



**AMERICAN
PSYCHOLOGICAL
ASSOCIATION**

August 14, 2026

Lyric Jorgenson, PhD
NIH Associate Director for Science Policy
NIH Office of Science Policy
6705 Rockledge Drive
Suite 630, MSC-7985
Bethesda, MD 20892-7985

**Re: Request for Information on Measuring and Rewarding Scientific Impact
(Notice Number: NOT-OD-26-087)**

Via email: SciencePolicy@od.nih.gov

Dear Dr. Jorgenson:

The American Psychological Association (APA) appreciates the opportunity to provide input to the National Institutes of Health (NIH) regarding how scientific impact should be measured, recognized, and incentivized in modern biomedical and behavioral research via the Request for Information (RFI): Measuring and Rewarding Scientific Impact, Notice Number: [NOT-OD-26-087](#). The APA is the nation's largest scientific and professional nonprofit organization representing the discipline and profession of psychology, as well as 190,000 members and affiliates who are clinicians, researchers, educators, consultants, and students in psychological science.

APA welcomes NIH's efforts to update how scientific impact is measured and rewarded. Research assessment should reflect the practices that support credible science, including rigorous methods, transparency, collaboration, mentorship, and public benefit. As behavioral, cognitive, developmental, mental health, and social sciences research continue to inform health and well-being, NIH has an opportunity to reward not only productivity, but quality, credibility, openness, mentorship, collaboration, and real-world value.

Academia's reliance on quantitative indicators, including publication counts, journal impact factors, and citation metrics, creates incentives that can work against research quality. Assessment systems should move from asking how much research is produced to asking how well research is conducted. APA supports NIH's interests in responsible research assessment that prioritizes research quality, rigor, transparency, credibility, and societal value over simple measures of productivity. Psychology has been a leader in developing methods that support this shift. Assessment and incentive systems should recognize contributions that strengthen the credibility and usefulness of science, including transparent reporting, data and code sharing, preregistration where appropriate, replication efforts, methodological

APA.ORG
APASERVICES.ORG

Advocating for APA members and psychology

750 First Street, NE
Washington, DC 20002-4242

202.336.5800
202.336.6123 TDD

innovation, team science, and shared research resources. Several initiatives currently exist that aim to address this challenge.

For behavioral and social science research, quality indicators should recognize contributions that improve credibility, reproducibility, and cumulative nature of scientific knowledge, including the development of validated measures, replication studies, open research resources, methodological advances, and efforts that improve research transparency and accessibility. Psychological science has been at the forefront of developing and evaluating approaches to replication, reproducibility, research transparency, and credibility assessment, and these advances demonstrate that scientific quality can be measured in ways that are more closely aligned with NIH's mission than traditional publication-based metrics alone.¹

Accordingly, APA recommends that NIH take a multidimensional approach to scientific impact that balances traditional indicators, such as publications and citations, with evidence of rigor, reproducibility, responsible data and software sharing, effective training and mentorship, meaningful collaboration, translation, and contributions to public trust. No single metric can capture the full value of scientific work. NIH should therefore avoid replacing one narrow set of incentives with another and instead develop flexible, field sensitive approaches that recognize diverse forms of contributions across career stages, institution types, research methods, and communities served.

Systematizing Confidence in Open Research and Evidence (SCORE) Credibility Assessment

APA recommends that NIH recognize open science practices as meaningful contributions to scientific impact. Sharing data, code, software, research materials, protocols, measures, models, and analytic plans can accelerate discovery, enable independent verification, reduce duplication, and increase the public return on federal research investments.²

The Score initiative is one of the largest efforts ever to evaluate the credibility of research findings in social and behavioral science. The project was developed through a DARPA-funded collaboration involving the Center for Open Science, academic researchers, and technology partners. The SCORE program underscores that scientific trustworthiness does not require certainty. It depends on the research enterprise's capacity to reduce uncertainty through credible evidence, transparent critique, and ongoing assessment. Common proxies for credibility, such as peer-reviewed publication, journal prestige, and citation counts are easy to measure but weak indicators of research quality because they do not examine the research itself. According to Nosek, B. A. et al, credibility assessment should instead ask whether the findings deserve serious consideration and whether the evidence is strong enough to guide future research, resource allocation, policy, practice, or translation.³

