Meta-Psychology, 2024, vol 8, MP.2023.3797
https://doi.org/10.15626/MP.2023.3797
Article type: Commentary

Open data: Not Applicable
Open materials: Not Applicable
Open and reproducible analysis: Not Applicable
Open reviews and editorial process: Yes
Preregistration: No

Edited by: Daniel Lakens
Reviewed by: Jakob Fink-Lamotte, Holger Brandt
Analysis reproduced by: Not Applicable
All supplementary files can be accessed at OSF:
https://doi.org/10.17605/OSF.IO/27NF6

Responsible assessment of what research? Beware of epistemic diversity!

Sven Ulpts[1]
[1]The Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University

Schönbrodt et al. (2022) and Gärtner et al. (2022) aim to outline in the target articles why and how research assessment could be improved in psychological science in accordance with DORA, resulting in a focus on abandoning the impact factor as an indicator for research quality and aligning assessment with methodological rigor and open science practices. However, I argue that their attempt is guided by a rather narrow statistical and quantitative understanding of knowledge production in psychological science. Consequently, the authors neglect the epistemic diversity within psychological science, leading to the potential danger of committing epistemic injustice. Hence, the criteria they introduce for research assessment might be appropriate for some approaches to knowledge production; it could, however, neglect or systematically disadvantage others. Furthermore, I claim that the authors lack some epistemic (intellectual) humility about their proposal. Further information is required regarding when and for which approaches their proposal is appropriate and, maybe even more importantly, when and where it is not. Similarly, a lot of the proposed improvements of the reform movement, like the one introduced in the target articles, are probably nothing more than trial and error due to a lack of investigation of their epistemic usefulness and understanding of underlying mechanisms and theories. Finally, I argue that with more awareness about epistemic diversity in psychological science in combination with more epistemic (intellectual) humility, the danger of epistemic injustice could be attenuated.

Keywords: responsible research assessment, epistemic injustice, epistemic diversity, epistemic (intellectual) humility

Schönbrodt et al. (2022) and Gärtner et al. (2022) aim to improve researcher assessment in psychological science according to the San Francisco Declaration on Research Assessment (DORA) towards more appreciation of quality with rigor and away from quantity of production. Although I appreciate the attempt to improve psychological science and the critique of the Journal Impact Factor (JIF) as an indicator for research quality, I think there are some crucial problems in their proposal. The unbalanced focus on methodological rigor, and the problematic undervaluation of theory in their proposal, the potential for gamification of the proposed procedure, as well as the important point that open science and responsible science (conduct) are two different things were already noted (see Dames et al., 2023). Therefore, my main focus will be a critique of their lack of humility and narrow understanding of how to conduct research in psychology and the resulting danger in neglecting the epistemic diversity within psychological science.

Psychological Studies not Psychology

The authors mention criteria for research rigor such as replication (reproducibility), computational correctness, statistical power, preregistration, and open data. However, those are not general indicators of research quality, they are indicators of research quality mostly appropriate for a certain approach to knowledge production (paradigm, system of practice, research program whichever philosopher of science you currently subscribe to). Those criteria might mostly be relevant for rather quantitative and statistical modes of knowledge production (ways of knowing) which presuppose some form of positivism. Importantly, the criteria can become very quickly almost meaningless in rather qualitative research that assumes some kind of interpretivism, constructivism or hermeneutics. Significantly, these approaches to knowledge production have their own criteria and terminology for the quality of research and practices to ensure said quality (e.g., criteria: transferability, dependability, confirmability & credibility; practices: positionality, reflexivity, member checking & thick description) that are (more)

tailored towards the combination of epistemological and ontological assumptions in those approaches (see e.g., Pownall, 2022; Stahl and King, 2020). Similarly, the proposal from the authors seems to be targeting research where writing has the purpose of reporting, but in rather theoretical, conceptual and hermeneutical work writing itself is largely the research product which therefore has to be judged based on different criteria because the act of writing also has different purposes (Penders et al., 2020). Hence, the authors neglect the plurality of approaches to knowledge production (ways of knowing)within psychological science and assume a fantastic uniform discipline that never existed (see e.g., Koch, 1993; Malich and Rehmann-Sutter, 2022). Furthermore, there seems to be some nuance missing when it comes to claims concerning the appropriateness and applicability of some open science practices across the scientific landscape in psychology (see e.g., Guzzo et al., 2022). For some research traditions and in some contexts, openness might not just be easier but also more appropriate than for others. This can be among other things for ethical reasons, or due to the availability of technology, infrastructure, and socioeconomic circumstances (Leonelli, 2022). This last point also demonstrates that general demands about how to conduct and report research neglect the situatedness of research in a specific larger context. Consequently, always scoring for open science practices such as open data and preregistration during research and researcher assessment would just give more privilege to some already privileged scientists and systematically ostracize others instead of opening up science and making it more diverse and inclusive. These are some of the reasons why the phrase open science buffet was introduced (Whitaker & Guest, 2020). For the above mentioned reasons, the proposed implementation of changes in the assessment of researchers in psychological science by Schönbrodt et al. (2022) and Gärtner et al. (2022) although well intended has the potential to systematically disadvantage a certain group of researchers due to their circumstances and their approaches to knowledge production. Therefore, the proposal from the authors implies the danger of committing epistemic injustice in psychological science (see Penders et al., 2019). Hence, the diversity within psychological science might make it necessary that the appropriateness and feasibility of criteria, such as the ones mentioned in the beginning of this paragraph, have to be evaluated on a case-by-case basis instead of pressured onto all of psychological science top-down.

