

Research assessment using a narrow definition of “research quality” is an act of gatekeeping: A comment on Gärtner et al. (2022)

Thomas J. Hostler¹

¹Manchester Metropolitan University

Gärtner et al. (2022) propose a system for quantitatively scoring the methodological rigour of papers during the hiring and promotion of psychology researchers, with the aim of advantaging researchers who conduct open, reproducible work. However, the quality criteria proposed for assessing methodological rigour are drawn from a narrow post-positivist paradigm of quantitative, confirmatory research conducted from an epistemology of scientific realism. This means that research conducted from a variety of other approaches, including constructivist, qualitative research, becomes structurally disadvantaged under the new system. The implications of this for particular fields, demographics of researcher, and the future of the discipline of psychology are discussed.

Keywords: open research, research assessment, epistemology, metrics, qualitative

Schönbrodt et al. (2022) and Gärtner et al. (2022) – “the taskforce” – correctly identify that changes to current researcher evaluation are needed to make publishing as many novel – but not necessarily rigorous – papers a less productive strategy for a successful career in science (see Bakker et al., 2012). The most direct way in which research evaluation is practically imposed on researchers is in hiring and promotion criteria. Gärtner et al. (2022) propose a new framework for hiring and promotion practices in psychology, where a candidate’s outputs are quantitatively scored against a number of criteria for determining methodological rigour. These scores can be used to shorten an initial pool of candidates to a manageable shortlist to proceed to further, qualitative criteria for making a final decision on employment. This commentary argues that the quantitative criteria proposed by Gärtner et al. are based on definitions of quality that are applicable only to specific epistemological positions and methodological approaches, and therefore structurally disadvantage the numerous other approaches in psychology by rating them as relatively ‘lower quality’.

Epistemologies and Methodologies in Psychology

Disciplines considered ‘hard sciences’ such as biology typically adopt a unitary epistemology of scientific realism, subscribing to a view of a ‘true’ reality independent of the researcher, and utilise homogeneous scientific methodology - including hypothesis testing and quantitative data collection - to investigate this reality. In contrast, psychology spans a broad range of episte-

mological positions and methodologies. Although these different elements are often incorrectly conflated (Syed and McLean, 2022), certain epistemological positions and methodologies are usually associated together as a broad approach or ‘research paradigm’ (Morgan, 2007), two of which are particularly common in psychology.

First, there are psychological researchers who, like biologists, also adopt an epistemology of scientific realism and use quantitative methodology (an approach often incorrectly referred to as ‘positivism’; although ‘post-positivism’ is more accurate; see Michell, 2003). An example of such an approach would be in cognitive psychology, with a researcher who assumes a ‘true’ underlying cognitive architecture that is the same across most human beings. This researcher may design experiments to test theories about this architecture, and employ deductive hypothesis testing procedures on quantitative data to falsify or support her predictions.

A second broad paradigm in psychology is the use of qualitative methods alongside an epistemology of constructivism-interpretivism, which assumes no single ‘true’ reality, and posits instead that reality is constructed and experienced differently for different individuals. An example of this approach might be a health psychologist interested in the lived experiences of people who have gone through a traumatic medical procedure. She may use qualitative methods such as interviews to collect data and try to interpret this inductively to come to an understanding of the individual’s unique experience. The huge scope of the subject of psychology – the mind and behaviour of human beings - leaves

ample room for multiple approaches with varying combinations of epistemology and methodology, dependent on the research question and purpose of the research. The value of any specific approach to psychological research is a complex issue that rests heavily on one's attitude towards methodological pluralism (see Madill and Gough, 2008). Crucially though, the correct way to assess the rigour (and by extension quality) of any piece of research must be based on an understanding of the unique characteristics of its epistemology and methodology.

