Reviewing journal papers effectively, efficiently and fairly

David J. Pannell

Centre of Environmental Economics and Policy, School of Agricultural and Resource Economics, University of Western Australia, Crawley, WA 6009, Australia

E-mail address: david.pannell@uwa.edu.au

Paper presented at the 59th Conference of the Australian Agricultural and Resource Economics Society (AARES), Rotorua, 10-13 February 2015

9 February 2015
Working Paper 1501
School of Agricultural and Resource Economics
http://www.are.uwa.edu.au

© Copyright remains with the authors of this document.
Reviewing journal papers effectively, efficiently and fairly

David J Pannell

School of Agricultural and Resource Economics, University of Western Australia

Abstract: Academic journals rely on researchers agreeing to act as reviewers (or referees) for submitted papers. Researchers who publish papers themselves benefit from the contributions of those who review their papers, and so have a moral obligation to provide reviewing services for journals. However, reviewing is time consuming, and the total number of papers published continues to increase, so editors have found it increasingly difficult to obtain reviews. Further, the quality of reviews provided by reviewers varies greatly. This paper addresses various issues related to the journal refereeing process, including: tips on how to go about conducting a high-quality review efficiently and fairly; what can and cannot realistically be expected of referees; advice on dealing with ethical issues in refereeing; problems with the existing peer-review system used most commonly by journals; and a look at novel review systems.

Key words: publication, research, refereeing,

JEL classifications: A10, B40
Reviewing journal papers effectively, efficiently and fairly

David J Pannell

School of Agricultural and Resource Economics, University of Western Australia

Introduction

Research disciplines rely on peer review as their main mechanism for assuring the quality of published research. However, the peer-review process has been widely criticised. One common criticism is that it is not very reliable. It allows too many poor-quality papers to be published, and it results in the rejection of too many good-quality papers. The latter point is also related to another criticism of peer review – that it is inherently conservative. It has been observed that relatively innovative papers can be more difficult to get published than papers that push boundaries to a much more limited extent. There have been at least two examples of research papers that were initially rejected by journals but eventually won the Nobel Prize for Economics (Gans and Shepherd 1994). A third criticism is that the review process can be very slow (Pannell 2002).

These problems with the peer review process largely result from the design of the system and the incentives it creates for reviewers. To start with, conducting a review in response to an editor’s invitation is (a) voluntary, (b) unpaid, and (c) costly to the reviewer in terms of time and energy. Given these characteristics, reviewers have a limited incentive to accept the invitation to review, mainly depending the hope that they can learn something interesting or useful. If they decline, this contributes to the duration of the overall process. This lack of incentive has always existed but, interestingly, invited reviewers seem to have become more willing to decline review invitations since the introduction of web-based editorial management systems for most journals around a decade ago. These days it is common for editors to have difficulty finding a sufficient number of competent reviewers for some papers. Perhaps the more impersonal nature of web-based systems has made it easier for people to decline, or maybe the continuing growth in the number of papers submitted means that reviewers are more overloaded with reviews to undertake and have to decline more invitations.

Those people who do agree to review a paper are asked to do so within a limited time – these days often four weeks. However, they have little explicit incentive to meet this deadline, and some reviewers are very slow to meet their agreed commitment. If editors preferentially select reviewers who respond quickly (which some editors certainly do), and reviewers realise this, there is an incentive for them to respond slowly.

Of course, papers are eventually reviewed, because reviewers respond to other less-explicit incentives, perhaps including: dedication to the research discipline or the journal in question, the opportunity to learn something new and interesting\(^1\), a willingness to undertake tasks that one is personally asked to do, and a fear that the editor in question will not be willing to

\(^1\) Not a benefit that arises with all papers that need to be reviewed, particularly those submitted to lower ranked journals.
accept our future research papers if we do not do our fair share of the reviewing. Thus, the common-property nature of the research-publishing enterprise is at least partly manageable via social norms and peer pressure, consistent with Ostrom (2005).

Also consistent with Ostrom is the observation that, even with social norms and peer pressure in place, the functioning of the system is imperfectly. Indeed, there is plenty of published evidence that it can be highly imperfect (Rennie 1999; Smith 2010).

