



Center for
Scientific Review

CSR Advisory Council Reviewer Evaluation Working Group Update

Lynn M. Yee, MD, MPH

Thomas J. Watkins Memorial Professor
of Obstetrics and Gynecology

Northwestern University Feinberg
School of Medicine

Delia Olufokunbi Sam, Ph.D.

DNDA Division Director
September 23, 2024



Acknowledgment: Reviewer Evaluation Working Group

CSR Advisory Council



Karen Anderson, M.D. Ph.D.
Arizona State University



Anton Bennett, Ph.D.
Yale University



Matthew Carpenter, Ph.D.
Medical University of South
Carolina



Narasimhan Rajaram, Ph.D.
University of Arkansas at
Fayetteville



Lynn Yee, M.D. M.P.H.
Northwestern University
Co-Chair

Ad-hoc



Catherine Fox, Ph.D.
University of Wisconsin–Madison



James Moody, Ph.D.
Duke University



Daniel Woo, M.D.
University of Cincinnati

NIH Staff



Stephanie Constant, Ph.D.
OD



Delia Olufokunbi Sam, Ph.D.
CSR
Co-Chair



Miriam Mintzer, Ph.D.
CSR



Elena Smirnova, Ph.D.
CSR

OUR CHARGE

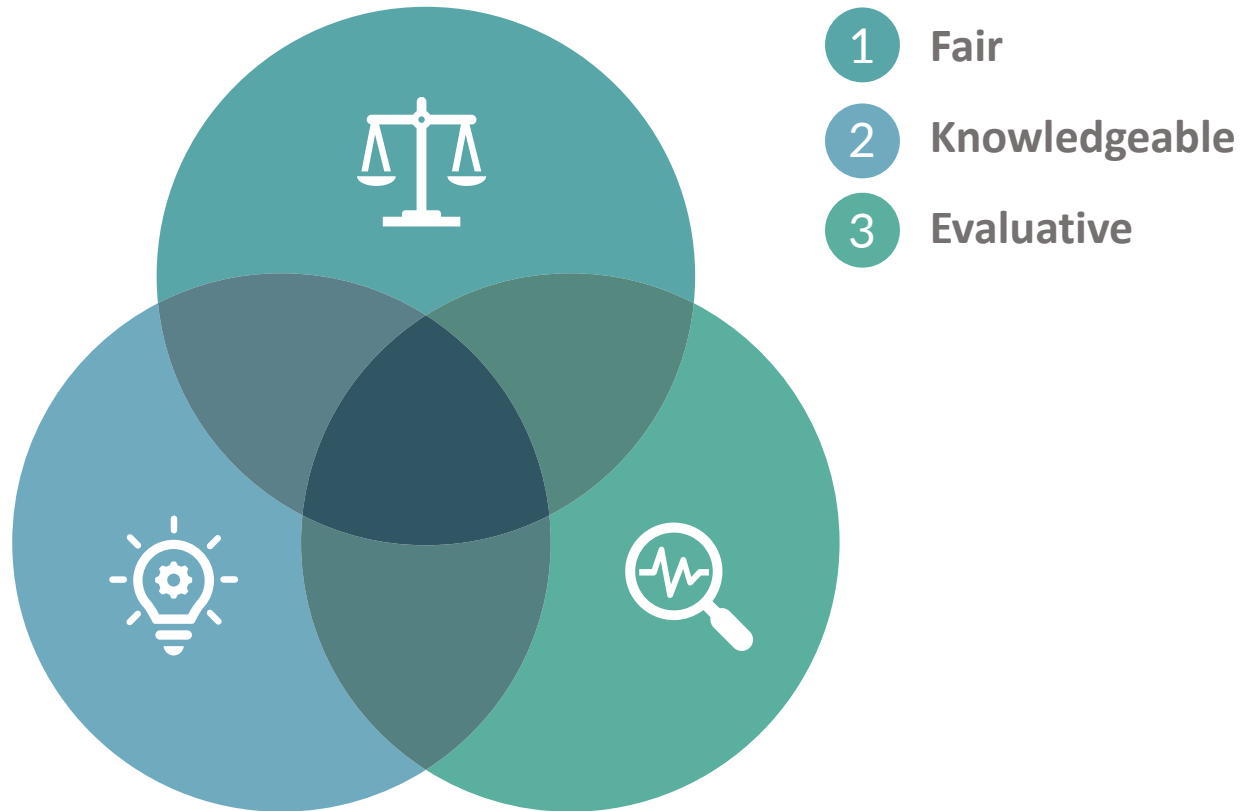
Develop a practical framework for evaluating reviewer performance – one that allows for a uniform, structured, consistent, and transparent approach to performance assessment and improvement.