Definitions of Quality from a Narrow Approach to Research

Although the taskforce's motives may be sound, a critical issue with Gärtner et al.'s system is that the proposed standards for assessing methodological rigour apply only to a narrow methodological and epistemological approach to psychological research. Gärtner et al. refer to the criteria as pertaining to "empirical" work, but the proposed standards for assessing rigour include the existence of pre-registrations (criteria 13+14), the computational reproducibility of analyses (criteria 9), as well as the presence of replication studies (criteria 16). These are all methodological procedures which have been developed from a single narrow paradigm of quantitative methodology, a-priori confirmatory hypothesis testing, and an epistemology of scientific realism. This means they do not apply or are ill-fitted to assessing the rigour (and by extension quality) of research conducted using alternative approaches.

In particular, the criteria are unsuitable for a wide breadth of research from the "psychological humanities" (Teo, 2017) which use qualitative approaches. Specifically, the criteria of pre-registration, data sharing, computational reproducibility, and replication may all be either inappropriate or difficult to apply to qualitative methodology and constructivist-interpretivist epistemology. First, the usefulness of pre-registrations for qualitative research is contested (Coffman and Niederle, 2015) and if they are useful it is for different reasons than for quantitative research (Haven et al., 2019). As such, qualitative preregistrations do not have a direct link to 'methodological quality' in the same way that quantitative preregistrations do, and the specific facets in Gärtner et al.'s scoring system of "hypotheses" "power analysis" and "operationalizations" are not applicable.

Second, both technological barriers to anonymizing typical qualitative data such as audio and video (Weitzenboeck et al., 2022) and ethical issues with sharing sensitive data or obtaining consent to share it (Jacobs et al., 2021) mean sharing qualitative data is significantly harder and more time consuming, if

not impossible in many situations. Third, the criterion of "correctness of computational reproducibility" is inapplicable to data analysis conducted from a subjective, constructivist-interpretivist epistemology. Finally, the idea that research conducted from interpretivist positions is (or should be) "replicable" is debated (Tuval-Mashiach, 2021), meaning the existence of "pre-registered replication studies" is rare in qualitative enquiry. If replications of qualitative studies are performed, their link to 'research quality' is tangential from an ontological perspective (Pratt et al., 2020).

Structurally Disadvantaging Qualitative Researchers

The application of these criteria to assessing research quality means that researchers using qualitative approaches submitting "empirical" articles to Gärtner et al.'s system would often score 0, which is the proposed number of points awarded even if a specific criterion is deemed (and justified) as "non-applicable" to a particular paper. In contrast, quantitative researchers are more likely to score highly and be shortlisted during the application process given it is much more common, relevant, intuitive, and easier for them to pre-register their studies, share data, and whose work by its very nature is more likely to involve replication attempts. The inevitability of Goodhart's law also means that 'open-washing' studies by sharing incomprehensible (yet technically reproducible) datasets and analysis code would be an easy way for quantitative researchers to "game" Gärtner et al.'s assessment criteria and score 'research quality' points, to gain advantages over qualitative researchers who perhaps cannot ethically share their data. Simultaneously, qualitative researchers would be incentivised to waste time pre-registering their studies (for debatable methodological benefit) to score points.

Implementing a scoring system in hiring criteria based on post-positivist, quantitative assumptions is representative of the wider challenges faced by qualitative and humanities scholars. Existing systems for assessing and evaluating research such as ethical review procedures are typically based on assumptions of scientific realist epistemology and utilise terminology and norms from quantitative methodology (Musoba et al., 2015). Qualitative researchers are often cast as "guilty until proven innocent" by such assessment criteria (Macdonald and Carnevale, 2008), requiring them to justify their methods to those from a post-positivist background until they are seen as scientifically acceptable. This trial is epitomized in Gärtner et al.'s scoring system by requiring qualitative researchers to justify why pre-registration or open reproducible scripts are "not applicable" for their papers.

