INDIAN JOURNAL OF LIBRARY SCIENCE RESEARCH & INFORMATION TECHNOLOGY
ISSN- 3048-4677

VOL 3 Iss. No. 2 (Jul-Dec 2026)
https://www.bmsgroup.in/indian-journal-of-library-science-research-information-technology/
DOI: 10.5281/zenodo.22086139

Beyond Citation Counts: A Green Research Assessment Framework for Sustainable and Responsible University Evaluation

Swati Gupta[1]
Research Scholar
Department of Library and Information Science
Banasthali Vidyapith, Tonk, Rajasthan
Email: pgpha23029_swati@banasthali.in

Dr. Praveen Babel[2]
Research Supervisor,
Department of Library and Information Science
Banasthali Vidyapith, Tonk, Rajasthan
Email: praveenbabeludr@gmail.com

Abstract
Research output, citation counts, and peer credibility are frequently used to evaluate universities. These metrics demonstrate research exposure, but they may not accurately reflect how responsibly a university conducts its everyday operations, research planning, or work sharing. Although they can occasionally overlook the underlying significance of research itself, sustainable rankings help set goals for community involvement, research accomplishments, and teaching excellence. Environmental sustainability and responsible behaviors are carefully considered in the evaluations of the Green Research Assessment Framework (GRAF). It is founded on a comprehensive analysis of academic publications, worldwide assessment instruments, categorization schemes, and research investigations. Relevance, measurability, comparability, integrity, vulnerability to manipulation, and reporting difficulties are the six main lenses through which each piece of data is examined. Additionally, the framework addresses four primary areas: cooperation, sustainability in research publications, accessibility, and the environmental impacts of research activities—emphasizing transparency, responsibility, and advancement. Every indication consists of an entity, a parameter, a norm supported by evidence, and a safeguard. Effectiveness and transparency are clearly distinguished by GRAF, which also considers many fields and organizations, eliminates untrustworthy or out-of-date data, and emphasizes finding missing information rather than assigning a zero. The primary objective is to provide a user-friendly dashboard that offers a thoroughly verified overall grade. Bibliometrics, web-based results, sustainable accounting, and responsible metrics are the four main categories of evidence that are combined in this method. Additionally, the framework provides libraries, research centres, and ranking organizations with a detailed pilot plan. To make sure everything goes successfully, it's crucial to carefully test this method across several institutions and systems before implementing it extensively.

***Keywords**: Sustainability Ranking, Indicators, Web Evidence, Higher Education, Publication, Metric*

INTRODUCTION

Academic departments at higher education institutions consistently publish important research publications in their specialties, demonstrating their continued output. Measurable factors of academic success are frequently emphasized by universities. Because of their affordability and comparability, metrics including publication numbers, citation impact, journal repute, and survey results are commonly used. However, real academic excellence is not always correlated with how convenient it is to use these measurements. A citation, for instance, just indicates that a publication has been cited in other academic works.

The quality, public value, environmental advantages, or ecological costs of research endeavours cannot be adequately communicated by numerical measurements alone. As a result, the Leiden Manifesto (Hicks et al., 2015), the San Francisco Declaration on Research Assessment (DORA, 2012), and the Agreement on Reforming Research Assessment (Coalition for Advancing Research Assessment [CoARA], 2022) all support evaluation approaches that take multiple viewpoints, contextual factors, and transparency into account. In these frameworks, expert opinion is meant to be supplemented rather than replaced by numerical indications.

An additional level of complexity is added when investigating environmental sustainability. Academic institutions are urged to produce information on important topics such as resource consumption, pollution, biodiversity loss, climate change, and the politics of a just transition. However, research operations might result in travel emissions, depend on laboratory and digital infrastructure, and frequently demand energy and materials. Consequently, a university may write a lot about sustainability yet fail to reveal its own environmental effects. Conversely, a university may enhance estate management yet maintain a research program with little emphasis on sustainability. These difficulties might not be adequately represented in rankings based only on publication counts or campus sustainability ratings.

