

Research assessment: new panels, new luck?



Published November 6, 2024

Citation

Brembs, B. (2024, November 6). Research assessment: new panels, new luck?. *Bjoern.brembs.blog*. <https://doi.org/10.59350/6mhrb-kd877>

Keywords

Science Politics, DFG, Grants, Journal Rank, Metrrics

Copyright

Copyright © None 2024. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

For 14 years, the main research funding agency in Germany, the German Research Foundation (DFG) has stated in its guidelines that submitted grant proposals will be assessed primarily on the basis of their *content*, rather than counting the applicants' previous publications. However, not all of DFG's panels seem to be on board.

In the so-called "normal application procedure" of the DFG, research grant proposals are evaluated by a study section panel after formal peer-review. This panel then recommends the proposed projects to the main funding committee either unchanged or with a reduced budget or not at all. In times like these, when the number of eligible applications always exceeds the budget, it is not uncommon to find budget cuts even for approved applications. So when one of my own grant proposals (an extension of a previously funded grant) was evaluated recently, I wasn't surprised to find that one of the two doctoral positions I had requested had been cut, rendering the proposed project unfeasible. This wasn't the first time that such cuts had forced us to use the approved funds for a different project.

In this case, however, one aspect was different from all previous similar cases, which irritated me quite a bit. The "Neuroscience" panel, which is responsible for my proposals, provided a total of two sentences to justify the cutback:

"However, we consider the progress of the first funding period to be rather modest. Even taking the pandemic conditions into account, the output of [...] the first funding phase [...] with one publication in Journal X, one preprint [...] and four abstracts can only be considered moderate."

(my translation, [German original here](#))

What is so irritating about these two sentences is the emphasis on "content assessment" in the DFG guidelines for research assessment. The *content* of the two publications, which the panel considers "moderate" in number, represents no less than the answer to a research question that first brought me into the neurosciences as a student – and that I have fought hard to answer for [thirty years now](#). Such a "content assessment" may be irrelevant to the "Neuroscience" panel despite the DFG guidelines, but for our research it was the big breakthrough after three decades of painstaking work.

Apart from the question of why a panel of elected professors is needed to count the publications of an applicant, the DFG has been keenly aware of the problems that arise when using publication metrics (such as counting publications or publication venues) since at least 2010. In the following, it is important to distinguish between the DFG as a collective organization (it is classified as a registered non-profit in Germany) with its head office and employees on the one hand, and its reviewers, study section panels and other committees on the other. The positions of individual committees or members of the DFG do not always necessarily have to correspond to the position of the DFG as an organization. When "the DFG" is mentioned in this post, I'm referring to the collective organization. In all other cases, members or committees are explicitly named.

Pernicious incentives

Fourteen years ago, the DFG changed the rules of its research assessment to limit the use of metrics. At the time, they cut down to ten the maximal number of publications that can be submitted in support of a grant proposal, and this was their justification:

In the course of performance assessment [...] it has become increasingly common to create quantitative indicators based on publication lists and to allow these to replace a content-based assessment of scientific work. This puts a great deal of pressure on scientists to publish as many papers as possible. In addition, it repeatedly leads to cases of scientific misconduct in which incorrect information is provided in the publication list regarding the status of publications. [...] The DFG regrets these developments [...]. However, it sees itself as obliged [...] to emphasize that scientific content should be the deciding factor in evaluations in DFG procedures. The limitation of publication information to a smaller number of publications is associated with the expectation that these will be appropriately evaluated in terms of content during the review and funding decision-making process. [1]

For 14 years now, applicants have therefore only been allowed to list a maximum of ten publications in their CVs when applying to the DFG. The aim here is that scientific and non-numerical criteria should play the decisive role in research assessment. This goal was apparently so important and central to the DFG that these ideas were even incorporated into the DFG's 2019 code of conduct, "Guidelines for Safeguarding Good Scientific Practice" (the "Kodex"):

Performance evaluation is based primarily on qualitative criteria, with quantitative indicators only being included in the overall evaluation in a differentiated and reflective manner. [2]

It was therefore only consequential that the DFG signed the "San Francisco Declaration on Research Assessment" (DORA) in 2021 [3] and, by doing so, committed itself to...

... not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions. [4]

This also emphasizes that when the DFG talks about "quantitative indicators", they are not only referring to just the number of publications, but also to the reputation of the respective publication venues ("impact factors"). Although this had already been implied by the primacy of "scientific content" since 2010, it has now been publicly clarified once again with signing DORA.