Could Gärtner et al.'s criteria be adapted to include

separate quality criteria for qualitative or constructivist research to overcome this problem? Whilst guidelines for assessing the rigour of qualitative research do exist (Johnson et al., 2020) their use as formal criteria to be used to judge overall research “quality” is not widely accepted. Schönbrodt et al.’s assertion that methodological rigour “goes a long way” to establishing whether research has “a high intrinsic quality” is directly challenged by researchers from qualitative approaches who argue that a focus on methodology can “over-simplify and inappropriately standardize the complex and non-formulaic nature of qualitative inquiry, promoting the notion of a fixed relationship between research practice and knowledge generated” (Eakin and Mykhalovskiy, 2003). These authors argue that a formulaic approach to assessing the “methods” used misunderstands the nature of qualitative research and the type and quality of knowledge generated from it, which acknowledges the roles of context, subjectivity, and ‘substantive judgement’ - concepts that by definition resist quantification.

Gatekeeping Psychological Research

Gärtner et al. acknowledge that “it is always possible to find examples” of research that is not rewarded by their scoring criteria, but that empirical (i.e., quantitative) articles represent “the vast majority of publications in our field”. Whilst it may be true in sheer volume that more research is published using quantitative compared to qualitative approaches, this simplification does not take into account that different research paradigms are associated with different fields of psychology. Many subdisciplines of psychology have shared philosophical roots with critical theoretical work in the humanities, e.g. feminist psychology, critical race psychology, and critical disability studies in psychology, and so qualitative and non-empirical work is more common in these fields. Applying criteria for assessing methodological quality that disadvantages certain types of research methodology is therefore an act of gatekeeping that by extension values certain fields of psychology over others. Furthermore, as many of these subdisciplines have a strong tradition of employing a participatory and advocational ethos, the type of researcher working in these disciplines often (although not exclusively) shares the demographic characteristics of the area of study (i.e., women, people from ethnic minority backgrounds, people with disabilities). By extension, gatekeeping these areas of research by judging them to be of lower ‘rigour’ is also structurally disadvantaging researchers from these groups.

Psychology as a discipline is often described as a “STEM” subject (Science, Technology, Engineering, and Maths), yet this is not a universal opinion. In the

UK, the Quality Assurance Agency that defines the academic standards of degree courses and the ‘nature of the subject’ specifically acknowledges that psychology uses a diverse array of methodologies including qualitative methods (Q.A.A., 2019). Indeed, even if the common description of psychology as a ‘science’ may reflect the current high proportion of quantitative, post-positivist research on the subject of the human mind and behaviour, a future transition of psychology into a ‘pure’ post-positivist science is far from an inevitability. Interest in and acceptance of qualitative research is increasing (Gough and Lyons, 2016) and there have been renewed calls for more engagement with the ideas and methods of qualitative paradigms in the discipline (Gough and Madill, 2012; Teo, 2017). Simultaneously, there have been devastating critiques of the epistemological and theoretical foundations of much of quantitative psychology (Eronen and Bringmann, 2021; Richters, 2021; Uher, 2023). A coherent and productive future for psychological research will likely involve embracing and utilizing epistemological positions and methodologies beyond scientific realism and quantification (Valsiner, 2020; Sanbonmatsu et al., 2023). Marginalizing work using qualitative approaches is likely to be harmful to the discipline as a whole in the future and represents a form of gatekeeping with significant long-term implications.

It may be that the taskforce and/or the German Psychological Society support the view that psychology should be a STEM subject and that only post-positivist, quantitative, scientific work should be considered “rigorous”. If this is the case, then they should explicitly state this as an aim in their guidelines for research assessment and acknowledge the implications of the likely outcomes of Gärtner et al.’s current narrowly defined rigour criteria. However, it could also be an oversight. It is clear that the context to these developing guidelines for research quality is the replication crisis, and the alarmingly high proportion of non-replicable research conducted in experimental quantitative psychology over the past two decades. It is a worthwhile pursuit to not reward researchers conducting sloppy, non-reproducible quantitative science with promotions, and using assessment criteria based on open science practices specifically designed to weed out this type of research, Gärtner et al.’s proposals would likely be effective. However, by targeting this type of research they also inadvertently keep out a wide range of other approaches to research that do not fit the quality criteria, including psychological humanities, qualitative methodologies, and other non-quantitative approaches not discussed in depth here (e.g. clinical case studies). Applying a ‘one-size-fits-all’ approach to promoting open sci-

ence practices has been widely criticized by numerous commentators (e.g., Bazzoli, 2022; Bergmann, 2023; Field et al., 2021; Huma and Joyce, 2022), but by indiscriminately rewarding their use in hiring practices for all psychology researchers (and thereby punishing their absence), Gärtner et al.'s proposal epitomizes exactly this ethos.