Notwithstanding these difficulties, sustainability rankings have broadened their focus to offer a more complete picture. Setting and infrastructure, energy and climate change, waste, water, transportation, education, and research are only a few of the ways UI Green Metric examines institutions (UI Green Metric, 2025). In the meanwhile, the Times Higher Education Impact Rankings evaluate colleges according to how they support the Sustainable Development Goals by compiling data from outreach, research, stewardship, and instruction (Times Higher Education, 2025). Raising institutional understanding of sustainability has been greatly aided by both programs. However, research comparing these rankings reveals only a tenuous connection between environmental sustainability initiatives and traditional academic prestige (Muñoz-Suárez et al., 2020). According to Boiocchi et al. (2023) and Galleli et al. (2022), methodological critiques frequently highlight persistent problems such self-selection bias, the difficulty of creating fair comparisons, disputed weighting schemes, a lack of attention to institutional context, and the difficulty of confirming self-reported data.

The disparity of indicators, rather than their scarcity, is the primary problem. Bibliometric analysis reveals the topics a university studies and the ways in which its findings are disseminated. Open-science metrics assess the reusability and accessibility of research. Research activities' ecological impact is measured by environmental accounting. Whether an organization reveals its objectives, strategies, and advancements may be seen on public websites and sustainability reports. These four categories are often evaluated independently, and transparency is frequently confused with actual performance. These categories should be included in a reliable green evaluation of research while following responsible metrics guidelines.

To assess institutions, this study presents a Green Research Assessment Framework that revolves around three key questions:

1. Without limiting sustainability to publication counts or campus activities, what essential elements are required to assess the environmental sustainability of a university's research endeavours?

2. How can bibliometric data, administrative records, and web-based information be used efficiently while maintaining appropriate denominators, means of verification, and safeguards against manipulation?
3. Rather than merely highlighting the shortcomings of league table rankings, how could we normalize and display indicators in a way that supports expert judgment? This study's main contribution is methodological- it offers a framework rather than ranking universities. The upcoming sections outline a pilot-ready indicator structure, scoring system, evidence protocols, and validation plans. It's important to note that the suggested weights are not meant to be universally applicable, nor does public disclosure automatically demonstrate true environmental performance.

2. Conceptual foundations

2.1 Responsible research assessment

The core tenet of responsible evaluation is that an indicator only gives a partial picture of an activity and does not fully capture its extent. The use of journal-level measures to assess the calibre of individual work is criticized by DORA (2012). Field normalization, local relevance, openness, data verification, and frequent review are all emphasized in the Leiden Manifesto (Hicks et al., 2015). CoARA (2022) also supports these ideas, recognizing a wider range of outputs, methods, and academic responsibilities and promoting qualitative evaluation backed by careful application of quantitative data.

The evaluation of sustainability is significantly impacted by these promises. Regardless of their actual environmental performance, major institutions in publication-intensive subjects tend to be rewarded when SDG-related articles are recorded in absolute numbers. Additional biases resulting from language, database breadth, geography, and academic field are included in citation-based measures. The release of a single aggregated ranking makes it difficult to determine if a university's excellent performance is a result of its extensive emissions reporting, its production of sustainability research, or just the size of its internet presence. As a result, GRAF views disaggregated reporting, mission relevance, normalization, transparent definitions, and data quality as necessary design restrictions rather than optional improvements.

2.2 From sustainable universities to sustainable research

Aspects including teaching, research, campus operations, governance, and the relationships between universities and society are typically the focus of study on sustainable universities. A useful difference between an institution's actions and their effects on society, the economy, and the environment is made by Findler et al. (2019).

This is an important distinction. An real decrease in emissions or resource use is a result, whereas an institutional policy or program is an output. According to Ceulemans et al. (2015), sustainability reporting can help organizations learn, but procedures vary widely and often don't lead to noticeable improvement.

Universities are gathering environmental data and sharing their practices as a result of green rankings. Despite their similarities, institutional sustainability and research sustainability are not the same. Information from teaching, housing, administration, and research activities is included in campus data on trash, water, power, and transportation. Simultaneously, research-related concerns frequently go beyond the purview of campus-wide reports, such as laboratory energy costs, fieldwork, conference and collaboration travel, high-performance computing, data storage, purchases related to research, open access initiatives, data sharing, and the main focus of the research itself. Research-specific frameworks should make use of campus data where appropriate, but they shouldn't presume that research activities are the only reason for all aspects of institutional success.