Guiding Principles

- Key words – **efficient, effective, and transparent**
- Should be tied to reviewer training and reviewer recognition – but focused on evaluation.
- Goal is to consider the characteristics of a strong review
- Not designed to compare reviewers to each other
- Not designed to penalize reviewers – focus is on helping reviewers improve
- Framework should supplement (and incorporate) what SROs are currently doing.
- Use conceptual framework developed by internal committee as a starting point to develop an actionable, measurable evaluation process

CSR Internal Evaluation Committee Conceptual Framework

Domains that characterize a reviewer's performance



The mission of the Center for Scientific Review (CSR) is to see that NIH grant applications receive

fair, independent, expert, and timely scientific reviews — free from inappropriate influences —

so NIH can fund the most promising research.

Reviewer Evaluation Framework

Fair evaluation of Scientific Merit

Provides Evaluative/Critical Assessments

Scores reflect original evaluative judgments and are appropriately weighted to strengths and weakness

Clear and concise presentations focused on score-driving strengths and weaknesses

Critiques are substantive, specific, and provide sufficient rationale to discern evaluation

Adheres to Established Review Guidelines

Adheres to scoring guidelines; consistency in scoring across all applications

Makes impartial, objective assessment, only using review criteria

Assigned reviewers preliminary scores align with oral comments

Critique addresses all core criteria and review guidelines

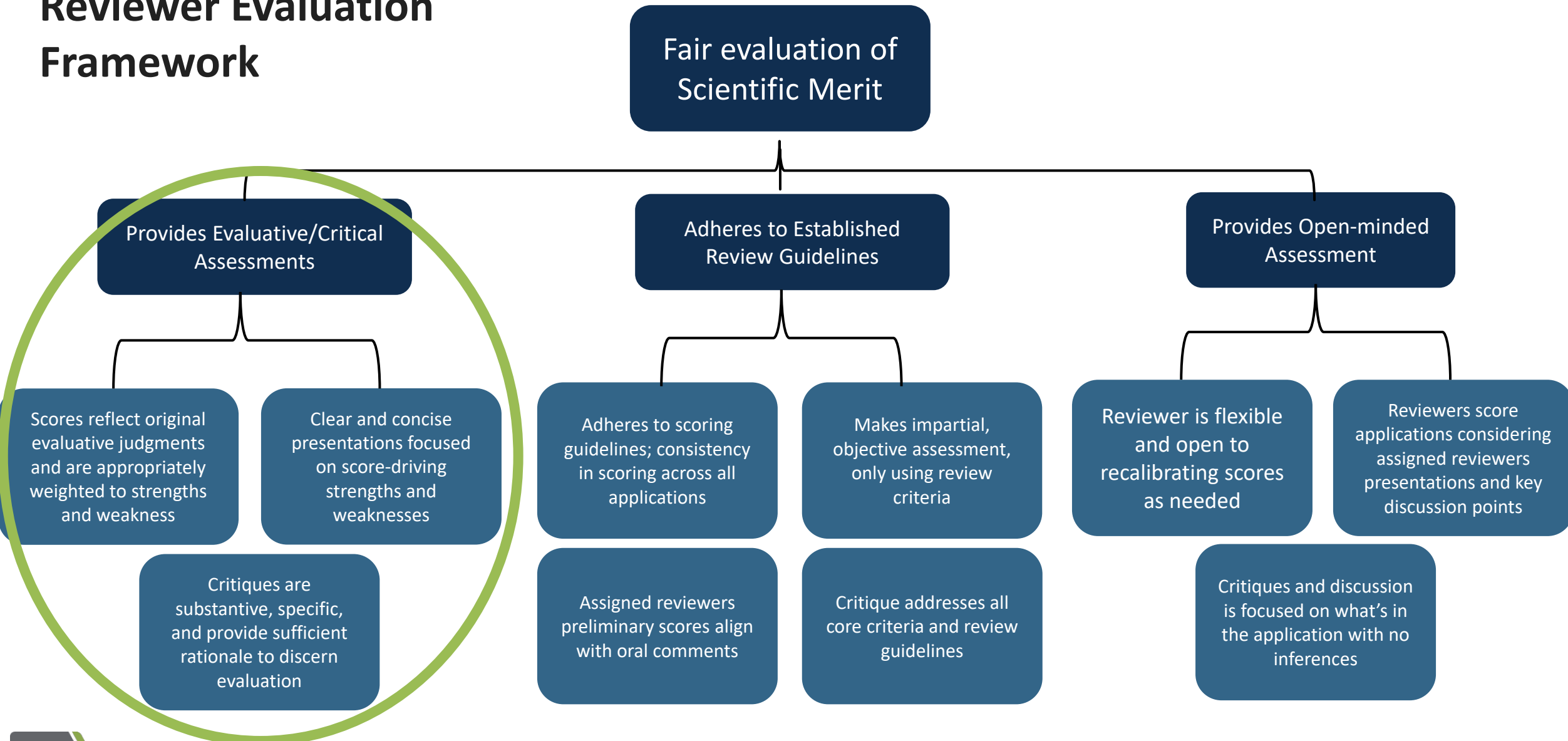
Provides Open-minded Assessment

Reviewer is flexible and open to recalibrating scores as needed

Reviewers score applications considering assigned reviewers presentations and key discussion points

Critiques and discussion is focused on what's in the application with no inferences

Reviewer Evaluation Framework



DOMAIN 1

Reviewer provides evaluative and critical assessments

Indicator 1

Written critiques are substantive, specific, and provide sufficient scientific rationale to discern evaluation.

Qualitative Measures

SRO

- Critique is focused on evaluative assessment and is not descriptive.
- Assessment is specific and avoids vagueness.
- Statements are supported by clear explanation/rationale.

Automated

- Machine learning can look for descriptive information or copy and paste from application. *Flags for specific language not associated with scores. AI could be trained on prior critiques.*