Hiring and promotion criteria are the most effective form of gatekeeping: those without the resources of a job in academia are not just marginalized but are prevented from participating at all in academic discourse. As such it is vital to make sure that criteria are fair for assessing the quality of all types of research, and by extension the researchers who conduct these.

Author Contact

email: t.hostler@mmu.ac.uk
ORCID: 0000-0002-4658-692X

Acknowledgements

I would like to thank Geoff Bunn for his helpful feedback on the commentary.

Conflict of Interest and Funding

I declare no conflict of interest for this commentary. There was no funding attached to this work.

Author Contributions

I am the sole author of the article.

Open Science Practices

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

References

- Bakker, M., Dijk, A., & Wicherts, J. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, 7(6), 543–554. <https://doi.org/10.1177/1745691612459060>
- Bazzoli, A. (2022). Open science and epistemic pluralism: A tale of many perils and some. *Industrial and Organizational Psychology*, 15(4), 525–528. <https://doi.org/doi:10.1017/iop.2022.67>
- Bergmann, C. (2023). The buffet approach to open science. <https://cogtales.wordpress.com/2023/04/16/the-buffet-approach-to-open-science/>
- Coffman, L., & Niederle, M. (2015). Pre-analysis plans have limited upside, especially where replications are feasible. *Journal of Economic Perspectives*, 29(3), 81–98. <https://doi.org/10.1257/jep.29.3.81>
- Eakin, J., & Mykhalovskiy, E. (2003). Reframing the evaluation of qualitative health research: Reflections on a review of appraisal guidelines in the health sciences. *Journal of Evaluation in Clinical Practice*, 9(2), 187–194. <https://doi.org/10.1046/j.1365-2753.2003.00392.x>
- Eronen, M., & Bringmann, L. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, 16(4), 779–788. <https://doi.org/10.1177/1745691620970586>
- Field, S., Ravenzwaaij, D., Pittelkow, M., Hoek, J., & Derksen, M. (2021). Qualitative open science – pain points and perspectives. <https://doi.org/10.31219/osf.io/e3cq4>
- Gärtner, A., Leising, D., & Schönbrodt, F. (2022). Responsible research assessment II: A specific proposal for hiring and promotion in psychology (Preprint) [PsyArXiv]. <https://doi.org/10.31234/osf.io/5yexm>
- Gough, B., & Lyons, A. (2016). The future of qualitative research in psychology: Accentuating the positive. *Integrative Psychological and Behavioral Science*, 50(2), 234–243. <https://doi.org/10.1007/s12124-015-9320-8>
- Gough, B., & Madill, A. (2012). Subjectivity in psychological science: From problem to prospect. *Psychological Methods*, 17(3), 374–384. <https://doi.org/10.1037/a0029313>
- Haven, T., Grootel, & L, D. (2019). Preregistering qualitative research. *Accountability in Research*, 26(3), 229–244. <https://doi.org/10.1080/08989621.2019.1580147>
- Huma, B., & Joyce, J. (2022). 'one size doesn't fit all': Lessons from interaction analysis on tailoring open science practices to qualitative research. *British Journal of Social Psychology*, 00, 01–15. <https://doi.org/10.1111/bjso.12568>
- Jacobs, A., Büthe, T., Arjona, A., Arriola, L., Bellin, E., Bennett, A., Björkman, L., Bleich, E., Elkins, Z., Fairfield, T., Gaikwad, N., Greitens, S., Hawkesworth, M., Herrera, V., Herrera, Y., Johnson, K., Karakoç, E., Koivu, K., Kreuzer, M., & Yashar, D. (2021). *The Qualitative Transparency Deliberations: Insights and Implications*. *Perspectives on Politics*, 19(1), 171–208. <https://doi.org/10.1017/S1537592720001164>
- Johnson, J., Adkins, D., & Chauvin, S. (2020). A review of the quality indicators of rigor in qualitative