2.3 Bibliometric evidence and its limits

Researchers can examine research output, cooperation trends, open-access status, citation effect, and discipline mix at the institutional level using bibliographic databases. Although Scopus and Web of Science remain the main resources for these kinds of analyses, their coverage differs according on the subject, language, area, and kind of document (Pranckutė, 2021). There are further challenges involved in mapping the Sustainable Development Goals (SDGs). Depending on the mapping strategy used, different suppliers and research teams use different query sets and classification techniques, leading to different publishing sets for the same SDG (Raman et al., 2023). As a result, SDG-related publishing shares are only significant if they are accompanied by comprehensive details on the mapping version, database utilized, extraction date, document genres, affiliation disambiguation techniques, and an examination of relevant uncertainties.

Citation impact does not constitute an effective indicator of sustainability on its own. Though these factors may not always translate into environmental benefits, articles are often mentioned because of their contentious nature, the vastness of the area, or the availability of helpful approaches. The influence of field- and year-normalized citations can show how much other academics interact with the work in a sustainability portfolio, which is instructive. However, other evidence—such as citations in policy papers, inclusion in technological standards, use in clinical or professional guidelines, patents, or explicit acceptance by public authorities and community organizations—is needed to demonstrate social influence. Bibliometric evidence should not be the main emphasis of GRAF, but rather one of its components.

2.4 Web evidence, disclosure, and the greenwashing problem

By giving employees, students, donors, and local communities access to information like sustainability policies, greenhouse gas inventories, objectives, statistics, and progress reports, institutional websites help companies become more accountable to the public. The availability, accessibility, dependability across time, and interconnection of this data are assessed using webometric techniques (Aguillo et al., 2006; Thelwall, 2018). Methodological notes, yearly data tables, boundary definitions, archival reports, procedures for responsible assessment, and details about the responsible office are all important resources when it comes to research sustainability.

Easy access to information does not always mean that an organization is operating well. Even a visually pleasing sustainability site could not show real progress or hide insufficient data. On the other hand, some underfunded institutions may be achieving important progress but failing to appropriately share it. In order to overcome this problem, GRAF uses evidence-quality multipliers, independently assesses the quality of disclosure, and makes sure that, where possible, outcome indicators are generated from time-series data. It doesn't take into consideration the quantity of webpages since big websites, redundant material, and search engine quirks might make these measures deceptive and don't necessarily improve accountability.

3. Method

3.1 Design

The design used was a structured integrative framework-synthesis. This strategy was appropriate given the data that was available, which included methods from ranking organizations, peer-reviewed studies, and research assessment criteria. Neither estimating the impact of an intervention nor performing a thorough analysis of a particular genre of literature were the goals. Instead, the goal was to pinpoint important ideas, deal with related measurement problems, and convert these ideas into useful indicators.

3.2 Source identification and eligibility

On July 14, 2026, a search was completed. The official websites of research assessment initiatives, intergovernmental organizations, and ranking suppliers were examined in addition to publisher and DOI information. The evaluation concentrated on four primary areas: webometrics and sustainability reporting; SDG publishing mapping and societal effect; sustainability assessment and environmental indicators; and research assessment and university rankings. We also looked at references in important foundational books.

Sources that covered at least one of the following subjects were included: responsible use of research indicators; measurement of sustainability at universities; identification of sustainability research using bibliometrics; quality of public reporting and evidence; or comparison and normalization of indicators. Official documents and peer-reviewed studies were given priority. Excluded sources included news stories, commercial summaries, social media posts, Wikipedia entries, untraceable reports, duplicates, and sources with insufficient bibliographic information. While earlier research was included if it established enduring methodologies or ideas, recent sources were mostly chosen to represent current methods.

3.3 Coding and Framework Development

So as to perform a comprehensive and meticulous evaluation, we thoroughly assessed each source using six key criteria. Construct relevance: What does the measure indicate in terms of sustainable research? Operational clarity: Are the time period, boundary, denominator, and numerator all clearly stated?

