Panel-Based Evaluation of Interdisciplinary Researchers: A Protocol for National Agencies

A. Rivero and A.I. Scaffold

2026

Abstract

National evaluation agencies face a structural challenge when assessing researchers classified as "Interdisciplinary": standard single-discipline panels cannot fairly evaluate portfolios that span multiple fields. We propose an evaluation protocol based on a three-component indicator panel — disciplinary diversity, network coherence, and cross-field effect — that characterizes the type and degree of boundary-crossing in a researcher's work. Using mock researcher profiles, we demonstrate how the panel informs committee composition and distinguishes genuine integrators from polymaths and misclassified specialists. We identify six failure modes in interdisciplinary evaluation and propose safeguards for each. The protocol is designed to complement, not replace, expert judgment.

Introduction

National evaluation agencies periodically assess researchers for career advancement, funding eligibility, and institutional accreditation. These assessments typically rely on disciplinary panels: a committee of experts in the researcher's field evaluates their publication record, citations, and broader contributions. This procedure works well when the researcher's portfolio falls within a single recognized discipline.

A problem arises when a researcher is classified in the science group "Interdisciplinary" — a category that exists in several national systems (e.g., Spain's ANECA, Italy's ANVUR) for researchers whose work does not fit any single disciplinary panel. Standard evaluation procedures assign reviewers from a single discipline, creating a structural mismatch: the committee lacks expertise in part of the researcher's portfolio, or worse, applies disciplinary norms (publication venues, citation rates, methodological standards) that are inappropriate for cross-disciplinary work.

This note proposes an evaluation protocol that uses a three-component indicator panel to characterize interdisciplinary researchers and inform committee composition. The panel, developed in a companion review (Rivero, 2026), combines disciplinary diversity (Δ), network coherence (S), and cross-field effect (E). We demonstrate the protocol on mock researcher profiles, identify failure modes, and propose safeguards.

Operationally, we treat "interdisciplinary" as an integration claim, not just a breadth claim: high input diversity can represent either genuine integration or disconnected multidisciplinarity, and the protocol is built to distinguish these cases explicitly.

The Indicator Panel

We briefly summarize the three panel components; formal definitions and validation appear in the companion review.

Diversity (Δ): The Rao-Stirling index measures the disciplinary spread of a researcher's cited references, incorporating variety, balance, and disparity. High Δ indicates that the researcher draws on a broad range of fields.

Coherence (S): The mean linkage strength measures the average pairwise bibliographic coupling between a researcher's publications. High S indicates that publications share references across category boundaries — the researcher's diverse inputs are woven into a unified research program.

Cross-field effect (E): The fraction of citations received from outside the researcher's primary category. High E indicates that the researcher's work produces impact across disciplinary boundaries.

The key insight is that no single component suffices. Diversity alone cannot distinguish a genuine integrator (Δ high, S high, E high) from a polymath who publishes in multiple unrelated fields (Δ high, S near zero, E low). The full triple is needed.

Evaluation Protocol

Step 1: Compute the Panel

For each researcher classified as "Interdisciplinary," the agency computes (Δ, S, E) from publication and citation data. This requires:

- A disciplinary classification of cited references (e.g., Web of Science subject categories or Scopus ASJC codes).
- A pairwise similarity matrix between categories.
- Citation data for the cross-field effect.

Step 2: Classify the Profile

The panel values are compared against classification thresholds:

Classification	Δ	S	E
Genuine integrator	≥ 0.40	≥ 0.30	≥ 0.30
Polymath (non-integrative)	≥ 0.40	< 0.15	< 0.15
Specialist (reclassify)	< 0.35	any	any
Provisional (CI overlap)	boundary-overlap	boundary-overlap	boundary-overlap
Ambiguous (requires full panel review)	all other combinations		

These thresholds are illustrative and should be calibrated against empirical distributions before operational deployment. When confidence intervals overlap multiple rows, the case should be treated as provisional and sent to full-panel qualitative review.

Step 3: Compose the Committee

Procedure. Given researcher r with category proportion vector p_r (the fraction of references in each category):

1. Identify the primary categories: $K_r = \{i : p_{r,i} \geq \tau\}$, where $\tau = 0.15$.
2. For each category $i \in K_r$, include at least one evaluator from discipline i.
3. Include at least one evaluator with demonstrated cross-disciplinary expertise ($\Delta > 0.40$).
4. Committee size: $|K_r| + 1$ (disciplines represented plus a cross-disciplinary chair).
5. If profile classification is provisional (CI overlap), add one methods evaluator tasked with uncertainty and robustness review.

The committee composition is thus data-driven: it reflects the researcher's actual disciplinary profile rather than an arbitrary assignment.

Demonstration on Mock Profiles

We illustrate the protocol with three mock researchers, all classified as "Interdisciplinary" by the agency.

Researcher	Δ	S	E	Classification
Dr. A (integrator)	0.558	0.733	0.600	Genuine integrator
Dr. B (polymath)	0.562	0.000	0.063	Polymath
Dr. D (specialist)	0.288	0.881	0.211	Specialist (reclassify)

Dr. A publishes across condensed matter physics, physical chemistry, and molecular biology. Her publications share references across category boundaries ($S = 0.733$), and her work is cited across disciplines ($E = 0.600$). The panel correctly identifies her as a genuine integrator. Note that Dr. A and Dr. B have nearly identical diversity scores (0.558 vs. 0.562), yet receive opposite classifications — diversity alone is insufficient. The committee should include evaluators from her three primary fields plus a cross-disciplinary chair (4 members).