Indicator 2

Assigned scores (both overall and factor) reflect evaluative judgements and are appropriately weighted to application's noted strengths and weaknesses.

Qualitative Measure

SRO

- High factor strength/overall impact scores are aligned and have evaluative comments that reflect clear strengths
- Low scores are aligned with evaluative comments that reflect clear weaknesses.
- Moderate scores are appropriately justified by evaluative comments reflecting the relative strengths and weaknesses.

Indicator 3

Oral presentations are clear and concise, focused on overall impact and score-driving strengths and weaknesses.

Qualitative Measures

SRO/Chair

Reviewer presentation addresses potential impact of work in the field with attention to the importance of the research and rigor/feasibility, notes main strengths and weaknesses, and aligns with assigned scores.

Survey

Post meeting survey or poll of reviewers to identify who stood out and who might need more training.

Reviewer Evaluation Framework

Fair evaluation of Scientific Merit

Provides Evaluative/Critical Assessments

Scores reflect original evaluative judgments and are appropriately weighted to strengths and weakness

Clear and concise presentations focused on score-driving strengths and weaknesses

Critiques are substantive, specific, and provide sufficient rationale to discern evaluation

Adheres to Established Review Guidelines

Adheres to scoring guidelines and is consistent in scoring across all applications

Makes impartial, objective assessment, only using review criteria

Assigned reviewers preliminary scores align with oral comments

Critique addresses all review factors and adheres to review guidelines

Provides Impartial and Open-minded Assessment

Reviewer is flexible and open to recalibrating scores as needed

Reviewers score applications considering assigned reviewers presentations and key discussion points

Makes impartial, objective assessment, only using review criteria

Critiques and discussion is focused on what's in the application with no inferences

DOMAIN 2

Reviewer adheres to established NIH review guidelines.

Indicator 1

Written critique addresses review factors (simplifying review language) and additional review criteria and adheres to review guidelines.

Qualitative Measures

SRO/Chair

- Critiques address each factor as well as additional review criteria
- Adheres to NIH grant and review policy (e.g., unique requirements for various review mechanisms; need not be hypothesis driven).