Lack of Humility

To attenuate the potential danger of epistemic injustice perpetrated by such "improvements" on psychological science there seems to be a need for some epistemic (intellectual) humility (Hoekstra & Vazire, 2021; Merton, 1957). Part of humility is to be honest about the limits regarding the knowledge one claims, but also embedding those claims within the existing network of knowledge in the literature in form of referencing and signaling what and who came before us. Another component of humility, which can be understood as being a part of openness and transparency, is being honest about the scope and generalizability of one's claims. In the case of the proposal in the target articles this would mean that there should be a clear statement of where, when and especially for what kind of research these assessment practices are appropriate and applicable, but maybe even more importantly there needs to be a statement for which kind of research the proposal is inappropriate. Furthermore, since the research into the epistemic effects of proposed changes and criteria in the reform movement, such a reproducibility and preregistration are quite scarce and research that has been conducted presents less than promising findings (see e.g., Van den Akker et al., 2022) the authors of the target articles should be clearer about the fact that at this point such changes in the scientific ecosystem are probably nothing more than trial and error, especially when applied to the whole of psychological science. Furthermore, as hinted at before, part of humility is also acknowledging in which broader context your work is situated. There is quite some recent literature, research, and other agreements about responsible metrics (see e.g., "Agreement on reforming research assessment," 2022; Rushforth and Hammarfelt, 2022). The authors might want to consider relating their proposal more to recent development in that area. It can become problematic to ignore recent developments other than DORA considering that it is a decade old. Another omission of the authors is their somewhat unreflective and general use of the term reproducible. Reproducibility does not have an agreed upon meaning within science nor psychological science (see e.g., Matarese, 2022). For example, Nosek et al. (2022), who the authors of the target articles cite, use the concept of process reproducibility, which is quite similar to what Goodman et al. (2016) call methods reproducibility, but there is also amorass of meanings of reproducibility that contradict each other (Barba, 2018). Hence, if reproducibility is used as an indicator for research quality (I here ignore the issue of applicability or appropriateness of reproducibility as such a criterion for a specific approach to knowledge production) and subsequently as a criterion

for research(er) assessment, then the authors have to, at least, define the term or elaborate on which meanings of reproducibility are meant as an indicator/criterion and which are not. Such conceptual clarity could help to ensure that the tool is used during promotion and hiring as the authors intended, but also improve comparability of assessments that are based on their proposal.

This comment is by no means a defense for the use of the impact factor as an indicator for research quality, it is just a comment with which I try to show some nuance and complexity that is usually ignored or hidden in the reform movement due to a focus on a specific quantitative and statistical way to do research with the (sometimes implicit) aim to only produce a certain kind of knowledge as well as a tendency towards activism without a proper investigation or understanding of underlying principles, theories and mechanisms (see e.g., Devezer et al., 2021). Therefore, we have to ask ourselves at what point do too narrowly focused but then overgeneralized guidelines and recommendations regarding scientists' behavior and evaluation with the aim of improving science become careless, uninformed, and more dangerous than the current practices these proposals are supposed to rescue science from.

What now?!

I do not have any concrete and easy step-by-step recommendation for where to go from here, but I might have some vague directions. Right now, the authors of the target articles do not address responsible research in psychological science, but only the responsible conduct of a certain kind of research based on a very narrow understanding of what psychological science is. Not just a more inclusive awareness of epistemic diversity and the situatedness of research within psychological science is needed, but also an actual understanding of and investigations into the changes we promote. To actually conduct ourselves responsibly we have to be more humble and more honest about the circumstance that at this point a lot of these proposed improvements might be nothing more than trial and error and their (appropriate) scope are probably quite narrow or unknown. These directions could be important to attenuate the danger of externally forcing an unwarranted unification based on a narrow understanding of what constitutes (rigorous) research on to psychological science and the resulting epistemic injustice.

Author Contact

Correspondence concerning this article should be addressed to Sven Ulpts, The Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University, Bartholins Allé 7 building 1332, 8000 Aarhus C, Denmark. Email: su@ps.au.dk

Acknowledgements

I thank the Research Integrity peer group at The Danish Centre for Studies in Research and Research Policy at Aarhus University for insightful comments and critical discussions on an earlier draft.

Conflict of Interest and Funding

There are no conflicts of interest. There was no specific funding for this research.