¹ Gärtner A, Leising D, Schönbrodt FD (2024) Towards responsible research assessment: How to reward research quality. PLoS Biol 22(2): e3002553. <https://doi.org/10.1371/journal.pbio.3002553>

² Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422-1425. <https://doi.org/10.1126/science.aab2374>

³ Nosek, B. A., et al. (2026). *Reimagining and diversifying assessment of the credibility of research findings* (abjur_v1). MetaArXiv. https://osf.io/preprints/metaarxiv/abjur_v1

SCORE tested whether human judgements and automated methods could predict replication outcomes and provide explainable assessments. The results show modes of progress but also indicate that neither human nor machine predictions are yet accurate enough to replace replication studies. For NIH, this shows that credibility indicators require sustained investment, validation, transparency, and caution before they are used to guide incentives or decision-making.

According to Nosek, B. A., et al., replicability is only part of a broader credibility landscape that includes transparent peer evaluation, engagement with prior knowledge, bias control, measurement precision and reliability, accountability, and whether research claims are proportionate to the evidence. Emerging modes such as Lifecycle Journal, an open-science scholarly publishing platform designed to evaluate research across the entire research lifecycle, allow multiple independent review services to assess discrete features of credibility, such as code execution, data sharing quality, consistency between research plans and outcomes, reporting standards, and statistical errors. A more diversified assessment system could complement peer review and help researchers, reviewers, policymakers, journalists, and the public evaluate research quality more effectively.⁴

Core Indicators of Scientific Impact

APA supports NIH's efforts to recognize rigor, reproducibility, replicability, robustness, and credibility as important dimensions of scientific impact. Psychological science has contributed significantly to the study of research credibility and has demonstrated that confidence in findings can be strengthened through complementary approaches such as transparent reporting, independent replication, reproducibility checks, robustness analyses, preregistration, and structured expert assessment.

Drawing on the literature, APA recommends that NIH distinguish between reproducibility, robustness, replicability, and trustworthiness when developing assessment frameworks. This distinction is important for implementation.⁵ Treating these concepts as interchangeable may create confusion and unintended incentives. In addition, different research questions may require different forms of assessment. Assessment systems should encourage investigators and reviewers to identify the most appropriate methods for evaluating confidence in findings and reward contributions that make research more transparent, interpretable, and verifiable.⁶

Nosek, B. A., et al., propose a concise glossary to improve precision and communication about research quality and evidence assessment.⁷ The glossary is directly relevant because assessment systems can create unclear incentives when key concepts are treated as interchangeable. Accordingly, APA recommends that NIH use precise terminology in application instructions, review criteria, reviewer

⁴ Nosek, B. A., et al. (2026). *Reimagining and diversifying assessment of the credibility of research findings* (abjur_v1). MetaArXiv. https://osf.io/preprints/metaarxiv/abjur_v1

⁵ Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73, 719-748. <https://doi.org/10.1146/annurev-psych-020821-114157>

⁶ Nosek, B. A., et al. (2025). A brief glossary of terms about repeatability: Replicability, robustness, and reproducibility (mqfp4_v1). MetaArXiv. https://doi.org/10.31222/osf.io/mqfp4_v1

⁷ Nosek et al. (2025). A Brief Glossary of Terms about Repeatability: Replicability, Robustness, and Reproducibility (No. mqfp4_v1). *MetaArXiv*. https://doi.org/10.31222/osf.io/mqfp4_v1

guidance, progress reports, and renewal decisions so that investigators and reviewers understand whether they are assessing reproducibility of analyses, robustness of conclusions across defensible analytic choices, replicability with independent data, or broader credibility and trustworthiness.

APA recommends that NIH invest in the science of credibility assessment itself. Nosek and colleagues' work on replicability, robustness, and reproducibility demonstrates that scientific impact depends not only on producing new findings, but also determining whether findings are reliable, interpretable, reusable, and capable of informing cumulative knowledge. SCORE shows that scalable indicators of replicability, robustness, reproducibility, transparency, and other credibility dimensions must be validated before they are used as incentives. NIH should support research, infrastructure, and pilot programs that test whether proposed indicators are reliable, explainable, field-sensitive, and useful to reviewers and decision-makers. Vazire's work reinforces this point by showing that quality uncertainty erodes trust when consumers of science lack the information needed to evaluate claims, and that the credibility revolution may reduce apparent individual productivity while increasing genuine scientific progress.^{8 9} Policies for measuring and rewarding impact should be evaluated empirically, implemented flexibly, and designed to encourage ongoing scrutiny of whether new metrics actually improve science.

