

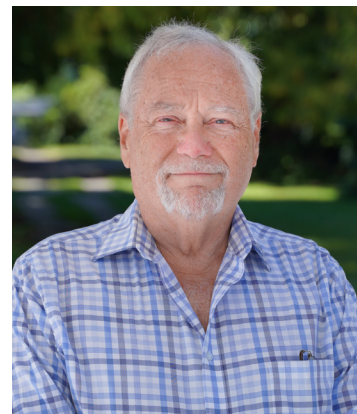
Reflections on Peer Review

A flawed process and... we still own it

By Dr. Wolfgang Linden (Vancouver)

General Observations

A desire to write this essay emerged from over four decades of being a manuscript reviewer and of course needing many reviews myself. It should be obvious but is still noteworthy that both authors and reviewers come from the same pool! They are not principally hostile to each other and each understands the other side. I never tracked how many reviews I have completed (nor how many I needed) but (averaging two to three reviewers per manuscript) I think I wrote over 400 reviews and needed a correspondingly larger number written by my peers (thank you, by the way). Even now, while officially retired, I write 15 to 20 reviews per year. However flawed the system is... and it is! (see 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, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0014331>)— I found reviewers rarely outright nasty; yes, they occasionally make critical comment on content that they simply read wrong (or missed), and some reviewers are annoying nitpickers... but on the whole I am grateful for how they helped improve the clarity and strength of my work.



Dr. Wolfgang Linden
Emeritus Professor
University of British Columbia

The trends in peer review management are ugly. Friends who are editors or associate editors struggle with finding reviewers, typically needing five attempts or more to recruit a single acceptance. And this trend is worsening. Apparently there is a growing skew in that researchers need a lot of reviews but do not carry their equivalent load while on the other side, being asked to provide reviews. In my opinion, this system is riding on the precipice of collapse and kind acknowledgments and data banks of reviewer activity seem to have little, though not unappreciated, impacts.

Wearing my Reviewer Hat

Should you have the good (or bad, you decide) fortune of having me review your work, be aware of the following. I will make a deliberate effort to find laudable features worth mentioning.

The prime driver of my final recommendations is whether the work is important for real world decisions and/or moves our field forward theoretically. I have limited praise for questionnaire studies in college students because the generalizability is rather low and many findings cannot be replicated elsewhere. On the other hand, I will fight for a high quality replication study and encourage full exploitation of data via secondary analyses provided the methods (which cannot be changed) are a strong match for the study rationale. On a different topic, it pains me having to tell

...cont'd

Peer Review: A flawed process and... we still own it

authors that their question may be 'an old hat' and despite possibly high internal validity the study is just not interesting any more. Fair enough, this is an opinionated decision but this criticism will keep articles out of highly ranked journals.

My perspective of reviewing has been distinctly shaped by research we did three decades ago (Linden W, Wen FK & Craig KD, (1992), Contributions of reviewer judgments to editorial decision-making for the Canadian Journal of Behavioural Sciences: 1985-1986. *Canadian Journal of Behavioural Sciences*, 24, 433-441). We had analyzed relationships among reviewers' appraisals and ultimate editorial decision-making for 120 manuscripts submitted to the *Canadian Journal of Behavioural Science*. Reviewers' ratings on eight standard evaluation criteria were evaluated for interrater agreement. The same criteria were used to predict the same reviewer's overall recommendation, and the reviewers' ratings were used to predict the editor's final decision to accept, reject, or request revisions. Interrater agreement on specific indices of manuscript quality were quite (and embarrassingly?) low. A principal components analysis revealed that reviewers effectively evaluated two factors: internal validity and importance of scientific contribution. Reviewers' ratings on specific criteria were correlated with moderate strength to their own final recommendation, and their recommendations were predictive of the editor's decision, explaining 63% of the variance. The bottom line here is that reviewers may identify different weaknesses in a given manuscript and that is often a good thing. Such variation does make them look disagreeable, but I find they largely agree on whether a study makes a significant contribution to the field.

The bottom line here is that reviewers may identify different weaknesses in a given manuscript and that is often a good thing.



A maybe odd observation is that younger researchers are generally very good at identifying relevant current literature, largely because they masterfully handle the literature search process on their own computers. I similarly benefit from this because I can get almost any fairly recent article in full length without lifting my butt out of the office chair; however... published articles are entered in electronic archives for only about the last three decades. I occasionally encounter authors who claim that a given study has never been undertaken before except that I know (because I am an old dog) that this is not true; I might have hardcopies of these papers in a drawer (fewer and fewer of course) or remember that 'xxx' and 'yyy' did a string of relevant and informative studies on this very topic in the sixties... but these are not locatable in e-archives.

...cont'd

Peer Review: A flawed process and... we still own it

Here are a few do's and don'ts (to get me suckered in... or thrown off)

Do

- Invest greatly in an introduction that shows the field's status quo about the most important questions, so that the development of the rationale is easy to follow and accept.
- Use the words "effect of" only if you conducted an intervention or controlled experiment.
- Use the term "predict" only if it was truly a longitudinal study with repeated measurements.
- Use structured guidelines for reporting (CONSORT or similar) wherever they exist.
- Write clear and informative abstracts (first impressions); what I first survey in a new manuscript is, in order: [1] the abstract, [2] relevant results tables, and [3] the beginning of the discussion where results are summarized.
- Provide effect sizes even if you stuck a good-old $p < .05$ approach overall.

Do not

- Simply state that your selected questionnaires "are reliable and valid." I see red when I read this. I recall one paper where a two-item questionnaire was used as the main measure to tap into an aspect of spirituality (already a tough concept to quantify) and the reader (me) was told it was reliable and valid; no detail offered. When I challenged the author/s, neither reliability nor validity information was provided in the revised manuscript but instead the reader was told that: "but other researchers have used this before." Guess what? This one went down in flames. As we all learned there are different types of reliability and even more types of validity and there is no consensus how much and what kind of validity is good enough. As for reliability, there is at least some consensus on internal consistency ($> .7$ ok, and $> .8$ good). And, as we were all taught in our undergraduate course on testing, a test cannot be valid when it is not reliable. Aside from these pivotal but generic requirements, authors must justify that their choice of measures is valid for their chosen population and study question.
- Covary out everything just because your stats package allows you to. Justify your choices.
- For the case of secondary analyses, do not report critical information on psychometrics or the design of the original study (upon which your secondary analyses rest) in unpublished work or obscure journals that I cannot locate.
- For a review, do not call it "comprehensive" if you excluded all publications that were not written in English; such a review is selective, not comprehensive. English being my second language, it is one of my pet peeves.

That's it... (for now)! Dialogue welcome.