- research. *American Journal of Pharmaceutical Education*, 84(1), 7120. <https://doi.org/10.5688/ajpe7120>
- Macdonald, M., & Carnevale, F. (2008). Qualitative health research and the IRB: Answering the “so what?” with qualitative inquiry. *Journal of Academic Ethics*, 6(1), 1–5. <https://doi.org/10.1007/s10805-007-9051-7>
- Madill, A., & Gough, B. (2008). Qualitative research and its place in psychological science. *Psychological Methods*, 13, 254–271. <https://doi.org/10.1037/a0013220>
- Michell, J. (2003). The quantitative imperative: Positivism, naïve realism and the place of qualitative methods in psychology. *Theory Psychology*, 13, 5–31. <https://doi.org/10.1177/0959354303013001758>
- Morgan, D. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1, 48–76. <https://doi.org/10.1177/2345678906292462>
- Musoba, G., Jacob, S., & Robinson, L. (2015). The institutional review board (IRB) and faculty: Does the IRB challenge faculty professionalism in the social sciences? the qualitative report. <https://doi.org/10.46743/2160-3715/2014.1413>
- Pratt, M., Kaplan, S., & Whittington, R. (2020). Editorial essay: The tumult over transparency: Decoupling transparency from replication in establishing trustworthy qualitative research. *Administrative Science Quarterly*, 65(1), 1–19. <https://doi.org/10.1177/0001839219887663>
- Q.A.A. (2019). Subject Benchmark Statement—Psychology. <https://www.qaa.ac.uk/docs/qaa/subject-benchmark-statements/subject-benchmark-statement-psychology.pdf>
- Richters, J. (2021). Incredible utility: The lost causes and causal debris of psychological science. *Basic and Applied Social Psychology*, 43(6), 366–405. <https://doi.org/10.1080/01973533.2021.1979003>
- Sanbonmatsu, D., Cooley, E., & Posavac, S. (2023). The institutional impact of research challenges and constraints on psychology and other social and behavioral sciences. *New Ideas in Psychology*, 70. <https://doi.org/10.1016/j.newideapsych.2023.101014>
- Schönbrodt, F., Gärtner, A., Frank, M., Gollwitzer, M., Ihle, M., Mischkowski, D., Phan, L., 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>
- Syed, M., & McLean, K. (2022). Disentangling paradigm and method can help bring qualitative research to post-positivist psychology and address the generalizability crisis. *Behavioural Brain Sciences*, 45, 32. <https://doi.org/10.1017/S0140525X21000431>
- Teo, T. (2017). From psychological science to the psychological humanities: Building a general theory of subjectivity. *Review of General Psychology*, 21(4), 281–291. <https://doi.org/10.1037/gpr0000132>
- Tuval-Mashiach, R. (2021). Is replication relevant for qualitative research? *Qualitative Psychology*, 8(3), 365–377. <https://doi.org/10.1037/qup0000217>
- Uher, J. (2023). What’s wrong with rating scales? psychology’s replication and confidence crisis cannot be solved without transparency in data generation. *Social and Personality Psychology Compass*. <https://doi.org/10.1111/spc3.12740>
- Valsiner, J. (2020). From clay feet to new psychology: Starting the move. *Integr. psych. behav.*, 54, 515–520. <https://doi.org/10.1007/s12124-020-09564-x>
- Weitzenboeck, E., Lison, P., Cyndecka, M., & Langford, M. (2022). *The GDPR and unstructured data: Is anonymization possible?* International Data Privacy Law. <https://doi.org/10.1093/idpl/ipac008>