APA recommends that NIH incentivize practices that increase confidence in findings, including well-powered designs, transparent analytic decisions, preregistration where appropriate, clear reporting of methods and limitations, independent replication, and evaluation of whether findings hold across analytic choices, populations, measures, and contexts. These practices are especially important in fields that inform clinical care, public health, education, and policy, where premature or overstated claims have significant downstream consequences, despite the frequent need for rapid decision-making and response.

At the same time, APA encourages NIH to implement rigor and reproducibility indicators in ways that are appropriate. Not all research questions are suited to the same forms of replication or preregistration, and exploratory, qualitative, community-engaged, longitudinal, and observational research may require different standards of transparency and robustness. NIH should support discipline-specific guidance and avoid rigid requirements that could unintentionally discourage innovation, exploratory discovery, and more research with hard-to-reach populations.

APA recommends that NIH incorporate these indicators into grant review, progress reporting, and career advancement mechanisms through narrative contribution statements and portfolio-based assessments that allow investigators to show the significance of their most meaningful work. In psychology and behavioral science, especially valuable contributions include demonstrating that findings replicate across diverse populations and settings, developing culturally valid measures, creating reusable research resources, advancing methods for studying behavior and mental health, and generating evidence that informs clinical practice, prevention, education, workplace well-being, and public policy.

⁸ Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra: Psychology*, 3(1), Article 1. <https://doi.org/10.1525/collabra.74>

⁹ Vazire, S. (2018). Implications of the Credibility Revolution for Productivity, Creativity, and Progress. *Perspectives on Psychological Science*, 13(4), 411-417. <https://doi.org/10.1177/1745691617751884>

The association further recommends that NIH also recognize contributions that strengthen the research enterprise itself, such as leadership in multi-site collaborations, training the next generation of scientists, and advancing open and transparent research practices. Rewarding the practices that produce trustworthy, reproducible, and actionable knowledge would strengthen stewardship of public research investments, accelerate scientific progress, and enhance public trust in biomedical and behavioral research.

Psychology-specific research indicators include:

- Independent replications of influential findings and interventions.
- Evidence that findings are robust across populations, contexts, measures, and analytic approaches.
- Preregistration and registered reports, when appropriate to research design.
- Transparent reporting of methods, limitations, analytic decisions, and effect sizes.
- Development and use of rigorous standards for causal inference in behavioral and social science research.
- Clear differentiation among reproducibility of analyses, replicability of findings, and robustness across analytic choices, samples, measures, and contexts.
- Use of repeatability as an umbrella term while separately assessing reproducibility, robustness, and replicability.
- Validated, explainable credibility indicators that assess research features directly rather than relying only on journal prestige, citation counts, or publication status.

Reward Open, Transparent, and Reusable Research Outputs

APA supports NIH's interest in recognizing open and reusable research outputs as meaningful scientific contributions. Sharing data, code, software, measures, protocols, materials, and analytic workflows increase the value of public research investments by enabling verification, reuse, collaboration, and cumulative knowledge building.

NIH should evaluate both the availability and usability of these resources, including documentation quality, accessibility, responsible governance, and evidence of reuse by others. At the same time, sharing requirements should remain flexible enough to accommodate privacy concerns, Tribal sovereignty, community governance, consent limitations, and other ethical obligations.

Consistent with the Transparency and Openness Promotion (TOP) framework, a set of guidelines developed by the Center for Open Science and a coalition of journal editors, researchers, funders, and scientific organizations to improve transparency, credibility, and reproducibility of research, NIH should adopt a tiered approach that encourages transparency while recognizing the differences across research designs and disciplines. Tiered standards would allow NIH to recognize meaningful progress toward transparency across diverse research designs, including experimental, observational, qualitative, longitudinal, clinical, and community-engaged research. This approach would also help ensure that data, code materials, protocols, and analysis plans are shared when appropriate, while preserving necessary protections for privacy, consent, Tribal sovereignty, community governance, and sensitive human subjects data.

