

Responsible research assessment in the area of quantitative methods research: A comment on Gärtner et al.

Holger Brandt¹, Mirka Henninger², Esther Ulitzsch³, Kristian Kleinke⁴, and Thomas Schäfer⁵

¹University of Tübingen

²University of Zurich

³IPN—Leibniz Institute for Science and Mathematics Education

⁴University of Siegen

⁵MSB Medical School Berlin

In this commentary, we discuss the proposed criteria in Gärtner et al. (2022) for hiring or promoting quantitative methods researchers. We argue that the criteria do not reflect aspects that are relevant to quantitative methods researchers and typical publications they produce. We introduce a new set of criteria that can be used to evaluate the performance of quantitative methods researchers in a more valid fashion. We discuss the necessity to balance scientific expertise and open science commitment in such ranking schemes.

Keywords: Quantitative methods researchers, hiring criteria, open science

Gärtner et al. (2022) propose a new structured ranking procedure for hiring processes in psychology where they focus on three types of research contributions, namely journal articles, published data sets and research software (see also Schönbrodt et al., 2022). The algorithm they propose will provide scores for journal articles only if they include empirical data. They also value the publication of data sets if they follow the FAIR (Findability, Accessibility, Interoperability, and Reuse) format and the production of statistical packages. At the same time, the algorithm ignores alternative types of publication such as simulation studies, meta-analyses, theoretical contributions or literature reviews. The above authors propose that the resulting scores should be used by hiring or promotion committees to assess the scientific rigor before an actual review of the candidates begins. Although we welcome the general approach of a more structured assessment of research performance, the proposed strategy for the appointment process is largely inappropriate in the area of quantitative methods research. In the following, we first discuss some reasons why we disagree with the procedure and then propose an alternative approach.

Quantitative Methods Research

In Germany, virtually all psychology departments include a professorship for quantitative research methods. While some professorships are hybrids that combine, for example, cognitive or social psychology with quantitative methods, the majority of positions is focused on the

development and evaluation of quantitative methods. Research in quantitative methods differs from research in other psychological disciplines that focus on empirical research by collecting data in order to test hypotheses or theories that are relevant from a substantive point of view.

Instead, quantitative methods researchers focus on conceptual developments of new methods, for example, for situations when traditional methods are deemed inadequate. Besides many others, current research areas (in German speaking countries) include machine learning based approaches for social sciences that are interpretable (Henninger et al., 2023), Bayesian modeling and model selection (Heck et al., 2022), or new approaches for intensive longitudinal or other complex data (e.g., Nestler & Humberg, 2022; Orzek & Voelkle, 2023).

When quantitative methods researchers present their new methods, they typically test them with simulation studies and illustrate them with existing data sets from collaborators or publicly available data sets. They do not collect data themselves or write articles that could be subsumed under empirical articles as is necessary for the scoring algorithm to receive any points. Thus, when applying the new scoring algorithm, quantitative methods researchers will not receive a valid ranking of their excellency, but the majority of quantitative methods researchers may end up with uninformative zero points.

In the following, we propose alternative evaluation criteria for quantitative methods researchers. We will focus on the evaluation of journal articles, research soft-

ware contributions, and open science aspects.

Evaluation of Journal Articles

We propose three categories of journal articles that substitute the algorithm in Table 1 by Gärtner et al. (2022). These three types of articles are better capable to discriminate quantitative methods researchers in hiring or promotion processes.

1. **Methodological development:** Conceptual developments that propose new methods are likely to be the most important type of research article. The major parts of such articles include the conceptualization of the method, the derivation of their statistical features and interpretation of parameters, and the evaluation of the plausibility of their underlying assumptions. Most of these articles include a simulation study that highlights the properties of the method, and a re-analysis of already available empirical data set for illustration purposes.
2. **Performance evaluations of methods:** Many quantitative methods researchers provide information about the (relative) performance of existing approaches. The major statistical tool for such evaluations are simulation studies, where artificial data from a known population model are analyzed with different, potentially competing methods. Applied researchers can use the results from such studies to determine, for example, (a) which of the available methods are the most suitable to detect relevant effects, (b) to test if methods are robust to violations of underlying assumptions, or (c) determine necessary sample size conditions. Since the true underlying population model is unknown in empirical data, simulation studies are mandatory for the valid application of methods.
3. **Method dissemination:** A third type of journal articles focuses on publishing software itself or tutorial-type papers that are didactic illustrations of existing approaches for applied researchers. Open source software publications are helpful to provide information about new implementations of existing approaches. Tutorials are step-by-step descriptions that often include software code and helpful information about the interpretation of complex models. In recent years, both types of articles increased in relevance, since statistical methods became more sophisticated and often go beyond what can be feasibly taught during a Bachelor's or Master's Psychology program.

We believe that a rating based on these categories will help discriminate the potential of quantitative methods researchers. As an optional rating scheme, we propose the scheme shown in Table 1.

Evaluation of Research Software Contributions

The development of software packages is an important part of quantitative methods research. We agree with Gärtner et al. (2022), in most points, for example, with regard to "one-shot" descriptions vs. continuous package maintenance, the use of open reproducible scripts, and the reusability indicator (Table 3 in their paper). One major concern is the scoring and the weight that statistical packages received in their article. From our point of view, journal articles should be weighed at least equally to the production of statistical software. Scientific expertise that is relevant for a professorship in quantitative methods should not be outweighed by a more practical skill set of package writing. Conceptual and theoretical contributions that are original and move the scientific field forward should be the main focus of any hiring algorithm.