Perhaps the most important failing is the quality of reviews conducted. As a reviewer, one often gets to see the reviews conducted by other reviewers when a revised paper is sent back for re-review. Sometimes I see that another reviewer has identified problems that I failed to identify, and very often I see that other reviewers have failed to identify problems that to me seem obvious. Rennie (1986), an experienced journal editor himself, reflected on the common failure of the peer review system to detect problems with papers.

There seems to be no study too fragmented, no hypothesis too trivial, no literature too biased or too egotistical, no design too warped, no methodology too bungled, no presentation of results too inaccurate, too obscure, and too contradictory, no analysis too self-serving, no argument too circular, no conclusions too trifling or too unjustified, and no grammar and syntax too offensive for a paper to end up in print. (Rennie 1986, p. 3291)

Considering the many reviews that I have seen (as co-reviewer, editor or author), my view is that a minority of reviewers provide high-quality reviews – reviews that identify the important problems with a paper, avoid mistaken or irrelevant comments, and provide good advice on how to improve the paper. This is partly explicable in terms of the voluntary, unpaid and costly nature of the review process. It also in part reflects the limited supply of competent reviewers.

Given the small sample of reviews upon which editors must base their decision (usually two or three), there will inevitably be a significant number of papers that receive no high-quality reviews. For example, if 40% of reviews are of high quality, each paper receives three reviews, and good-quality reviews are independently distributed amongst papers, then over 20% of papers would receive no high-quality reviews. If each paper receives two reviews, then over 35% of papers would receive no high-quality reviews. There is also heterogeneity amongst researchers in judgements about what constitutes good research, contributing further to the randomness of the system.

To test the ability of reviewers to detect known errors, the editors of the British Medical Journal “took a 600 word study that we were about to publish and inserted eight errors (Schroter et al. 2008). We then sent the paper to about 300 reviewers. The median number of errors spotted was two, and 20% of the reviewers did not spot any. We did further studies of deliberately inserting errors, some very major, and came up with similar results.” (Smith, 2010).

A fourth characteristic of the existing peer-review system is that senior researchers have a disproportionate influence on the outcomes. For understandable reasons, editors tend to seek the advice of senior researchers as reviewers. The advantage of their knowledge and

---

2 In my judgement, this rarely influences acceptance decisions, but some reviewers may suspect that it could.
3 They may not be. In reality, editors are likely to try to include at least one competent reviewer for each paper.
experience outweighs the disadvantage of their lack of time, so that they are more likely to provide high-quality reviews than are their more junior colleagues. However, a likely consequence of this is a bias towards conservativeness. Senior researchers are relatively likely to have a strong commitment to existing theories, methods and paradigms, making them somewhat resistant to papers that conflict with currently disciplinary norms.

Reflecting on this issue, Armstrong (1982) proposed a set of rules for increasing the likelihood of acceptance of a paper under the existing peer-review system. According to Armstrong, authors should not (i) pick an important problem; (ii) challenge existing beliefs; (iii) obtain surprising results; (iv) use simple methods; (v) provide full disclosure; or (vi) write clearly. To the extent that this is true, it indicates a serious problem for research disciplines, reducing the importance and usefulness of their results.

In this paper, my main aim is to increase the number of high-quality reviews provided to journals by giving advice to reviewers on how to complete the task effectively, efficiently and fairly. Increasing the competence of reviewers may increase the acceptance of invitations to review, reduce the time taken to review, and increase the quality of completed reviews. Increasing the number of competent reviewers amongst less senior researchers may help to broaden the pool of favoured reviewers, lessening the conservatism of the system.

My secondary aim is to briefly discuss alternative approaches to reviewing that have been discussed and trialled by some journals, in attempts to avoid some of the perceived problems of the existing system. I will describe how these alternative systems affect the incentives faced by reviewers and discuss some of their pros and cons relative to the traditional system.

A focus on the peer-review process may have additional benefits, in terms of an improved understanding of the research process. Tracey Brown, Managing Director of UK science charity Sense About Science has observed that, “understanding peer review is key to developing informed opinions about scientific research”.