NIH should also treat open research outputs as substantive scholarly contributions. Investigators should be able to document the quality, accessibility, reuse, and scientific value of shared datasets, measures,

protocols, software, code, analytic workflows, and other research materials in bio-sketches, progress reports, grant renewals, and career-stage assessments. Recognizing these outputs would align individual incentives with the public value of research that is verifiable, reusable, and cumulative.

Psychology-specific research indicators include:

- Public sharing of validated measures, intervention manuals, study protocols, code, and datasets where appropriate.
- Development of research infrastructure, repositories, and community resources that benefit the field.
- Demonstrated reuse of shared resources by other investigators.
- Use of tiered transparency standards that recognize appropriate sharing of data, code, materials, protocols, analysis plans, and documentation across different research designs.

Value Quality Over Quantity

As noted above, APA recommends that NIH reduce reliance on publication counts, journal prestige, and citation-based indicators when evaluating scientific impact. Although such measures can provide useful contextual information, they should not serve as the only proxy for scientific value.

Assessment systems should instead emphasize the quality, credibility, transparency, rigor, and cumulative contribution of scientific work. Activities such as replication, correction of errors, development of methods and measures, open science practices, development of community resources, and meaningful translation often create substantial scientific value when they do not generate large numbers of publications.

Narrative contributions statements and portfolio-based assessment approaches may provide a more accurate picture of scientific impact by allowing investigators to describe their most significant contributions rather than simply listing outputs.¹⁰

APA recommends that NIH also protect creativity and foundational exploration while strengthening rigor. Assessment frameworks should recognize exploratory and high-risk research when investigators are transparent about uncertainty, clearly distinguish exploratory from confirmatory analysis, and describe how early-stage findings will be tested, replicated, or refined. This would help avoid a narrow compliance culture in which researchers pursue only low-risk, easily preregistered questions, while still advancing NIH's goals of rigor, reproducibility, and public trust.¹¹

NIH should reward research practices that make scientific claims easier to evaluate. Impact indicators should recognize transparent disclosure of study design, measures, sampling decisions, exclusions, analytic flexibility, limitations, and departures from preregistered plans when applicable. These practices would help reviewers, other scientists, policymakers, and the public distinguish well-supported claims from more tentative findings and would strengthen trust in NIH-supported science.¹²

¹⁰ Vazire, S. (2018). Implications of the Credibility Revolution for Productivity, Creativity, and Progress. *Perspectives on Psychological Science*, 13(4), 411-417. <https://doi.org/10.1177/1745691617751884>

¹¹ Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra: Psychology*, 3(1), Article 1. <https://doi.org/10.1525/collabra.74>

¹² Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra: Psychology*, 3(1), Article 1. <https://doi.org/10.1525/collabra.74>

Recognize Mentorship, Training, and Team Science as Scientific Contributions

APA strongly supports NIH's inclusion of mentorship, training, and team science as dimensions of scientific impact. These activities contribute directly to the integrity, sustainability, and the effectiveness of the biomedical and behavioral research enterprise but are often undervalued in publication-based assessments.

NIH should recognize evidence of effective mentorship, trainee development, inclusive research environments, methodological and ethical training, collaborative leadership, shared infrastructure development, and contributions to large interdisciplinary projects. Evaluation criteria should remain flexible enough to accommodate differences in career stage, institutional context, and research settings.¹³

Psychology-specific research indicators include:

- Development of rigorous methodological and ethical training programs.
- Contributions to building and sustaining the behavioral and social science workforce.
- Successful mentorship of trainees and early-career investigators.

Public Impact Should Include Health, Trust, Accessibility, and Community Benefit

APA strongly agrees that scientific impact should ultimately be connected to NIH's mission to generate knowledge that enhances health, lengthens life, and reduces illness and disability. Public impact should be understood broadly to improvements in health outcomes, clinical practice, prevention, education, public policy, community well-being, accessibility of scientific knowledge, and public trust in science.

For psychological and behavioral science, public impact may include validated interventions, improved mental and behavioral health services, better measurement tools, evidence-based prevention strategies, reduced stigma, improved adherence to care, enhanced learning and development, strong community resilience, and more effective health communication. NIH should recognize that translation may occur through clinical guidelines, public health programs, community partnerships, education, workforce training, policy implementation, and dissemination of accessible evidence—not only through commercialization or patents. It often requires sustained engagement, adaptation, partnership, and communication beyond publication.