Evaluation: Does the metric take into consideration variations in mission, field mix, and institutional size?

Evidence quality: Is the measurement repeatable, current, accessible to the general public, and subject to independent confirmation? Behavioural risk: Is the metric susceptible to manipulation or the creation of perverse incentives? Data burden: Is it possible to get the evidence without rejecting organizations with poor analytical capabilities?

If a candidate indicator provided information not covered by existing indicators, represented a novel idea, and was supported by convincing evidence, it was retained. Measures that overlapped were combined. Institutional slogans, unadjusted totals, and indicators based only on the volume of web material were not included. To ensure adherence to the concepts of contextualization, transparency, diversity, and review, the remaining indicators were examined.

3.4 Scope and unit of assessment

The evaluation focuses on the university's research activities for a certain reporting year within a specific organizational boundary. The framework requires that the metric be classified as campus-wide rather than exclusively research-focused when research-specific environmental data cannot be separated from the institution's overall data. Every year, operational activities are recorded. In order to minimize oscillations and guaranty that the data is still relevant to current policy, publication indicators are generated from a four-year rolling period.

4. Results: The Green Research Assessment Framework

4.1 Architecture

There are fifteen indicators and four dimensions in the Green Research Assessment Framework (GRAF) (see Table 1). It functions primarily as a dashboard for profiles. The focus of institutional research, the methods for knowledge application and dissemination, the environmental impacts of research operations, and the reliability of institutional administration and progress reporting are all related to particular questions. Maintaining the separation between these domains guaranties that excellent performance in one does not mask shortcomings or insufficient information in other areas.

Table 1. Green Research Assessment Framework

Dimension and indicator	Operational measure	Preferred evidence	Principal safeguard
A. Sustainability orientation of the research portfolio (30%)			
A1. Sustainability-relevant output share	SDG- or environment-relevant outputs divided by all indexed outputs in the same window	Versioned bibliographic query and a disambiguated affiliation set	Report database and query sensitivity; raw counts not permitted
A2. Field-normalised scholarly uptake	Mean normalized citation score of the sustainability portfolio	Article-level citations by field and year	Citations must never be read as environmental benefit
A3. Interdisciplinary integration	Diversity and cross-field integration within sustainability outputs	Article-level subject categories and reference diversity	Correct for the granularity of the classification scheme
A4. Documented policy or practice use	Share of sustainability outputs with traceable use in policy, standards, guidance, patents, or community practice	Persistent links and documented cases	Identifiable use required; media attention alone does not qualify
B. Openness, collaboration, and societal usability (20%)			
B1. Open-access availability	Openly accessible sustainability outputs divided by all sustainability outputs	DOI-level OA status and repository records	Distinguish lawful OA routes; report metadata coverage
B2. Open research objects	Outputs with accessible data, code,	Repository identifiers and data-	Permit justified ethical, legal, and

	protocols, or materials where sharing is appropriate	availability statements	community restrictions
B3. Equitable sustainability collaboration	Collaboration with lower-resourced regions and non-academic partners, reported by contribution role	Affiliations, contribution statements, partnership records	No reward for token authorship or extractive partnerships
C. Environmental footprint of research processes (25%)			
C1. Research greenhouse-gas intensity	Scope 1–3 emissions attributable to research per research FTE and, secondarily, per unit of research expenditure	Inventory with stated organizational and operational boundaries	Publish the allocation method and an uncertainty interval
C2. Research-travel intensity	CO_2e from grant-funded and institution-funded research travel per research FTE	Travel records and stated emission factors	Include rail and ground travel; flight-only accounting undercounts
C3. Laboratory and computing efficiency	Energy or CO_2e intensity for laboratories, core facilities, high-performance computing, and major data services	Metered data or a transparent estimation model	Do not compare unlike facility types without stratification
C4. Circular research practice	Proportion of priority research waste and procurement streams covered by reduction, reuse, recovery, or responsible disposal	Procurement and waste audits	Measure outcomes and material quantities, not the presence of a policy
C5. Verified improvement trajectory	Change in C1–C4 against a fixed baseline and a science-aligned target	At least three comparable reporting points	Disclose boundary and method changes; baseline resets prohibited
D. Transparency, governance, and continuous			