**Advice for reviewers**

Assuming that we maintain the current system of peer review, how can reviewers do their job to a high standard? The following sub-sections provide advice on various aspects of the process.

*Ethics*

Reviewers should decline invitations to review papers by authors with whom there is a conflict of interest. Clearly, this is only possible to ascertain in cases where the authors’ identities are revealed to the reviewers, either because the journal practises single-blind reviewing, or the identity of the author(s) is apparent to the reviewers for some reason. If you know who the authors are, can you can answer “yes” to one or more of the following questions?

- Have you had significant and acrimonious disagreements with the authors in the past?
- Are the authors and you co-investigators on a current research project?
- Have the authors and you jointly published an article in the past 5 years?
- Are you close personal friends with or closely related to one of the authors?
If so, you should consider declining. Reviewers must attempt to be impartial when evaluating a manuscript. Although it is difficult to be completely objective when assessing a paper that may not coincide with one's own beliefs or values, nevertheless, a reviewer must always strive for that goal. If a reviewer cannot separate the evaluation process from a desire to advocate a preferred theory or to reject the manuscript out-of-hand on philosophical grounds, then the reviewer should disqualify himself or herself from that review.

Do not allow the manuscript to be reproduced while in your custody. You must not use the manuscript for your personal advantage in any way. You cannot cite it or use its contents in any way until/unless it is published. If it is not published but you wish to use it, you need to contact the author (e.g. via the editor if the journal uses double-blind reviewing).

**Strategy**

Here is a strategy that I have found works for me. This strategy is my attempt to strike the right balance between efficiency and thoroughness.

Work from a printed copy. This makes it easier to make notes on it as you go, and to read it in places where you would not take a computer. I don’t have an iPad, but that may be a viable alternative to paper.

Read through the paper once, making notes at the level of detail that is appropriate for that paper. The better the paper, the finer detail you should go to. For a paper that is excellent and needs only very minor revisions to be accepted, note problems at a detailed level, including punctuation and spelling issues if you notice them. For a paper that is terrible, you would only note major issues, or even major groups of issues.

If you feel that you have understood the paper sufficiently, write your review immediately after your first reading. Otherwise, leave it for a day or so and come back to it. Read problem sections again, and then write your review.

When selecting your recommendation, consider the prestige of the journal and the usual standard of articles it includes. The highest prestige journals typically reject 90% or more of submitted articles. You are unlikely to be asked to review for such a journal early in your publishing career, unless you have managed to publish in one yourself. The least selective journals probably accept somewhere around 50-70% of submitted articles (after revision of course), but most journals probably accept between 20 and 50% of submissions.

Respond as quickly as you can manage. Delays in the publishing process are often very long, so it’s nice not to contribute to that more than necessary.

**Report**

Always number your comments. If you have multiple related comments, number them separately. This allows the editor and the authors to easily refer to your comments in further correspondence.

Don’t be excessively negative. Start by saying something positive about the paper, no matter how difficult this is. Say as many positive things as the paper deserves (or one more than that if it doesn’t deserve any) before you get into criticisms.
Some reviewers provide an overview of the paper. Personally, I don’t think this necessary in most cases, as the editor will read the paper. If I do include an overview, it is only a sentence or two long.

After your positive comments and your overview comment, start with your most important points. Note any concerns or problems at the big-picture level. This could include issues relating to the overall approach, the statistical methods, the interpretation of results, the quality of presentation or writing. If there are any concerns that you believe are essentially unfixable without more-or-less re-doing the research, you should say so clearly early in your report.

Justify all criticisms by specific references to the text of the paper or to published literature. Don’t just make vague or general criticisms. Back them up by giving the specific cases where they occur.

Note the page number and the line number for each comment if possible, or just the line number if lines are numbered sequentially throughout the paper, not separately for each page.

After you’ve given you most important comments, move to the less-important ones, which might be matters of detail.

A typical pattern for me is to write around three to five relatively important comments and 20 or more less-important comments.