Implementation Should Minimize Unintended Consequences

APA encourages NIH to implement new impact measures cautiously and evaluate them regularly for unintended consequences. Indicators should be piloted across disciplines, institution types, and career stages before broad adoption. NIH should monitor whether proposed measures improve the assessment of rigor, transparency, and scientific value without creating excessive administrative burden, discouraging innovation, or disadvantaging researchers working in resource-constrained settings or with sensitive data.

¹³ Gärtner, A., Leising, D., & Schönbrodt, F. D. (2024). Towards responsible research assessment: How to reward research quality. *PLoS Biology*, 22(2), e3002553. <https://doi.org/10.1371/journal.pbio.3002553>

Successful implementation will require reviewer guidance, stakeholder feedback, and investments in the infrastructure needed to support replication, data stewardship, software maintenance, training, and methodological support.¹⁴

Recommended Indicators and Incentives

APA urges NIH to elevate rigor, transparency, collaboration, and public impact by operationalizing new incentives through application instructions, review criteria, reviewer guidance, progress reports, renewals, bio-sketches, final reports, and career-stage assessments.

Responsible Research Assessment

- Reduce reliance on publication counts and journal-level metrics.
- Use narrative contributions statements and portfolio-based evaluations emphasizing rigor, transparency, and cumulative value.

Rigor and Credibility

- Reward reproducibility, robustness analyses, replication, transparent reporting, appropriate preregistration, adequate methods and analyses,
- and evidence of generalizability.

Open and Reusable Research

- Recognize sharing of data, code, measures, software, protocols, and research materials, with appropriate privacy and governance protections.

Mentorship and Team Science

- Recognize mentorship quality, trainee and workforce development, methodological and ethical training, inclusive research environments, collaborative leadership, shared infrastructure, and interdisciplinary teamwork.

Innovation and Translation

- Support credible high-risk research and recognize contributions to policy, practice, implementation, and public health impact.

Evaluation and Continuous Improvement

- Pilot, validate, and periodically reassess indicators to ensure they are reliable, fair, field sensitive, and resistant to unintended consequences.

APA commends NIH for seeking input on how to better measure and reward scientific impact. The biomedical and behavioral research enterprise depends on incentives that support credible, transparent, collaborative, and socially beneficial science. By broadening assessment beyond narrow productivity measures and recognizing rigor, reproducibility, openness, mentorship, collaboration, and public benefit, NIH can strengthen the reliability of scientific findings, improve stewardship of public research investments, and reinforce public trust in science. Should you have any questions, please contact Angela L. Sharpe, MG, Senior Director, Congressional & Federal Relations at asharpe@apa.org.

¹⁴ Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422-1425. <https://doi.org/10.1126/science.aab2374>

Sincerely,



Adrienne Stith, PhD
Executive Lead Psychologist, Science
American Psychological Association

References

Abatayo, A. L., et al. (2026). Assessments of credibility in the social and behavioural sciences (7u58q_v1). MetaArXiv. https://osf.io/preprints/metaarxiv/7u58q_v1

Aczel, B., Szaszi, B., Clelland, H. T., et al. (2026). Investigating the analytical robustness of the social and behavioural sciences. *Nature*, 652(8108), 135-142. <https://doi.org/10.1038/s41586-025-09844-9>

Alipourfard, N., et al. (2021). *Systematizing Confidence in Open Research and Evidence (SCORE)* (46mnb_v1). SocArXiv. <https://doi.org/10.31235/osf.io/46mnb>

Miske, O., Abatayo, A. L., Daley, M., et al. (2026). Investigating the reproducibility of the social and behavioural sciences. *Nature*, 652(8108), 126-134. <https://doi.org/10.1038/s41586-026-10203-5>

National Academies of Sciences, Engineering, and Medicine. 2025. *The Science and Practice of Team Science*. Washington, DC: The National Academies Press DOI: <https://doi.org/10.17226/29043>.

Nosek, B. A., et al. (2026). *Reimagining and diversifying assessment of the credibility of research findings* (abjur_v1). MetaArXiv. https://osf.io/preprints/metaarxiv/abjur_v1

Tyner, A. H., Abatayo, A. L., Daley, M., et al. (2026). Investigating the replicability of the social and behavioural sciences. *Nature*, 652(8108), 143-150. <https://doi.org/10.1038/s41586-025-10078-y>