The DFG's 2022 position paper "Scientific Publishing as the Basis and Design of Scientific Assessment" [5] doubles down on these concepts:

The central task of the funders – and as such, of course, also of the German Research Foundation – is therefore to ensure that the evaluation of scientific performance is first and foremost based on the scientific content. The reputation of the publication venues and

bibliometric indicators are therefore to be removed from the canon of official evaluation criteria, where they exist, and their practical use is to be minimized.

And to make it absolutely clear once again that this was exactly what the above quote from the 2019 Kodex intended to mean:

A focus on bibliometrically oriented assessment of scientific performance at the level of individuals sets incentives for behavior contrary to the standards of good scientific practice as defined by the DFG Kodex.

After all these developments, it was not surprising that the DFG also became a founding member of CoARA in 2022, which entails the following “Core Commitment”:

Abandon inappropriate uses in research assessment of journal- and publication-based metrics [6]

This multitude of documents and text passages serves to document that the guidelines and efforts of the DFG as an organization wrt research assessment have been very clear and consistent over the last 14 years. One could summarize them as: “It is not in line with our concept of good scientific practice to count the number or reputation of publications or to place them at the center of research assessment.” It needs to be emphasized that this is a solidly evidence-based policy: both the reputation of journals and the number of publications correlate with unreliable science [7]. Thus, this development within the DFG over the last 14+ years did not arise out of some overregulatory zealots putting bureaucracy before science, but out of using the best available evidence in pursuit of the best possible scientific practice.

Study sections not included

The DFG is not alone in these developments. It is working within a pan-European phalanx of research organizations, whose perhaps greatest success to date has been to convince the EU Council of Science Ministers that there now is sufficient evidence to push ahead with a far-reaching reform of the scientific publishing system [8]. We live in times when one cannot praise such consistently evidence-based policy enough. What the DFG has achieved here is groundbreaking and a testament to their scientific excellence. The DFG is thus demonstrating that it is a progressive organization, spearheading good scientific practice, embedded in a solid evidence base and international cooperation. It is therefore completely understandable that applicants now would assume that their performance assessment by the DFG is no longer based on quantitative indicators, but on scientific content.

But it may seem as if the DFG did not quite anticipate the reactions of their panels? Or maybe the opposition to evidence-based research assessment described above was just an exception in a single, extreme study section panel? If an article in the trade magazine “LaborJournal” is anything to go by, such reactionary views may still be widespread among DFG panels. Last year, the DFG panel “Zoology” issued a statement in the LaborJournal that they do not feel bound by the DFG’s evidence-based guidelines [9]:

It is not easy to select the most suitable journal for publishing research results in a changing publishing system. We would like to share some thoughts on this, so that the expectations of decision-making bodies can also be included in the applicants' publication strategy.

Even though it is not explicitly worded as such, all applicants nevertheless immediately understand that the considerations that follow these sentences imply instructions on where to publish in order to meet the expectations of decision-making bodies such as the “Zoology” panel. To avoid any misunderstandings, the “Zoology” panel disambiguates:

Therefore, you should mainly publish your work in journals that enjoy a good reputation in the scientific community!

Everybody knows what the words “good reputation” stand for. After all, several studies have found that precisely this “good reputation” correlates exceedingly well with the impact factor [10] that the DFG has committed itself to not using with DORA/CoARA – and which in turn correlates with unreliable science [7]. And as if to ensure that really everybody gets the message, the “Zoology” panel re-iterates again which journals they are recommending for applicants to maximize their chances of getting funded:

Therefore, publications in selective journals with a high reputation will continue to be an indicator of past performance.

At least the author of these lines is tempted to continue the sentence with: “... no matter what the DFG thinks, decides or signs”.

The panel managed to make it crystal clear to all applicants what they were referring to, without having to mention any red-flag words such as “impact factor” or “h-index” (plausible deniability). The term “dog whistle” was coined for such an approach: coding controversial statements such that the target audience understands exactly what is meant – without provoking opposition from anyone else.

Perhaps not surprisingly, the “Zoology” panel does not shy away from applying their recommendations in their funding decisions. For example, the panel rejected the first grant proposal from an early career researcher (ECR), although it agreed with the two reviewers that the project itself deserved funding. Among the reasons the “Zoology” panel listed for nevertheless rejecting the grant proposal, it cited the ECR’s merely “average publication performance” as the deciding factor, without dealing with the content of the publications and disregarding that the relevant time period not only included the Covid pandemic but also the ECR’s parental leave. It is hard to imagine a more pernicious demonstration that where and how much one publishes still remains an essential funding criterion for this panel, no matter the DFG guidelines.

