

# Responsible Research Assessment requires structural more than procedural reforms.

Gidon T. Frischkorn<sup>1</sup>

<sup>1</sup>Department of Psychology, University of Zurich

In their target articles, Schönbrodt et al. (2022) and Gärtner et al. (2022) propose new metrics and their practical implementation to improve responsible research assessment. Generally, I welcome the inclusion of open science and scientific rigor into evaluating job candidates. However, the proposed reform mainly focuses on the first stage of selecting candidates who then continue towards a second stage of in-depth evaluation of research quality. Yet, this second selection stage is underdeveloped but likely more critical concerning responsible research assessment and hiring decisions. I argue that an adequate assessment of research quality at this second stage requires the representation of specific knowledge in the subfield of a discipline that the candidate should be hired for by the hiring committee. This is rarely achieved given the current structural organization of departments, especially in German-speaking countries, and potentially explains the reliance on suboptimal indicators such as h-index and Journal Impact factor. Therefore, I argue that responsible research assessment requires structural reform to ensure that institutions have several researchers in permanent positions with specific knowledge in different subfields to provide an adequate and responsible assessment of research quality by hiring committees at all evaluation stages.

*Keywords:* Research Assessment; Structural Reform; Hiring Committees

Schönbrodt et al. (2022) acknowledge that “the assessment of scientific quality will always remain a challenge with imperfect solutions” (p. 3). Their proposal attempts to face this challenge by introducing reformed metrics and criteria for the evaluation of candidates to “safeguard minimal standards of scientific rigor”. Generally, including the evaluation of scientific rigor in hiring decisions is desirable. However, as outlined by Dames et al. (2023), any metric, no matter if its citation count or a score for the level of scientific rigor, is insufficient for fully capturing scientific quality or its pre-requisites. Thus, it is critical to conduct a comprehensive qualitative evaluation of a candidate’s scientific contributions during a second evaluation phase. Although Schönbrodt et al. (2022) recognize the importance of such an in-depth evaluation of candidates in a second evaluation phase, its procedure remains underdeveloped. Given the complexity of assessing scientific quality and the heterogeneity of scientific work in different subfields of psychology, the individual researchers serving on hiring committees and their knowledge of the respective subfield as well as their commitment to responsible research assessment will be most relevant. Thus, not only the evaluation procedure but also the composition of hiring committees will need to be reformed to ensure sufficient expertise in the subfield a

position is advertised for.

The proposal by Schönbrodt et al. (2022) and its implementation outlined by Gärtner et al. (2022) focuses mainly on the initial evaluation of job applicants at a first stage of evaluation and outlines potential criteria to identify candidates that meet a minimum threshold of methodological rigor in their academic work. As already indicated in their proposal (see Figure 2 on p. 6; Schönbrodt et al., 2022), likely a large proportion of candidates will already or rather sooner than later meet this minimal threshold, thus leaving most candidates for a more comprehensive in-depth evaluation at the second stage of evaluation. For a tenured position this means that the hiring committee needs to decide who of at least 30 – more likely well above 50 – applicants remaining after the first evaluation stage should be invited to an interview. This further underscores the significance of the second evaluation stage during which the hiring committee is responsible for identifying the most viable candidates for a given position.

At the second stage of evaluation, Schönbrodt et al. (2022) advise against the usage of metrics altogether and propose that it should primarily focus on an “in-depth discussion about how innovative and meaningful the research” of the different candidates is. This is a step into the right direction, but it is also a lofty and

labor demanding goal. Additionally, to adequately evaluate the significance and potential innovation of an applicant's work within a specific subfield of psychology (or any other discipline), it is crucial to possess specialized knowledge of the current state of science in that respective subfield. For example, cognitive or social psychologists might not be ideally suited to comprehensively evaluate the quality of research in clinical or educational psychology, and vice versa.<sup>1</sup> Therefore, the adequate assessment of the quality of research in any subfield of psychology requires sufficient representation of researchers within this subfield in hiring committees. Given the current state of structural organization of long-term researcher positions, this specific knowledge is rarely represented, and most institutions will not be able to fulfill this requirement with researchers from their own institution.

Currently, hiring committees for permanent positions primarily consist of professors with long-term commitments to the respective institute. In most departments in German-speaking countries, there is only one tenured professorship for each subfield. When a person leaves that position and needs to be replaced, the expertise for the respective subfield is therefore no longer represented at that institution. Thus, hiring committees lack specific expertise in the field for which the position is open, except for the inclusion of typically one external expert from the respective subfield. As a result, despite an in-depth review of the few ultimately shortlisted set of applicants, the selection of this small number of candidates is conducted without sufficient expertise in the field for which the position is to be filled.<sup>2</sup>

The consequence of this situation is that the critical second stage of evaluation is often also based on the evaluation of insufficiently valid metrics and subjective evaluations of research impact, such as the journal the research has been published in (Abele-Brehm & Bühner, 2016). Ironically, Schönbrodt et al. (2022) aimed to shift the focus away from exactly these metrics during research assessment. Whereas researchers in specific subfields of psychology are likely able to put this information into context (for a discussion of potential benefits, see Dames et al., 2023), for researchers outside the subfield it might be difficult to separate high quality research published in lower impact journals from less impressive work that has been published in a high-prestige journal, for whatever reason that might be. No matter how dedicated the members of the hiring committee might be, an adequate and in-depth evaluation of the quality of research relevant to the respective position at this stage will always require sufficient expertise in the subfield the position is advertised for.

