FABBS Response to Request for Information (RFI) on Proposed Changes to Reporting Outcomes from NIH Peer Review (NOT-OD-26-088)

Submitted online on October 13, 2026

Please provide input on proposed changes to reporting outcomes from NIH peer review (response box has 5000 character limit - can also submit PDF).

The Federation of Associations in Behavioral and Brain Sciences (FABBS) represents 35 of the nation's leading scientific societies in the psychological, cognitive, and behavioral sciences. We advance the sciences of mind, brain, and behavior; promote evidence-based policymaking; and bolster the rigor and impact of the federal scientific enterprise. With support from federal agencies, including NIH, FABBS scientists study critical health challenges such as maternal mortality, mental health conditions, and chronic disease. Our members also serve as peer reviewers in NIH's grantmaking process. We value the opportunity to comment on NIH's proposed changes to reporting outcomes from NIH peer review.

FABBS agrees that peer review information should be considered in its entirety when making funding decisions, along with other factors such as institute priorities and research portfolios. According to our members who have had grant proposals reviewed, that has long been the NIH process. It is unclear how this proposal will help achieve this goal; instead, it raises several concerns about the transparency, integrity, and value of the peer review process, if implemented. Additionally, FABBS is concerned that withholding the working group's analysis that led to the proposed changes as well as failing to include a plan to evaluate the new model lack transparency, raise questions, and undermine trust in NIH – a stated goal of NIH leadership. We expand on these concerns below.

(1) Categorization of impact scores decreases their usefulness and reduces precision, impacting how investigators move forward with their grant applications. The proposed change adds no new information to the review process while eliminating an important data point.

Although FABBS agrees with NIH that numerical peer review scores must be considered in context – e.g., they may not always distinguish between two closely ranked applications – they still provide useful information to researchers, institutions, and NIH staff. And even if imperfect, the solution is not to do away with them entirely while introducing even more uncertainty and imprecision. As currently proposed, the three categories are far too broad and risk grouping together grant proposals that are, in reality, extremely different in terms of quality. This is particularly problematic in the second category, "competitive," based on 26-50 percent of final overall impact scores. This category would include proposals that just barely miss out on

being "most competitive" as well as those that are only marginally better than "not discussed." Despite being in the same review category, one project could only need minor revisions to make the categorical jump while another needs a fundamental overhaul. Given mixed or even relatively poor qualitative feedback, an investigator may still choose to revise and resubmit if told their grant is "competitive," regardless of their actual prospects of success. As a result, most unfunded applicants in the top two categories will likely resubmit, potentially diverting their resources from more viable projects as well as increasing the load on volunteer reviewers and NIH staff.

Grant applicants respond to institute priorities and funding availability without any opportunity for input. By contrast, upon receiving their peer review score they have the ability to adjust and respond. Therefore, having maximum information enables them to best revise their proposal accordingly (or choose not to revise at all, and devote time and resources to more viable projects). The impact score is the only piece of data that reflects the judgment of the entire review panel, as qualitative feedback comes from just three reviewers, which can vary considerably and can give conflicting impressions of the project's viability. More detailed information on scientific merit is especially helpful for early-career investigators and their mentors, who can guide grant proposal revision more effectively if they know exactly where the proposal stands in comparison to peers'.

Impact scores are similarly useful to second-level reviewers, in combination with summary statements, providing information about the collective judgment of a larger group of experts as well as more detailed feedback from a smaller group.

(2) The proposal to categorize peer review impact scores – in reports to researchers, institutions, NIH staff, Advisory Councils (ACs), and agency leadership – reduces transparency and weakens accountability in the grantmaking process. This is especially problematic at a time when political interference in research funding – which Congress has clearly rejected – threatens our scientific ecosystem.

A stated goal of NIH's Unified Funding Strategy (UFS) is to prioritize scientific merit, with institutes, centers, and offices (ICOs) directed to consider peer review information in its entirety. The decision to categorize and hide impact scores directly contradicts this and reduces transparency in the funding process, making it more difficult to know how the agency reached a funding determination.

Making evidence of scientific merit readily available forces ICOs to be accountable for their funding decisions – if a highly scored proposal is not funded, the ICO must clearly explain the other factors that led to that decision. ICOs must similarly justify funding projects with (relatively) poorer merit scores. Maintaining the reporting of impact scores within the UFS can improve communication around funding decisions, which could help restore trust between the agency and stakeholders. In contrast, withholding the score from the public obscures deviations

between scientific merit and funding status, which then can go unexplained. Concealing scores also prevents accurate evaluation of funding decisions.