These examples show that neither the “Neuroscience” nor the “Zoology” DFG panels find anything wrong with setting precisely the incentives the DFG finds so objectionable: “A primarily bibliometrically oriented assessment of scientific performance at the level of individuals sets incentives for behavior contrary to the standards of good scientific practice.” On the contrary, it

appears as if none of the panels have developed any “*mens rea*” when it comes to quite openly – and in one case even publicly – violating long-standing DFG policies.

In science, the principle of [Occam's razor](#) applies: “Entities must not be multiplied beyond necessity”, entailing that, e.g., of two otherwise equivalent hypotheses, the simpler one should always be preferred. Probably somewhat less well known is [Hanlon's razor](#), which similarly requires a decision between hypotheses: “Never attribute to malice that which is adequately explained by incompetence”. Could it be that the panels were simply unaware of the DFG's guidelines and commitments? Is it possible that the last 14 years of developing these guidelines have passed the panels by without a trace? When I approached the DFG in this regard, the DFG employees I interacted with seemed slightly exasperated when they emphasized that of course all panels were thoroughly briefed before they start their work, that these briefings were a long-standing practice at the DFG and that they of course also included the rules for research assessment.

This seems to remove any last doubts: the DFG panels are all familiar with the guidelines and know that it does not correspond to the DFG's concept of good scientific practice to count the number or reputation of publications. What then could possibly motivate some DFG panels to publicly take up a position directly opposite of the DFG's established policies for research assessment? Ultimately, only the individuals on the panels can provide an accurate answer, of course. Until then, we can only speculate, but it would not be the first time that renowned researchers do not take it lightly when they are told what good scientific practice is or that their methods and views are outdated (see, for example, “methodological terrorists”, “research parasites” or “nothing in their heads” [11]).

Slow cultural change?

Following Occam and Hanlon, a straightforward interpretation of the above DFG panel behavior would be that some more reactionary panels simply do not accept that counting publications and reputation is now out of bounds due to DFG rules. If that interpretation were correct, the first round in such a power struggle would have gone to the panels. It seems the DFG is not willing (at least so far) to enforce their guidelines. One of the reasons I was given is that 14 years were too short a time frame and especially the DORA signature were only three years old. The expressed fear was that enforcement at such an early stage of the evaluation reform process could lead to a strong backlash from the panels, which the DFG wants to avoid at all costs.

In principle, the DFG would have ways and means of imposing appropriate sanctions for scientific misconduct by researchers or reviewers. The list of possible consequences is defined in the document “Rules of Procedure for Dealing with Scientific Misconduct” [12]. However, this document does not (yet?) list flouting DFG policies among the list of punishable actions. Perhaps the contents of this list ought to be re-considered? Maybe the fact that the DFG's “Research Integrity Team” did not even consider a ‘preliminary examination’ in this case tells us something about the priority the DFG internally assigns to the reform of research assessment?

As one more reason justifying the lack of enforcement of their guidelines, the DFG's "Scientific Integrity Team" stated that the panels who had made these decisions had just been dismissed after their terms had now ran out and that new panels had just been elected. This would provide the DFG with another opportunity to emphasize this topic during their briefings. Indeed, it was confirmed by members on these newly elected panels that the DFG specifically highlighted the assessment guidelines in the webinar training sessions for the new panels. The DFG is thus not complacent at all, but is opting for a slower, voluntary cultural change instead of effectively enforcing its guidelines.

While one can obviously sympathize with the DFG approach for any number of good reasons, for all applicants it means maximum uncertainty: Do the DFG guidelines apply, or don't they? Should you continue to keep the sample sizes small and thus publish faster – or should you aim for the necessary 80% statistical power after all? Do you make an ugly data point disappear so that the Nature editor also likes the result – or do you publish honest data in the "Journal of Long Titles"? Do you continue to pursue salami slicing – or invest more efforts into making your science reproducible? Do you adjust the p-value downwards and only upload tabular data – or do you implement Open Science with full transparency? And anyway: which of the possible projects you could apply for is the one where you could squeeze the most publications out of?

For ECR applicants without permanent positions, these are all existential questions, and the answers are now completely open again. These ECRs are generally more affected by all negative evaluations than tenured professors – and are therefore particularly vulnerable to this form of uncertainty. For those among them who wanted to rely on the 14 years of DFG practice, including their commitments under DORA/CoARA, the above-mentioned bibliometric assessments and the article by the "Zoology" panel in the Laborjournal must seem like an open mockery of good scientific practice. Maybe the applicants whose grant proposals will be judged by the panels mentioned here ought to take a close look at the lists of "Questionable Research Practices" [13], because although they increase the likelihood of unreliable science, they promise more and higher-ranking publications – as required by these panels.