Author Contributions

This is a sole-author paper.

Open Science Practices

This article is a commentary and is not eligible for any Open Science badges. The entire editorial process, including the open reviews, is published in the online supplement.

References

Agreement on reforming research assessment. (2022). https://coara.eu/agreement/the-agreement-full-text/

Barba, L. A. (2018). Terminologies for reproducible research. https://doi.org/https://doi.org/10.48550/ARXIV.1802.03311

Dames, H., Musfeld, P., Popov, V., Oberauer, K., & Frischkorn, G. T. (2023). *Responsible research assessment should prioritize theory development and testing over ticking open science boxes*. https://doi.org/https://doi.org/10.31234/osf.io/ad74m

Devezer, D. J., Navarro, Vandekerckhove, J., & Buzbas, E. O. (2021). The case for formal methodology in scientific reform. *Royal Society Open Science*, *8*(3), 200805–200805. https://doi.org/https://doi.org/10.1098/rsos.200805

Gärtner, A., Leising, D., & Schönbrodt, F. D. (2022). *Responsible research assessment ii: A specific proposal for hiring and promotion in psychology*. https://doi.org/https://doi.org/10.31234/osf.io/5yexm

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. (2016). What does research reproducibility mean? *Science translational medicine*, *8*(341), 341ps12–341ps12. https://doi.org/https://doi.org/10.1126/scitranslmed.aaf5027

Guzzo, R., Schneider, B., & Nalbantian, H. (2022). Open science, closed doors: The perils and potential of open science for research in practice. *Industrial and Organizational Psychology*, *15*(4), 495–515. https://doi.org/10.1017/iop.2022.61

Hoekstra, R., & Vazire, S. (2021). Aspiring to greater intellectual humility in science. *Nature Human Behaviour*, *5*(12), 1602–1607. https://doi.org/https://doi.org/10.1038/s41562-021-01203-8

Koch. (1993). "psychology" or "the psychological studies"? *The American Psychologist*, *48*(8), 902–904. https://doi.org/https://doi.org/10.1037/0003-066X.48.8.902

Leonelli. (2022). Open science and epistemic diversity: Friends or foes? *Philosophy of Science*, *89*(5), 991–1001. https://doi.org/https://doi.org/10.1017/psa.2022.45

Malich, L., & Rehmann-Sutter, C. (2022). Metascience is not enough — a plea for psychological humanities in the wake of the replication crisis. *Review of General Psychology*, *26*(2), 261–273. https://doi.org/https://doi.org/10.1177/10892680221083876

Matarese. (2022). Kinds of replicability: Different terms and different functions. *Axiomathes : Quaderni Del Centro Studi Per La Filosofia Mitteleuropea*, *32*(Suppl 2), 647–670. https://doi.org/https://doi.org/10.1007/s10516-021-09610-2

Merton, R. K. (1957). Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review*, *22*(6), 635–659. https://doi.org/https://doi.org/10.2307/2089193

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., et al. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual review of psychology*, *73*, 719–748. https://doi.org/https://doi.org/10.1146/annurev-psych-020821-114157

Penders, B., Holbrook, J., & de Rijcke, S. (2019). Rinse and repeat: Understanding the value of replication across different ways of knowing. *Publications (Basel)*, *7*(3), 52. https://doi.org/https://doi.org/10.3390/publications7030052

Penders, B., de Rijcke, S., & Holbrook, J. B. (2020). Science's moral economy of repair: Replication and the circulation of reference. *Accountability in Research*, *27*(2), 107–113. https://doi.org/https://doi.org/10.1080/08989621.2020.1720659

Pownall, M. (2022). *Is replication possible for qualitative research?* https://doi.org/https://doi.org/10.31234/osf.io/dwxeg

Rushforth, A., & Hammarfelt, B. (2022). *The rise of 'responsible metrics' as a professional reform movement: A collective action frames perspective*. https://doi.org/https://doi.org/10.31235/osf.io/cdmqz

Schönbrodt, F. D., Gärtner, A., Frank, M., Gollwitzer, M., Ihle, M., Mischkowski, D., & ... Leising, D. (2022). *Responsible research assessment i: Implementing dora for hiring and promotion in psychology*. https://doi.org/https://doi.org/10.31234/osf.io/rgh5b

Stahl, N. A., & King, J. R. (2020). Expanding approaches for research: Understanding and using trustworthiness in qualitative research. *Journal of Developmental Education*, *44*(1), 26–29.

Van den Akker, O., van Assen, M. A. L. M., Enting, M., de Jonge, M., Ong, H., Rüffer, F. F., & ... Bakker, M. (2022). *Selective hypothesis reporting in psychology: Comparing preregistrations and corresponding publications*. https://doi.org/https://doi.org/10.31222/osf.io/nf6mq

Whitaker, K. J., & Guest, O. (2020). #Bropenscience is broken science. *The Psychologist*, *33*, 34–37. https://www.bps.org.uk/psychologist/bropenscience-broken-science