Hiding impact scores also risks inviting the perception that NIH does not value peer review as highly as it claims. Reviewers may believe that their scores do not matter if their hard work is going to be reduced to a category that obscures meaningful distinctions between proposals. What purpose does the score provided by the full review panel serve if second-level review will meaningfully consider only qualitative feedback from just three members? Volunteer reviewers may be less likely to make the significant commitment to serve on study sections if their work is not properly valued.

(3) The NIH working group reportedly collected data and undertook a process that led to these recommendations. Unfortunately, NIH has yet to share the data and deliberations, providing few details about the process that produced the proposal. Stakeholders cannot properly evaluate the rationale for the changes.

We have fundamental concerns about the charge of the working group (WG), i.e., "recommending changes," suggesting that NIH had already determined to revise how it reports out peer review evaluations before the WG even started reviewing evidence. Typically, such a group would instead set out to evaluate the current model, i.e., what is and is not working, and go from there; instead, it seems that NIH already determined that current peer review reporting was a problem.

The RFI provides little information about the process that led to the decision to modify the first-level peer review process and change the outcomes reported to researchers, institutions, and NIH staff. The agency reportedly convened a WG of NIH leaders who were "charged with recommending changes in how NIH reports out peer review evaluations of scientific merit, to facilitate the Unified Funding Strategy." The RFI tells us that "the group reviewed literature on the predictive value of peer review scores for scientific outcomes, considered reporting approaches utilized by other funding entities, and examined NIH peer review data." The RFI provides no additional information about who these leaders were; what literature they reviewed; what considerations were made when comparing reporting approaches between agencies; and what data were analyzed and how.

As a result, external stakeholders cannot effectively evaluate the proposal and the few justifications supplied. For example:

- The RFI claims that reporting categories rather than impact score "aligns with the Unified Funding Strategy" but there is no explanation for why scores do not also align. It is possible that the WG considered this, but we cannot know without a public report.
- Similarly, the RFI claims that "removal of final overall impact scores also supports a focus on substantive comments from the assigned reviewers and the panel, elevating the role of scientific peer review" – once again, we cannot evaluate the accuracy of this claim. Did the WG review evidence that program officers (POs) and other NIH staff are

currently *not* considering the qualitative feedback? What is the evidence that POs regularly over-rely on impact scores? Prior to the UFS, not all ICOs relied strictly on paylines in their grantmaking (e.g., NIA has always considered factors such as institute priorities and portfolio gaps). Without clear evidence to the contrary, this seems like a non-issue that the proposal does not actually address.

- The comparisons made with other agencies' reporting approaches are also questionable due to the differences in grantmaking policies and application pools – for example, NIH handles tens of thousands more applications than NSF in a given cycle.
- Additionally, even if we accept that categorizing impact scores is appropriate in the grantmaking process, no explanation is provided for why the WG recommended the three categories it did. Will categories be computed within study sections, within review cycles, or across the agency?

FABBS encourages NIH to release a report of the WG's analysis and recommendations to mitigate concerns about its process and allow for a meaningful evaluation of the evidence on which it based these proposed changes. Based on a review of other NIH working groups, sharing these reports is standard practice.

(4) The proposal does not include an evaluation plan to determine if the new reporting model is an improvement over the current model.

The implementation of new policies requires evaluations to assess whether changes are achieving intended goals or having unintended consequences that outweigh the benefits. Nothing in the proposal indicates that NIH has a plan to evaluate this change to peer review outcome reporting. FABBS recommends that NIH consider the following metrics:

- Stakeholder (i.e., applicant) evaluations of the usefulness of the current report (including impact scores) versus the new report (excluding impact scores)
- Resubmission decisions – increased or decreased under the new model?
- Funding outcomes – how many proposals in the "competitive" category are being funded compared to the "most competitive" category?
- Administrative and operational burden for POs and other NIH staff – increased or decreased under the new model?
- Quality of peer reviewers' qualitative feedback – any changes under the new model?

FABBS urges NIH to carefully consider how it will evaluate this new reporting model before it is implemented, with clear benchmarks for success and failure relative to the current model. NIH might also consider a pilot program to test the new model before implementing it across the entirety of the agency.