It is reasonable to assume that institutions will want

to maintain their independence in hiring decisions. Therefore, the expertise for a specific subfield cannot be entirely outsourced to experts from other institutions. To sufficiently represent the necessary expertise represented by scientists who are committed to the development of the institute, any institute needs to have in-house expertise in each subfield beyond the professor who has left and is to be replaced. This can only be achieved by structural changes. Specifically, any subfield should not be represented only by a single long-term position, but it should rather be distributed across several researchers. The more an institute has multiple tenured scientists (e.g., lecturers, assistant, or full professors) in the different subfields of psychology, the more it will be able to provide expertise for the critical second stage of evaluation in hiring committees itself. Yet, the current profile of single professorships representing each subfield - prevents institutions from building a sustainable basis of expertise in each subfield of psychology that is urgently needed for a more adequate assessment of research quality at the stage of narrowing down many adequate candidates to the small pool of candidates that get invited to job talks.

The need to improve current assessment procedures alone may not justify such a significant change in institutional structures. Yet, at this point, the requirements for responsible research assessment converge with other important arguments for structural reform (Rahal et al., 2023; Tiokhin et al., 2021): Moving towards departments with several independent tenured scientists representing each sub-discipline has the advantage of creating more tenured or tenure-track positions in science overall and offering some degree of job security at an earlier career stage. This will likely contribute considerably to improving research quality. In addition, such a structural change would provide departments with more flexibility to develop areas of specialization in which they can build up a critical mass of independent researchers with similar expertise and interests. This, in turn, can foster fruitful collaborations within the institution.

<sup>1</sup>In heterogeneous subfields this might even count for different areas of research. For example, a researcher focusing on language comprehension might be challenged by evaluating the work of a researcher in perception or decision making, although both might count as cognitive psychologist.

<sup>2</sup>I acknowledge that the evaluation of research quality will not be the only relevant criterion for hiring decisions at this second stage. The hiring committee will likely also focus on the fit of different candidates to the institute and the profile the committee is looking for.

### Author Contact

Gidon T. Frischkorn; Address: Binzmühlestrasse 14/22, CH-8050 Zürich, Switzerland; Mail: gidon.frischkorn@psychologie.uzh.ch; ORCID: 0000-0002-5055-9764.

### Conflict of Interest and Funding

The author declares no conflict of interest.

### Author Contributions

Gidon T. Frischkorn: Conceptualization, Writing.

### Acknowledgements

I would like to thank Hannah Dames, Philipp Musfeld, Vencislav Popov, and Klaus Oberauer for the initial discussion of the two target articles, as well as feedback and comments on this commentary.

### Open Science Practices

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

### References

- Abele-Brehm, A. E., & Bühner, M. (2016). Wer soll die Professur bekommen? *Psychologische Rundschau*, 67(4), 250–261. <https://doi.org/10.1026/0033-3042/a000335>
- 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/10.31234/osf.io/ad74m>
- Gärtner, A., Leising, D., & Schönbrodt, F. (2022). *Responsible Research Assessment II: A specific proposal for hiring and promotion in psychology*. <https://doi.org/10.31234/osf.io/5yexm>
- Rahal, R.-M., Fiedler, S., Adetula, A., Berntsson, R. P.-A., Dirnagl, U., Feld, G. B., Fiebach, C. J., Himi, S. A., Horner, A. J., Lonsdorf, T. B., Schönbrodt, F., Silan, M. A. A., Wenzler, M., & Azevedo, F. (2023). Quality research needs good working conditions. *Nature Human Behaviour*, 1–4. <https://doi.org/10.1038/s41562-022-01508-2>
- Schönbrodt, F., Gärtner, A., Frank, M., Gollwitzer, M., Ihle, M., Mischkowski, D., Phan, L. V., Schmitt, M., Scheel, A. M., Schubert, A.-L., Steinberg, U., & Leising, D. (2022). *Responsible Research Assessment I: Implementing DORA for hiring and promotion in psychology*. <https://doi.org/10.31234/osf.io/rgh5b>
- Tiokhin, L., Panchanathan, K., Smaldino, P. E., & Lakens, D. (2021). *Shifting the level of selection in science*. <https://doi.org/10.31222/osf.io/juwck>