, but also in terms of reviewer time invested.
- It must work fast. The peer review system produces clear-cut decisions relatively quickly (in part because editors pester reviewers to deliver their reports), whereas many forms of communal assessment - such as the emergence of a statistically significant pattern of citations or expert recommendations - are likely to be slow and gradual by comparison. Perhaps a popularity index (for example a 'most emailed' list) would provide a quick readout, but there is a danger of runaway amplification - the so-called 'Matthew effect', recognized by Robert Merton almost 40 years ago⁴ and likely to be exacerbated in the era of digital communication
- It must be resistant to 'gaming' by authors. Of course, savvy authors already know how to work the current system, but the separation of powers between editors and anonymous reviewers does - I believe - preserve some integrity to the process. After ten years as an editor, one thing I feel sure of is that if any alternative system becomes influential in determining career success, authors will seek ways to manipulate it to their advantage

Evolution not revolution

Finally, as others in this debate have emphasized, there is plenty of room for improvement within the current system. Editors are made not born, and electronic tools can help them do a better job. For instance, manuscript tracking systems can provide feedback on where delays arise and where resources are being allocated. (When I was executive editor of the Nature research journals, I occasionally risked my colleagues' ire by creating scatter plots to show the relationship - or lack thereof -- between individual editors' workloads and their median delays in making editorial decisions.) Even better, editors would benefit from feedback on the quality of their decisions. It would be interesting (for instance) to look retrospectively at the citations to accepted versus rejected papers, and to see whether editors vary in their ability to pick the winners. I still don't know what's the most highly cited paper that I ever rejected, but it would probably be chastening to find out.

RELATED LINKS:

[Peer review congress](#)

[Faculty of 1000](#)

REFERENCES

1. Rennie, D. *J. Am. Med. Ass.* **287**, 2759-2760 (2002)..
2. Lawrence, P. A. [The politics of publication](#). *Nature* **422**, 259-261 (2003).
3. Revolutionizing peer review? *Nature Neuroscience* **8**, 397 (2005).
4. Vandenbroucke JP. Medical journals and the shaping of medical knowledge. *Lancet*. 1998;352:2001-6.
5. Merton, R. K. The Matthew Effect in Science *Science* **159**, 56-63 (1968).