improvement (25%)			
D1. Disclosure completeness and assurance	Coverage of required indicators, methods, limitations, and independent assurance	Public report, methods annex, assurance statement	Score disclosure separately from performance
D2. Data recency and persistence	Age of the latest evidence, plus availability of an archived time series	Dated, persistent institutional pages and files	Penalize stale or disappearing evidence
D3. Digital findability and machine readability	Steps from the institutional home or research page to the evidence; structured downloadable tables	Website audit using a fixed protocol	No page-count metric; test accessibility and link integrity
D4. Responsible-assessment governance	Published policy showing how sustainability evidence informs assessment without displacing peer judgement	Policy, implementation plan, review date, accountable office	Reward implementation evidence, not endorsement alone

Rather than being final numbers, the dimension weights are meant to serve as first guidance for prototype projects. The framework's emphasis on knowledge generation is shown in the 30% weight given to research direction. In order to prevent publishing performance from taking precedence over operational duties or jeopardizing the reliability of the evidence, environmental process and governance are each given 25% of the budget. The remaining 20% are societal usability and openness. Before implementing a weighted summary, institutions should ask pertinent groups for input and share the results of testing different weightings.

4.2 Measurement rules

Portfolio orientation: Instead of just reporting the number of publications, A1 must be provided as a proportion. All qualifying outputs related to the institution within the given time frame must be included in the denominator. The database used, the mapping vocabulary, the query version, the document kinds, the extraction date, and the affiliation resolution process must all be disclosed by institutions. A1 should be computed using at least two appropriate mappings on a subsample, or a sensitivity interval should be supplied, because to the inherent variability in SDG mappings.

A2 measures academic involvement only by using field- and year-normalized citation impact inside the sustainability portfolio. A3 acknowledges that sustainability concerns frequently span different disciplines; nonetheless, results should be evaluated within the context of the institution's goal. A4 addresses utilization beyond academic citation networks separately. For instance, a specialized environmental institute shouldn't be at a disadvantage just because its field portfolio is more limited than that of a comprehensive university.

Transparency and fair usage. Accessibility and repeatability are operationalized in B1 and B2. While recognizing that openness may have justifiable constraints, UNESCO's Recommendation on Open Science places a strong emphasis on accessibility, inclusivity, transparency, and sustainability (UNESCO, 2021). GRAF takes a similar stance, allowing reasonable limitations on intellectual property, privacy, endangered species, and Indigenous or community-governed knowledge. If the limitation is recorded together with a metadata record, an access method, or placement in a restricted repository, the output is not considered closed.

B3 is intended to stop international co-authorship from being seen as proof of fair cooperation. Partner type, geographic location, contribution roles, and, when supported by evidence, collaborative agenda-setting or benefit-sharing should all be included in reporting. Compared to a single count, presenting a distribution of these factors yields a more realistic portrayal.

Performance of environmental processes. Because research-specific environmental records are rarely kept up to date, C1 through C5 need substantial institutional development. C1 should explicitly state which Scope 3 categories are included and use a recognized greenhouse gas accounting standard. In cross-country comparisons, two denominators are recommended: research spending adjusted for purchasing power and full-time equivalent research staff. On its own, neither denominator is enough. Reporting both makes it clear if the labor scale or the finance scale affects the results.

When trip-purpose coding allows for such distinction, C2 separates travel connected to research. Because wet laboratories, observatories, field stations, and computing facilities have different energy profiles, C3 has to be stratified by facility type. C4 concentrates on material outputs in priority streams, such as research procurement, chemicals, laboratory plastics, and electronic equipment. Instead of an advantageous starting position, C5 acknowledges improvement in relation to a steady baseline.

Governance and transparency. The purpose of D1 through D4 is to stop public communication from serving as a stand-in for real performance. D1 records assurance and completeness. D2 keeps track of archive continuity and data freshness. D3 evaluates users' ability to find and utilize the evidence. D4 assesses how responsible assessment and research governance include sustainability. If a statement or policy is not supported by an implementation mechanism, a designated accountable office, a review cycle, and proof of practical use, it merely obtains a baseline score.