What message is the DFG sending to the future generation of researchers in Germany when it leaves such obvious non-compliance unanswered – and instead just hopes that the new panels might show a little more understanding for evidence-based policies or at least be a little bit more amenable to DFG webinars? What is the consequence if the goal of a research project is no longer to answer a scientific question, but rather to obtain a maximal number of publications and their highest possible ranking?

There is also the question what the organizations behind DORA and CoARA have to say about all this. When I asked them, both organizations indicated that the panels seemed indeed to have run afoul of DORA/CoARA agreements, but that they did not have the resources to investigate individual cases. Their resources were just enough to promote the reform of research assessment and to take care of members. The review and enforcement of the voluntary commitments had to be taken over by someone else.

Rewarding unreliable science

Everyone knows that voluntary commitments are useless if their violations remain without consequence. The behavior of the DFG panels is just one example of such toothless commitments. How will disillusioned researchers, who may have had high hopes for DORA/CoARA, react if the institutions' voluntary commitments ultimately turn out to be mere signal politics without consequences? The efforts to modernize research assessment, as described above, are based on the evidence that the race to submit more and more publications to the highest-ranking journals rewards unreliable science and punishes reliable science [7]. The elimination of the number of publications and the reputation of the journals from research assessment, the logic goes, would also eliminate significant drivers of unreliable science.

Ultimately, the aim of research assessment reform is to instill in authors and applicants the certainty that reliable science will now be rewarded – regardless of where and how much they publish. However, if the DFG does not soon win the panel lottery, the fear remains that eventually nobody feels bound by any such policies or obligations anymore. The resulting uncertainty among authors and applicants would completely undermine the international efforts to reform research assessment, at least in Germany. In this case, the only risk-averse strategy remaining for authors and applicants were, then as now, to publish as much as possible at the highest possible rank – with all the well-documented consequences.

This is a loosely translated version of my [German article in the Laborjournal](#).

References:

- [1] https://web.archive.org/web/20100304074136/http://www.dfg.de:80/foerderung/info_wissenschaft/info_wissenschaft_10_11/index.html
- [2] <https://www.dfg.de/resource/blob/173732/4166759430af8dc2256f0fa54e009f03/kodex-gwp-data.pdf>
- [3] <https://twitter.com/dorassessment/status/1402273251502571531>
- [4] <https://sfdora.org/read/>
- [5] <https://www.dfg.de/resource/blob/175770/5772e3980d4e81991dc716f94fbc2382/positionspapier-publikationswesen-data.pdf>
- [6] <https://coara.eu/news/reforming-research-assessment-the-agreement-is-now-open-for-signature/>
- [7] <https://www.frontiersin.org/articles/10.3389/fnhum.201800037/full>
<https://www.frontiersin.org/articles/10.3389/fnhum.2013.00291/full>
<https://royalsocietypublishing.org/doi/10.1098/rsos.160384>

[8] <https://www.consilium.europa.eu/en/press/press-releases/2023/05/23/council-calls-for-transparent-equitable-and-open-access-to-scholarly-publications/>

<https://www.dfg.de/de/service/presse/pressemitteilungen/2023/pressemitteilung-nr-16>

<https://www.coalition-s.org/wp-content/uploads/2023/05/JointResponse2CouncilScholCommConclusions.pdf>

[9] https://www.laborjournal.de/rubric/hintergrund/hg/hg_23_10_03.php

[10] <https://doi.org/10.1002/asi.4630330109>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC141186>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1773051>

<https://doi.org/10.1007/s11192-007-1924-4>

[11] <https://www.businessinsider.com/susan-fiske-methodological-terrorism-2016-9?op=1>

<https://www.thecut.com/2016/10/inside-psychologys-methodological-terrorism-debate.html>

<https://www.science.org/content/blog-post/attack-research-parasites>

<https://www.nationalgeographic.com/science/article/failed-replication-bargh-psychology-study-doyen>

<https://replicationindex.com/wp-content/uploads/2020/07/bargh-nothingintheirheads.pdf>

[12] <https://www.dfg.de/de/formulare-80-01-246936>

[13] <https://doi.org/10.1093/oso/9780190938550.003.0010>

<https://doi.org/10.1177/1084822320934468>

<https://www.aje.com/arc/questionable-research-practices>

<https://libguides.uvt.nl/researchintegrity/questionable-research>