

In the format provided by the authors and unedited.

Empowering peer reviewers with a checklist to improve transparency

Timothy H. Parker^{1,2*}, Simon C. Griffith³, Judith L. Bronstein³, Fiona Fidler^{4,5}, Susan Foster⁶, Hannah Fraser⁴, Wolfgang Forstmeier⁷, Jessica Gurevitch⁸, Julia Koricheva⁹, Ralf Seppelt^{10,11,12}, Morgan W. Tingley¹³ and Shinichi Nakagawa¹⁴

¹Department of Biology, Whitman College, Walla Walla, WA, USA. ²Department of Biological Sciences, Macquarie University, North Ryde, New South Wales, Australia. ³Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA. ⁴School of BioSciences, University of Melbourne, Melbourne, Victoria, Australia. ⁵History & Philosophy of Science, School of Historical & Philosophical Studies, University of Melbourne, Melbourne, Victoria, Australia. ⁶Department of Biology, Clark University, Worcester, MA, USA. ⁷Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, Seewiesen, Germany. ⁸Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, USA. ⁹School of Biological Sciences, Royal Holloway University of London, Egham, Surrey, UK. ¹⁰Department of Computational Landscape Ecology, UFZ – Helmholtz Centre for Environmental Research, Leipzig, Germany. ¹¹Institute of Geoscience and Geography, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany. ¹²iDiv – German Centre for Integrative Biodiversity Research Halle-Jena-Leipzig, Leipzig, Germany. ¹³Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA. ¹⁴Evolution & Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Randwick, New South Wales, Australia. *e-mail: parkerth@whitman.edu

Ten questions reviewers can use to improve transparency and reduce bias in the empirical literature.

Questions to promote transparent reporting of methods and results	
1	Were all sample sizes fully reported, including exact values for all subsets of data (e.g., each treatment group), and for all statistical analyses? →If 'no', request that authors provide this information.
2	Are the methods reported in sufficient detail to allow another researcher to gather the same data and run the identical analyses? →If 'no', request that authors provide the relevant information. →If you are uncertain about some aspect of the methods, state your uncertainty to the editor so that she or he can seek appropriate expertise as needed.
3	Are statistical results reported completely (considered in two parts below)?
3a	Are statistical results for each test reported in sufficient detail? What qualifies as 'sufficient detail' will differ among analyses. →If 'no', request that authors provide this information. →If you are uncertain, state your uncertainty to the editor so that he or she can seek appropriate statistical expertise as needed. Remember that you may be the only reviewer looking carefully at this aspect of the manuscript.
3b	Are results from all variables and from all models reported? Complete reporting should include results related to all variables examined in preliminary models and all results from exploratory analyses. →If 'no', request that authors provide this information. →If you are uncertain, ask the authors to declare in the paper that all exploratory analyses are reported in full. We recommend using the 'Standard Reviewer Statement for Disclosure of Sample, Conditions, Measures, and Exclusions' [see http://osf.io/hadz3].
Questions to check biases of reviewers and authors	
4	Were observers kept unaware of the experimental treatment imposed on the samples (e.g., organisms, plots) when recording observations or measurements so as to minimize unconscious bias? →If not stated, then request clarification in the manuscript of whether methods were adopted that reduced the possibility of unconscious bias influencing observations. →If no steps were taken to prevent observer bias, request that an explanation appear in the manuscript of how unconscious bias could have influenced observations.
5	Did the authors explain how sample size was decided (e.g., based on a priori power analysis or logistical constraints), or when an experiment with pre-set sample sizes was terminated? If sample size or the end of the experiment was not decided prior to the initiation of the study, was there a decision rule for when to cease data collection? →If not reported, request that authors provide this information. →If the stopping rule included iterative statistical tests or examination of patterns as data accumulated, request that authors acknowledge the bias resulting from this process.
6	Did the authors develop their analysis plan, including choices of variables, without looking at the data, for instance prior to gathering data or with a dummy data set? This is most easily determined by the existence of a pre-registered analysis plan. In the absence of pre-registration, a statement from the authors about the development of their analysis plan is still important. → If no, request that authors acknowledge the exploratory nature of their analyses and declare that they are reporting the complete set of results from all exploratory analyses.

	→ If authors deviated from their analysis plan, request an explanation of why and how they deviated from the plan.
7	How suitable do you find the research methods without considering the outcome? Evaluate the design and methods regardless of whether or not there was a finding of “statistical significance”, or whether or not the results conform to a predicted pattern. → If the methods appear to have been flawed, call attention to the problems and, if possible, recommend a better design. Deciding whether the problems with the methods are sufficient to justify a recommendation of rejection will require your expert judgement. → If uncertain about the suitability of some aspect of the methods, state your uncertainty to the editor so that she or he can seek appropriate methodological expertise as needed.
8	Are the sample sizes large enough to justify the authors’ conclusions? If presenting significance tests, how much power would this study have to detect statistically significant weak, moderate, and strong effects? Expectation of effect size can best be derived from average effect sizes presented in meta-analyses of similar topics. The effect size reported in the manuscript under review can be a poor estimate of the underlying effect size, especially if the sample size is small, which elevates sampling uncertainty. Statistical significance is a poor indicator of the reliability of an estimate across a wide range of sample sizes and common effect sizes. → If sample sizes are small in a system where effects are expected to be weak to moderate, request that authors avoid inferences based on threshold p-values, acknowledge uncertainty in effect size estimates, and acknowledge the need for further study. → Do not use sample size as a criterion for recommending publication unless you do so regardless of study outcome (i.e., regardless of reported effect size and regardless of the outcomes of tests for significance). → Do not use the failure to surpass a significance threshold as a reason to recommend rejection.
9	What does the size of the estimated effect (e.g., slope, correlation coefficient, difference in means) suggest about its biological or practical importance, and what does uncertainty around that effect estimate suggest about the estimate’s precision? → If the authors do not interpret their results in terms of the biological relevance of the effect and the uncertainty surrounding their effect estimate, request that they consider doing so.
10	How unexpected would you judge these results to be in light of prior empirically derived understanding? Effects that are more surprising in light of robust prior information are those that had a lower prior probability of being correct. → If a result is unexpected in light of prior evidence and is not supported by very strong new evidence (e.g., multiple lines of convincing evidence), do not recommend against publication on these grounds, but request that the authors acknowledge the tentative nature of their results.