There is no need to write a conclusion to your review. Just stop when you run out of comments.

If you have cited any literature, provide complete reference details.

Be careful not to identify yourself by your comments. Limit citations to your own published research. Sometimes it is appropriate to suggest your own work, but don’t over-do it. Be aware that if you do suggest your own work, you are increasing the probability that the author will identify you as the reviewer.

There is no fixed length for reviews. They are often between one and two pages long, but they can be somewhat longer than that if required, or shorter if not many comments are needed.

Checklist

In this section I suggest a set of questions that a reviewer could consider in the course of preparing the review report. You don’t need to provide comments on each of these points in every review, but I find it helpful to go through this list when formulating my review comments.

(a) Questions related to whether the paper is worth publishing at all

Is the research potentially worth publishing? It should be up to the authors to make the case that it is, and up to the reviewers to assess the case. Relevant questions include: Why is this work important? Why is it needed? What it is useful for? Which decisions would it influence? Who would use it? How much difference would it make?
Is the paper well written but not of sufficient interest or importance? So few papers I review are really well written that when I receive a well-written one to review, I can be lulled into being too uncritical about its content.

Is the paper methodologically innovative or mathematically impressive but not of sufficient interest or importance? Similar to the previous point, I can sometimes be seduced by a paper’s fancy methodology or impressive mathematics. But it is still important to ask whether the research question being addressed is sufficiently worthwhile. Is it a fancy model being used to address a trivial research question?

Is the analysis sufficiently well-grounded in the real world, or is it just an academic exercise? Some studies are over simplified, some rely on unrealistic assumptions, and some evaluate management or policy options that are unrealistic or implausible.

(b) Questions related to the quality of the research itself, apart from the write-up

Does the study employ a sound conceptual framework? Some authors make no attempt to build their analysis on a sound conceptual framework, but rely on an ad hoc approach that is not explained or justified in the paper.

Do the results presented pass a laugh test? Are they consistent with common sense? Are there any results that look illogical or inconsistent, making you worried about the data, the assumptions or the analysis?

If the study would need a control treatment or a baseline for comparison, does it include one? Is the control or baseline appropriately chosen?

Are there aspects of the study that could have biased the results? Issue could include sample selection, assumed parameters, method of analysis, choice of scenarios, wording of survey questions, simplifying assumptions, etc.

Have the assumptions used in the analysis been selected to achieve a particular result? For example, it is common to find analyses of optimal decisions under risk aversions that use exaggerated parameters of risk aversion – far greater than what the empirical evidence would support. Sometime I suspect that this has been done in order to make the study look more important or more interesting than it really is.

If the study includes a sensitivity analysis, is it a perfunctory sensitivity analysis? Sensitivity analysis is an extremely useful tool (Pannell 1997), especially for testing the robustness of conclusions. However, it needs to be done well, and often isn’t. Is it clear which conclusion are being tested in the sensitivity analysis, and is the sensitivity analysis done in way that actually does test the robustness of that conclusion?

(c) Questions about the write-up of the research

Is the written English adequate? I commonly recommend that authors should get expert help to improve their written English.

Could the text be made more readable? Is it needlessly obscure through the use of technical jargon or awkward sentence structures? Are there sentences that are too long? Do the paragraphs flow smoothly? Is there unnecessary repetition? Is the written expression clear and unambiguous? Clarity is vitally important. Whether or not you are an expert in the
subject discussed, you should understand the paper’s content. Do the authors run together long series of nouns. Here is a real example from one paper I reviewed: “Murray-Darling Basin sustainable agricultural water reallocation intervention approaches”. That is very hard on readers.

Assuming that the research is important enough to be potentially worth publishing, do the authors do a good job of motivating the paper and justifying the need for it? Do they convey answers to these questions asked above: Why is this work important? Why is it needed? What it is useful for? Which decisions would it influence? Who would use it? How much difference would it make?

Does the paper have clearly expressed aims, objectives or research questions? By the end of the Introduction, it should be clear what the authors aim to achieve with the paper. It should be stated quite explicitly. One purpose of doing this is to help the author achieve consistency between the title of the paper, the research questions articulated, the methods used, the results reported and the conclusions drawn. Check whether these things are consistent.