Open Science

Open Science is a vital aspect of responsible quantitative methods research. At the same time, it is important that hiring criteria are used that include and balance both open science commitment and scientific expertise that is reflected in scientific rigor and research quality (e.g., literature review, derivation of hypotheses, theory-building). We disagree with the scoring algorithm by Gärtner et al. (2022) that mostly ignores these aspects and only scores, for example, journal articles for their use of FAIR formats or preregistration. Instead, we propose to actively reward open science practices in addition to scientific rigor, by evaluating the above-mentioned criteria for journal articles also and explicitly with regard to FAIR formats of scripts and the use of open materials (see ID 11 and 12 in Table 1 in Gärtner et al., 2022) as well as open source programs. Reproducibility of simulation studies and feasible access by other researchers to new methods via open source software is an essential aspect of scientific methodological advancements.

Discussion

We believe that open science practices are a vital point in making hiring and promotion choices in general. Yet, the proposed scheme in Gärtner et al. (2022) cannot be applied to quantitative methods researchers as their research is not reflected in the proposed scoring

Table 1*Category scheme for journal publications*

Article type	Content	Contribution
Methodological development	Completely new approach for a relevant problem	Major
	Adaptation of existing approach that, for example, improves power or similar	Relevant
Performance evaluation	Variation of an existing approach, for example, for a specific data type	Minor
	Comparison of different approaches under a large set of conditions	Major
Method dissemination	Medium-sized simulation study for a specific set of conditions	Minor
	Open-source software publication	Major
	Tutorial paper with detailed software implementation	Major
	Short tutorial paper	Minor

algorithm. We proposed alternative criteria for quantitative methods researchers and provided information about relevant categories that may help to "filter" or rank excellency of quantitative methods researchers.

It remains debatable whether it is expedient to replace journal or citation metrics with open science metrics that may even be in contrast to the San Francisco Declaration on Research Assessment (DORA) and Coalition for Advancing Research Assessment (COARA) principles themselves. These principles include a qualitative assessment of scientific content as well as indicators of research impact such as influence on policy and practice (see [here](#)). Quantifying quality is problematic and simply scoring points based on a fixed template oversimplifies the topic under investigation.

The originality of ideas and contributions that advance knowledge should be judged as primary information for hiring processes as it is included in the COARA agreement (p. 3, see [here](#)). It is mandatory that such responsible research includes open science practices and it should be evaluated in light of these principles.

Finally, we would like to point out a few cautionary notes. While a transparent hiring process is important, a strictly formalized review process might also set the wrong incentives, for example, when young researchers aim at checking every box of a scoring algorithm, while ignoring other aspects or standards in their research domain. The use of fixed criteria carries the risk of narrowing down a scientific career to "profitable" strategies in scientific practice, at the expense of scientific creativity. Other aspects that may be influenced negatively by the proposed procedure address gender equality, equal opportunities, and inclusiveness. For example, potential gender bias may occur when using the proposed self-assessment of scientific rigor because female applicants may self-rate their performance worse than male applicants (Fletcher, 1999).

Author Contact

Holger Brandt: holger.brandt@uni-tuebingen.de (ORCID)

Mirka Henninger: mirka.henninger@uzh.ch (ORCID)

Esther Ulitzsch: ulitzsch@leibniz-ipn.de (ORCID)

Kristian Kleinke: kristian.kleinke@uni-siegen.de (ORCID)

Thomas Schäfer: thomas.schaefer@medicalschoo-berlin.de (ORCID)

Conflict of Interest and Funding

No conflict of interest to declare.
No specific funding for this research received.

Author Contributions

(in alphabetic order) Holger Brandt – Conceptualization and writing of the original draft. Mirka Henninger – Conceptualization and writing of the original draft. Kristian Kleinke – Review and editing. Thomas Schäfer – Review and editing. Esther Ulitzsch – Review and editing.

Open Science Practices

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

References

Fletcher, C. (1999). The implication of research on gender differences in self-assessment and 360 degree appraisal. *Human Resource Management Journal*, 9, 39–46. <https://doi.org/10.1111/j.1748-8583.1999.tb00187.x>

- Gärtner, A., Leising, D., & Schönbrodt, F. D. (2022). Responsible research assessment II: A specific proposal for hiring and promotion in psychology. <https://doi.org/10.31234/osf.io/5yexm>
- Heck, D. W., Boehm, U., Böing-Messing, F., Bürkner, P.-C., Derks, K., Dienes, Z., Fu, Q., Gu, X., Karimova, D., Kiers, H. A., et al. (2022). A review of applications of the bayes factor in psychological research. *Psychological Methods*. Available for free on *PsyArXiv*.
- Henninger, M., Debelak, R., Rothacher, Y., & Strobl, C. (2023). Interpretable machine learning for psychological research: Opportunities and pitfalls. *Psychological Methods*. <https://doi.org/10.1037/met0000560>
- Nestler, S., & Humberg, S. (2022). A lasso and a regression tree mixed-effect model with random effects for the level, the residual variance, and the autocorrelation. *Psychometrika*, 87(2), 506–532. <https://doi.org/10.1007/s11336-021-09787-w>
- Orzek, J. H., & Voelkle, M. C. (2023). Regularized continuous time structural equation models: A network perspective. *Psychological Methods*. <https://doi.org/10.1037/met0000550>
- Schönbrodt, F., Gärtner, A., Frank, M., Gollwitzer, M., Ihle, M., Mischkowski, D., Phan, L. V., Schmitt, M., Scheel, A., Schubert, A.-L., Steinberg, U., & Leising, D. (2022). Responsible research assessment I: Implementing dora for hiring and promotion in psychology. <https://doi.org/10.23668/psycharchives.8162>