4.3 Evidence-quality adjustment

Let z(i,j) be the context-normalized performance value on a 0–100 scale for institution i and indicator j. The value adjusted for evidence is:

$$a(i,j) = z(i,j) \times q(i,j) \times f(i,j)$$

where f is the freshness coefficient, and q is an evidence-quality coefficient. For instance, q = 0.75 for entire public evidence without external assurance, q = 0.50 for incomplete evidence, and q = 1.00 for independently guaranteed or reproducibly verifiable evidence. Unverifiable statements should not be regarded as measurable performance, but rather as missing data. The formula for freshness is $f = \exp(-\lambda t)$, where λ is predetermined, and t is the age of the evidence in years. Slow-moving policy initiatives should not deteriorate as quickly as operational measures. Transparent reporting of missing data is required. It is incorrect to assume that non-disclosure equates to zero performance or zero emissions. Dimension coverage is defined as follows when aggregation is required:

$c(i,d)$ = (weight of observed indicators in dimension d) ÷ (total planned weight in dimension d)

A dimension score should be withheld from publication if coverage falls below a predeclared threshold, provisionally set at 70%. Any overall score must include a coverage penalty and be presented alongside the unpenalized profile:

$$G(i) = \Sigma \text{ over } d \text{ of } [\, w(d) \times D(i,d) \times c(i,d)^\alpha \,]$$

In this formula, $D(i,d)$ denotes the evidence-adjusted dimension score, $w(d)$ is the dimension weight, and α determines the severity of the missing-data penalty. The values for $w(d)$, α, the normalization bounds, and any historization rules must be published and subjected to sensitivity analysis.

4.4 Reporting format

A four-dimensional dashboard with each indicator's score, uncertainty or sensitivity range, coverage, evidence grade, and trend is the suggested output. Before making any worldwide comparisons, institutions should be compared within well-defined mission and context groupings. A single ordinal rank should be avoided since little methodological adjustments might cause institutions to be rearranged and give the impression that they are more precise. Score bands and profiles are more defendable than a league table if a summary is needed.

5. Discussion

5.1 What GRAF adds

A research-specific dimension that is inconsistently addressed by current methods is introduced by GRAF. Research volume, influence, reputation, revenue, and internationalization are all prioritized in traditional rankings. Although the institutional viewpoint is broadened by sustainability rankings, they frequently mix together different purposes and heavily rely on self-reported data. The content of the research portfolio, the usability and accessibility of its knowledge, the environmental costs associated with its production, and the credibility of governance and reporting are the four categories of responsibility that GRAF integrates while maintaining its focus on the research enterprise.

Three recurrent category mistakes in the literature are also addressed by the framework. First, it makes a distinction between real sustainability performance and the quantity of publications pertaining to sustainability. Secondly, it distinguishes between significant institutional action and internet presence. Third, and perhaps most importantly, it acknowledges that methodological choices about borders, denominators, mapping algorithms, missing data, and weighting create aggregate scores rather than making them objective realities. The framework allows for critical evaluation of these features by preserving distinct dimensions and disclosing methodological decisions.

5.2 Implications for universities and information professionals

Instead of using the framework comparatively at first, universities should use it diagnostically. To identify which indicators are currently accessible, which are campus-wide, which may be linked to research activities, and which will need new data gathering methods, a pilot phase should start with data mapping. Research offices are able to link project and funding records to travel and deliverables. Inventory limits and reduction paths can be recorded by sustainability offices. Research-related expenses and material streams can be identified by the finance and procurement divisions. The open-access and research-object indicators are supported by institutional repositories.

Information specialists and libraries have a bigger impact than is usually acknowledged in the literature. Indicators A1–A4, B1–B2, and D2–D3 are directly supported by their proficiency in affiliation disambiguation, bibliographic databases, metadata quality, permanent identifiers, repository administration, and web information architecture. To make sure that modifications to database coverage, SDG mappings, or open-access status do not unintentionally skew trend lines, libraries can optionally keep a versioned methodology register. This positioning of the library as an assurance partner inside the institution's evidence base goes beyond just giving citation counts.