Are the aims, objectives or research questions explicitly addressed in the Conclusion (or the equivalent section, if there is no Conclusion section)? Ideally, you should be able to see in the Introduction what the authors aim to achieve, and then jump to the Conclusion and see what they actually achieved.

Is there any material presented that is extraneous to the stated aims, objectives or research questions? Authors often can’t resist including results or points of discussion that don’t actually relate to the stated aims, objectives or research questions of the paper. In general, they should be deleted.

Are the variables used clearly defined? If there are many variables used, it can be worthwhile including a table that describes them and gives their units of measure.

Do the authors present too many significant figures in the results? Usually, two significant figures is best. It makes it easy to compare results with each other, and it is sufficiently precise for the accuracy of the data and assumptions used. Never tolerate more than three significant figures.

Are the authors consistent in the labels and names used throughout the paper? Inconsistency in naming increases the risk of confusion for readers.

Are there too many acronyms? Unfamiliar acronyms can seriously reduce the readability of text. My rule of thumb is that a term should be used at least seven times in the paper to justify abbreviating it into an acronym.

Do the references cited really support the point being made? Some authors make a very specific statement and then support that statement with references that are broadly in the right area but don’t actually provide support for the very specific statement being made. This may not be obvious if you are not familiar with the references, but sometimes I get suspicious and go and check a couple of references.

Are statements about existing knowledge sufficiently supported by references? Any statement that could potentially be considered questionable should be supported by an appropriate reference.
Does the paper include sufficient information about the methods, data and assumptions for you to be able to make judgements about the quality of the results presented? If the information is voluminous, it might be best for the authors to put it in an online appendix of supplementary information. There is a growing trend for journals to publish online appendices of Supplementary Information. For example, at the American Journal of Agricultural Economics, it is no longer acceptable for authors to write that supplementary information is available on request. It must be provided with the submission, and is made available to reviewers. An even more stringent test is whether the information provided is sufficient to replicate the analysis. In principle it should be, and with the availability of online appendices, the practicality of making it sufficiently detailed is enhanced. Most journals do not actually require this more stringent question to be asked, but some do.

Does the paper comply with the word limit set by the journal?

Do you detect any plagiarism? There is no expectation for reviewers to check for plagiarism, but reviewers do sometimes identify cases of it.

(d) Questions about the presentation, interpretation and discussion of results

Are the results presented in a way that allows readers to understand and interpret them? Some authors devise very complex graphs that are very difficult to understand. Are the graphs and tables clearly and accurately labelled and captioned? They should be free-standing and not dependent on definitions or explanations provided in the main text. Can you suggest a better way to present the graphs? Are all the tables and figures really necessary?

Does the interpretation of a graph rely on colours? We may reach a time when all papers are read on screen in colour or printed on colour printers, but we are not there yet. It is still important for a graph to be easily interpretable when printed on a grey-scale printer. Authors need to use distinctive line thickness, shading, dashes or dots to distinguish series.

Is the interpretation of results sound? Sometimes an interpretation ignores additional relevant considerations, or the confidence placed in a conclusion is too strong given the evidence presented. Sometimes the author’s interpretation is not consistent with what it is in a graph or table.

Can you detect flaws in the model used or the statistical analysis? Altman (2002) found that flawed statistical analysis is common in medical journals.

Do the conclusions of the paper follow from the evidence presented? A common problem is for the authors to state conclusions that don’t actually follow from the evidence presented in the paper. They might be correct conclusions, but they are not conclusions of this research, because the research didn’t actually address those issues. Sometimes this occurs because the authors are trying to discuss the results in a relevant and interesting way, but they need to be careful not to go too far. Other times, the research does address the issues, but the authors express conclusions that seem to match their preconceptions rather than being supported by the evidence presented.

Do the authors provide intuitive explanations for results? Can you as a reader understand the underlying causes for the results from the explanations given? Are you convinced by the explanations? Are there alternative plausible explanations? Preferably, the explanations
provided should not be speculative but should be tested. If that is not possible, multiple speculative explanations should be provided.