Dr. B has published in five different fields, but each publication is a single-field contribution with no shared references across domains ($S = 0.000$). His work is cited almost exclusively within each publication's own field ($E = 0.063$). The panel identifies him as a polymath. He should be reclassified to his strongest field or evaluated separately in each field — the "Interdisciplinary" label is misleading.

Dr. D concentrates in condensed matter physics and materials science ($\Delta = 0.288$), neighboring fields with high mutual similarity. Her coherence is high because all publications draw on the same knowledge base. The panel identifies her as a misclassified specialist who should be evaluated by a standard disciplinary panel for condensed matter physics.

Failure Modes and Safeguards

We identify six failure modes that can arise in interdisciplinary evaluation, even when using the panel.

F1: Breadth-without-depth reward

Risk: A researcher with high Δ is rewarded for breadth regardless of coherence or impact.

Safeguard: Require $S \geq 0.30$ and $E \geq 0.30$ for "integrator" classification. High diversity alone does not qualify.

F2: Non-standard publication penalty

Risk: Interdisciplinary journals typically have lower impact factors than top disciplinary journals. Researchers publishing in such venues are penalized in impact-factor-based assessments.

Safeguard: Use field-normalized citation indicators. Do not compare impact factors across fields.

F3: Incommensurable citation norms

Risk: Citation rates differ by an order of magnitude across fields (e.g., mathematics vs. molecular biology). An integrator bridging such fields will appear underperforming by the norms of the high-citation field.

Safeguard: Normalize E and citation counts by field-specific baselines. The panel's E is already a fraction (not an absolute count), which mitigates this partially.

F4: Misclassification persistence

Risk: A specialist enters the "Interdisciplinary" category and remains there because reclassification is bureaucratically difficult.

Safeguard: Panel-based reclassification trigger: if $\Delta < 0.35$, recommend reclassification to the researcher's primary field.

F5: Early-career data sparsity

Risk: Junior researchers have too few publications for reliable panel values. Nakhoda, Whigham, and Zwanenburg (2023) showed that Rao-Stirling confidence intervals can span up to 0.6 points for small samples. Even at mid-career, individual-level estimates carry wider uncertainty than aggregate measures, making threshold-based classification inherently noisy.

Safeguard: Minimum publication threshold (e.g., $n \geq 15$). Below this threshold, present panel values with confidence intervals, flag them as provisional, and defer hard classification unless the full interval lies inside one profile region.

F6: Gaming via strategic co-authorship

Risk: A researcher adds co-authors from distant fields to inflate Δ without genuine integration.

Safeguard: Weight Δ by corresponding-author publications only. Cross-check against S: genuine integration produces high coherence, while strategic co-authorship does not.

Discussion

The protocol proposed here is designed to complement expert judgment, not replace it. The panel provides structured evidence about the type of interdisciplinarity, and the committee composition rule ensures that the right expertise is present. But the final evaluation decision remains with the committee.

Several limitations should be noted. First, the classification thresholds are illustrative and require calibration against empirical distributions of panel values across disciplines and career stages. Second, the protocol assumes access to citation data, which may not be available for all researchers (particularly in the humanities). Third, the mock profiles use simplified data; real researcher portfolios are messier.

Rafols (2019) argued that science indicators should be contextualized, multidimensional, and subject to stakeholder validation. The protocol follows this principle: it is multidimensional (three components), context-dependent (committee composition adapts to the researcher's profile), and transparent (thresholds and rules are explicit and auditable). The panel should remain vector-valued evidence (Δ, S, E), not a single composite score.

The fundamental insight is that "Interdisciplinary" is not a single category. It encompasses integrators, polymaths, and misclassified specialists — researchers with qualitatively different profiles that require qualitatively different evaluation approaches. The panel provides the resolution to make these distinctions.

Conclusions

We have proposed an evaluation protocol for national agencies assessing researchers classified as "Interdisciplinary." The three-component indicator panel (Δ, S, E) provides the structural characterization needed to distinguish integrators from polymaths and specialists, compose appropriate evaluation committees, and guard against six identified failure modes. The protocol is transparent, auditable, and designed to be adapted to specific national contexts. Empirical calibration of the classification thresholds against real researcher distributions is the natural next step.

References

- Nakhoda, M., Whigham, P., and Zwanenburg, S. (2023). Quantifying and addressing uncertainty in the measurement of interdisciplinarity. *Scientometrics*, 128:6107–6127.
- Rafols, I. (2019). S&T indicators in the wild: contextualization and participation for responsible metrics. *Research Evaluation*, 28(1):7–22.
- Rivero, A. (2026). Measuring interdisciplinarity: A multi-component indicator panel for research evaluation. Manuscript.
- Xiang, S., Romero, D. M., and Teplitskiy, M. (2025). Evaluating interdisciplinary research: Disparate outcomes for topic and knowledge base. *Proceedings of the National Academy of Sciences*, 122(16):e2409752122.