5.3 Risks and safeguards

Every indication system has an impact on behavior, frequently in unexpected ways. A sustainability publication aim, for instance, may promote shallow keywording. Inappropriate information disclosure by researchers may result from an open-data aim. Partnerships that only exist in theory could be rewarded by a cooperation measure. Universities that conduct energy-intensive and socially useful research are penalized by a carbon-intensity index. The website assesses the ability to communicate rewards, which may not always result in significant action. These dangers cannot be eliminated by formula refinement alone.

Instead, GRAF depends on procedural protections such as multiple denominators, expert review, mission-based interpretation, evidence grading, clear admission of uncertainty, limitations on aggregate, and a planned cycle of indicator review. In addition to disclosing who chose the indicators, whose interests were represented throughout the selection process, and how disagreements are settled, institutions ought to be able to update their data. A metric is not meant to provide a conclusive conclusion, but rather to stimulate investigation.

5.4 A staged validation agenda

Each of the four phases of a validation agenda must be finished before any subsequent application. A panel of academics, sustainability experts, librarians, research managers, students, and representatives from lower-resourced research systems evaluates content validity in the first step. A multi-country trial that assesses data accessibility, reporting burden, inter-rater reliability, and the viability of research-specific environmental attribution is the second step.

By examining both the predicted links between dimensions and the expected absence of associations among others, the third stage evaluates construct validity. Disclosure quality, for instance, should not be assumed to represent real emissions performance, but it is expected to correspond with the completeness of the evidence. The sequence is completed by the fourth stage, sensitivity analysis, which assesses the impact of denominators, thresholds, weighting schemes, database selection, and Sustainable Development Goal (SDG) mapping.

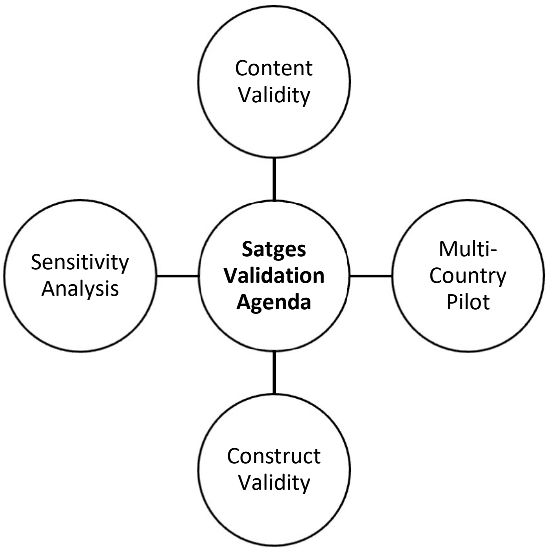

Figure 1.1 Validation agenda

If an indicator is inaccurate, duplicates another measure, encourages undesirable behavior, or unfairly disadvantages certain institutions for reasons unrelated to the aspect being assessed, it must be revised or removed. Validation is an ongoing process, not a one-time exercise. The framework must evolve as research methods, emissions accounting, and academic communication develop.

7. Conclusion

Adding green publication numbers to current rankings is insufficient to genuinely make research assessment more sustainable. Additionally, we should focus on four important areas: the subjects covered by your study, the fairness and openness of information sharing, the research's environmental impact, and how institutions manage and report these consequences. The GRAF framework's categorized, easily comprehensible indications aid in putting these concepts into practice.

This framework is intended to prevent the creation of an inaccurate impression of accuracy. It highlights missing data, separates disclosure from actual performance, employs fair comparison techniques, disregards faulty evidence, and allows users to combine findings if they so want. At the moment, it's mostly used for testing in various contexts and for assessing institutions. The framework can assist universities, funders, libraries, and ranking agencies in expanding their evaluation of research as it undergoes more testing and enhancements. It also serves as a reminder that expert opinion is supported by statistics, not the other way around.