*Have the authors identified and discussed their most interesting and important results?* Some authors present a lot of results but don't seem to recognise that some results are much more important (practically or academically) than others. If they have identified the most important results, are they right in their judgements about which is most important? Are the most important results discussed early on, or are they a bit buried in the discussion?

*Do the authors discuss the degree of confidence they have in the conclusions reached?* Is their degree of confidence justified by the evidence and analysis presented?

*In a study that involves economic optimisation or comparison of alternative management or policy options, do the authors discuss the possibility of flat payoff functions?* Elsewhere I have described how there are usually many sub-optimal solutions that are only slightly worse than the optimal solution (Pannell 2006). Many authors focus on the result that strategy A is optimal, but fail to point out that strategies B, C and D are only very slightly inferior. Given uncertainty in the analysis, B, C and D might actually be just as good as A. More generally, authors should address the question of how much their conclusions matter. Authors sometimes state that it is “important” to adopt a particular strategy, but fail to note there are other strategies that are only very slightly worse.

*Is the Discussion simply a repetition of aspects of the results?* This is a common weakness. Discussions should put the results in a broader context (e.g. how do they relate to previous research?), and consider their implications.

**Alternative systems for reviewing**

How might we go beyond relying on reviewers themselves to improve the quality of peer review? In response to the recognised weaknesses of peer review, a number of different systems of review have been proposed, and some have been trialled.

One set of ideas maintains a process of review prior to publication but increases the openness of the process. One example is the journal Atmospheric Chemistry and Physics, published by the European Geosciences Union. It has a two-stage review process, starting with an editor checking the paper for typing errors, clarity of figures, basic scientific quality, and consistency with the scope of the journal. After the authors have made necessary corrections, the paper is published as a discussion paper on the journal’s web site. It is then open for public review and discussion. Some of the reviews are invited by the editor, but anybody can post a review of the paper, or a comment on another review. Then the authors can revise the paper and prepare a response to the comments received. The editor reads all the material and makes a decision, potentially including requests for further revisions. If the paper is accepted, all the materials generated through the process are published, including both versions of the paper. The Wikipedia page on “Peer review” describes a number of variations on this type of approach used by other journals.

In terms of the incentives faced by reviewers, this type of approach seems quite similar to the traditional system. It is still voluntary, unpaid, costly and subject to biased and self-interested reviewers. The main difference is that the process is open, so that erroneous or biased review comments may be detected and countered by other reviewers. Whether this happens in
practice would depend on the number of reviewers who actively engage in the process. Looking at the Atmospheric Chemistry and Physics web site, it appears that a minority of papers receive additional reviews beyond those of the designated reviewers. Even without additional non-invited reviews, the fact that the process is open might cause reviewers to improve their behaviour (e.g. they may be less self-serving).

A more radical option is to publish everything and then let the community of scholars judge the quality of the work through a process of discussion, critique and use of the information (Smith 2010). This would involve changing the system from “filter then publish” to “publish then filter”. It would be assisted by some sort of public internet-based process. This would allow many readers to comment on a paper, either favourably or unfavourably, with good papers standing out through the levels of attention and praise that they receive over time.

Advantages of this approach include that review comments are public; that the outcome would be the consensus of many people rather than the views of two or three people; that the pressure on senior academics to review many papers in which they are not interested would diminish; and that it may be less expensive in terms of the transaction costs of the system. Also biased or self-interested reviewers cannot prevent the publication of high-quality papers. They can, of course, post adverse comments, but these can be countered by other more independent reviewers. Limitations could include that some published papers might never get a desirable level of scrutiny and critique; even those that did might not do so for a long while after publication; and biased and self-interested people might be most interested in posting reviews. Countering those possible limitations is the obvious success of Wikipedia and other similar approaches to crowd sourcing. Whether this approach would work for research publications is not clear, but I feel it is worth serious consideration.

Acknowledgments

Thanks to James Vercammen (currently an editor of the American Journal of Agricultural Economics) for his comments on a draft.

References