Automated System

- Checks to flag for incomplete critiques and application of inappropriate review criteria (e.g., flag for “preliminary data” for R21 application).

Indicator 2

Reviewer adheres to scoring guidelines and is consistent in scoring across all applications.

Quantitative Measures

Scoring Analysis

- Within reviewer cluster analysis; consistent outlier compared to other assigned reviewers.
- If Reviewer's score is consistently at the limit of the ranges of other scores (within meeting or across multiple meetings), potential opportunity for training.
- Scoring matrix in REV provides score ranges used by each reviewer in the meeting; can evaluate score inflation or compression.

Reviewer Evaluation Framework

Fair evaluation of Scientific Merit

Provides Evaluative/Critical Assessments

Scores reflect original evaluative judgments and are appropriately weighted to strengths and weakness

Clear and concise presentations focused on score-driving strengths and weaknesses

Critiques are substantive, specific, and provide sufficient rationale to discern evaluation

Adheres to Established Review Guidelines

Adheres to scoring guidelines; consistency in scoring across all applications

Critique addresses all core criteria and review guidelines

Provides Impartial and Open-minded Assessment

Reviewer is flexible and open to recalibrating scores as needed

Reviewers score applications considering assigned reviewers presentations and key discussion points

Makes impartial, objective assessments only using review criteria

Critiques and discussion is focused on what's in the application with no inferences

DOMAIN 3

Reviewer is impartial and open-minded.

Indicator 1

Reviewer (assigned) is flexible and open to recalibrating scores as needed.

Qualitative Measure

SRO and Chair

Appropriate rationale is provided for when scores change, and for when scores do not change.

Quantitative Measure

- Scoring analysis – consistent/repeated out of range scoring for panel on reviewers' assigned applications as a flag.
- Scoring analysis – compare change in preliminary and final scores to look for any patterns (flag).

Indicator 2

Provides impartial, objective assessments.

Qualitative Measure

SRO (critiques and discussion)

- No evidence of bias due to sex, ethnicity, age, language, career stage, institutional affiliation, geography.
- No evidence of scientific/methodological bias; opinions must be supported by scientific reasoning.
- Uses information within the grant application materials without making inferences/assumptions or relying on reputation of investigators or environment.

Automated

Machine learning can check for demographic references and uncivil language content?

Indicator 3

Reviewer (as unassigned reviewer) scores applications considering assigned reviewers' presentations and key discussion points.

Qualitative Measure

- **SRO/Chair**
 - ~~Assess reviewer level of engagement (difficult to define/measure)~~
 - Reviewer appropriately identifies conflicts (personal and scientific bias)
- **Self evaluation** (provide some data on engagement at the reviewer level but need to careful the focus is on the reviewer and not the study section).
 - **Post meeting Survey**
 - How often did you score the average of the assigned reviewers scores vs sided with one or another argument
 - To what extent did you feel like you were engaged (engagement score relative to others in the room)

Quantitative Measure

- Scoring analysis – average of primary reviewers scores consistently
 - Challenge - this might be problematic since it's hard to decipher what caused them to score the way they did

Final Reviewer Evaluation Framework

Fair Evaluation of Scientific Merit

Provides Evaluative/Critical Assessments

Scores reflect original evaluative judgments and are appropriately weighted to strengths and weakness

Clear and concise presentations focused on score-driving strengths and weaknesses

Critiques are substantive, specific, and provide sufficient rationale to discern evaluation

Adheres to Established Review Guidelines

Adheres to scoring guidelines; consistency in scoring across all applications

Critique addresses all core criteria and review guidelines

Provides Impartial Open-minded Assessment

Reviewer is flexible and open to recalibrating scores as needed

Reviewers score applications considering assigned reviewers presentations and key discussion points

Makes impartial, objective assessments only using review criteria

Discussion/Questions

Challenges

- Identifying appropriate, specific measures without becoming too narrow
- Balancing qualitative and quantitative measures
- Considering different perspectives for evaluation (i.e., SRO, chair, other reviewers, self)
- Considering efficiency, feasibility

Next Steps

- Finalize Domain 3 Indicator 3
- Develop a post meeting assessment based on identified measures
 - Simple 1-page assessment completed by SRO and Chair for each reviewer?

Feedback, concerns and questions?