References

1. Aguillo, I. F., Granadino, B., Ortega, J. L., & Prieto, J. A. (2006). Scientific research activity and communication measured with cybernetics indicators. *Journal of the American Society for Information Science and Technology, 57*(10), 1296–1302. https://doi.org/10.1002/asi.20433
2. Boiocchi, R., Ragazzi, M., Torretta, V., & Rada, E. C. (2023). Critical analysis of the Green Metric World University Ranking system: The issue of comparability. *Sustainability, 15*(2), 1343. https://doi.org/10.3390/su15021343
3. Ceulemans, K., Lozano, R., & Alonso-Almeida, M. D. M. (2015). Sustainability reporting in higher education: Interconnecting the reporting process and organisational change management for sustainability. *Sustainability, 7*(7), 8881–8903. https://doi.org/10.3390/su7078881
4. Coalition for Advancing Research Assessment. (2022). *Agreement on reforming research assessment.* https://coara.eu/agreement/the-agreement-full-text/

5. Declaration on Research Assessment. (2012). *San Francisco Declaration on Research Assessment.* https://sfdora.org/read/
6. Declaration on Research Assessment. (2024). *Guidance on the responsible use of quantitative indicators in research assessment.* https://sfdora.org/wp-content/uploads/2024/05/DORA_indicators_guidance.pdf
7. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research, 133,* 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070
8. Findler, F., Schönherr, N., Lozano, R., Reider, D., & Martinize, A. (2019). The impacts of higher education institutions on sustainable development: A review and conceptualization. *International Journal of Sustainability in Higher Education, 20*(1), 23–38. https://doi.org/10.1108/IJSHE-07-2017-0114
9. Gadd, E. (2021). Why global university rankings don't add up. *Frontiers in Research Metrics and Analytics, 6,* 680023. https://doi.org/10.3389/frma.2021.680023
10. Gadd, E., Holmes, R., & Shearer, J. (2021). Developing a method for evaluating global university rankings. *Scholarly Assessment Reports, 3*(1), 2. https://doi.org/10.29024/sar.31
11. Galleli, B., Teles, N. E. B., Santos, J. A. R. dos, Freitas-Martins, M. S., & Journeaux Junior, F. (2022). Sustainability university rankings: A comparative analysis of UI Green Metric and the Times Higher Education World University Rankings. *International Journal of Sustainability in Higher Education, 23*(2), 404–425. https://doi.org/10.1108/IJSHE-12-2020-0475
12. Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature, 520*(7548), 429–431. https://doi.org/10.1038/520429a
13. Muñoz-Suárez, M., Guadalajara, N., & Osca, J. M. (2020). A comparative analysis between global university rankings and environmental sustainability of universities. *Sustainability, 12*(14), 5759. https://doi.org/10.3390/su12145759
14. Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications, 9*(1), 12. https://doi.org/10.3390/publications9010012
15. Raman, R., Nair, V. K., & Nedungadi, P. (2023). Discrepancies in mapping Sustainable Development Goal 3 (good health and well-being) research: A comparative analysis of Scopus and Dimensions databases. *Sustainability, 15*(23), 16413. https://doi.org/10.3390/su152316413
16. Thelwall, M. (2018). Web indicators for research evaluation. Part 1: Citations and links to academic articles from the web. *El Profesional de la Información, 27*(5), 1019–1028. https://doi.org/10.3145/epi.2018.sep.07
17. Times Higher Education. (2025). *Impact Rankings 2025: Methodology.* https://www.timeshighereducation.com/world-university-rankings/impact-rankings-2025-methodology
18. UI Green Metric. (2025). *UI Green Metric World University Rankings guideline 2025.* https://uigreenmetric.com/wp-content/uploads/2025/10/UI-GreenMetric-Guideline-2025.pdf
19. United Nations Educational, Scientific and Cultural Organization. (2021). *UNESCO Recommendation on Open Science.* https://unesdoc.unesco.org/ark:/48223/pf0000379949
20. Ritter, K., Borst, C. W., & Chambers, T. L. (2018). Virtual Solar Energy Center Case Studies. The ASEE Computers in Education (CoED) Journal. http://asee-coed.org/index.php/coed/article/view/Ritter_Virtual.