METRICS not only measure what has been done, but influence what will be done. That is, they signal priorities, act as proxies, and shape trajectories. Funders use metrics to allocate grants, design requests for proposals, and track the responsible use of their funding. Recruiters rely on metrics to make hiring decisions. And, in colleges and universities, administrators at every level use metrics to evaluate and promote students, staff, faculty, and programs. In the academy, metrics are often powerful incentives, shaping researcher investments, behaviors, and careers [1].

While traditional research metrics like publication counts, journal impact factors, and tallies of grant income are bureaucratically useful, they do little to capture the aspects of research that are most important to society: identifying solutions to real-world problems, promoting science literacy and engagement, and fostering the next generation of change agents. Moreover, traditional metrics have built-in biases that have been shown to discriminate against certain research styles and topics, reward insularity over innovation, and promote research quantity over quality [2, 3].

Research that is co-produced and externally engaged is more likely to lead to societal impact, but is perceived by some as less likely to lead to scholarly definitions of impact. That is, the practice of engagement is at odds with traditional research metrics. Successful engagement requires long-term relationships, a commitment to co-development, and an orientation not to academia, but to products valued by communities and decision-makers. Building a culture of engagement therefore requires rethinking how we measure research impact in ways that better align scholarly and societal metrics.

How can institutions adopt a more comprehensive and pluralistic approach that leverages the power of metrics to promote engaged solutions to grand challenges?
BEYOND TRADITIONAL METRICS

Alternative metrics and approaches for measuring research impact have recently gained traction inside and outside the academy. Below, we summarize four alternatives that better capture societal impact and scholarly engagement than strict citation counts and other traditional metrics. Each comes with pros and cons.

Altmetrics

Algorithms can scrape publicly available Internet data to produce “altmetrics,” which stand as proxies for a research product’s degree of scholarly influence beyond academia. The company Altmetric provides a free tool for researchers to evaluate their online and social media presence, combining information including a given article’s clicks and downloads, social media and blog shares, and media mentions [4].

Pros: Altmetrics incentivize researchers to take ownership over the promotion and dissemination of their work and emphasize the contemporary importance of sharing and communicating research with (and within) the public. These scores may also indicate the relative relevance of scholarship to contemporary issues of public concern. Altmetrics also incentivize scientists to broaden their audiences [5], increasing public attention, university reputation, and attraction of students and researchers.

Cons: Altmetrics measure attention and popularity, not necessarily impact or influence. Social media-based metrics favor researchers with public profiles, are unable to differentiate between high- and low-quality research (for example, retracted studies often gain popularity following their retraction, and, for altmetrics, all publicity is good publicity), and can be manipulated [6]. Due to differential access to and use of varying web platforms, altmetrics are also biased toward and against scholars in certain countries [6].

Impact tracking

Consistent, systematic recording of research impact can complement traditional academic metrics. Researchers are encouraged to maintain an “impact file” in the same way they keep their CV or service activities up to date [7]. The impact file provides a ready resource for scholars asked to substantiate research impact, as when applying for grants or tenure. Among the items appropriate to an impact file are records of stakeholder meetings, engagements with outside organizations, invited talks, service on public advisory or non-profit boards, and documentation of the research’s translation into legislation, policy analysis, or management plans. Statements from external partners could support or validate the contents, via a list of references or the inclusion of external letters.

The impact file is also key to those asked, at the departmental and institutional levels, to complete annual surveys about their research and activities or track progress on impact metrics. Those seeking additional help with these annual reviews may look to templates like the Science +
Chapter Three: Metrics for Assessing Research Impact

Story “researcher thought leadership” survey [8]. This survey prompts scholars to consider (and increase) their publications and presentations for non-specialist audiences, media savviness and attention, and social media presence. Departments or institutions might also find such a survey fruitful if administered regularly over time, tracking shifts in impact activities.

**Pros:** Impact files incentivize researchers to invest in building long-term, responsible, and productive relationships with external partners. They provide ready evidence of impact for quick turnaround requests and can form the basis of annual, impact-oriented institutional reporting. Existing templates, such as the “I Want to Plan My Impact” organizer from Fast Track Impact, are available to researchers who wish to track external partners, outreach strategies, and indicators of external impact.

**Cons:** Careful and consistent impact tracking demands additional time, beyond that already invested in engagement and co-production, from both researchers and evaluators. Programs may need to build additional organizational capacity to administer surveys, solicit letters, or otherwise substantiate impact reports. External feedback is also burdensome for external stakeholders requested to validate research impact, and narrative reporting may be dismissed as too subjective or difficult to validate and compare across researchers or research programs. Finally, societal impact may take years or decades to demonstrate; impact files are, to some extent, misaligned with short-term milestones of grant reporting, evaluation, or promotion.

**Impact Compass**

The Impact Compass is a tool developed by the Stanford Graduate School of Business to assess individual or institutional performance across multiple metrics [9]. The tool helps students rank the social impact of potential employers, but can be adapted to measure research impact. The axes of the compass identify the values on which each assessment is based, scaled from one to three. The “fuller” the compass, the closer one is to their holistic goal. The impact compass could be adapted by individuals, as well as tenure committees, departments, or funding agencies, to assess multidimensional contributions of researchers, research programs, or institutional performance.

**Pros:** Standardized and simplified categories and scales improve the consistency of evaluation, leading to more transparent criteria for measuring impact. The tool is flexible; axes can be defined according to departmental priorities. The compass emphasizes multiple dimensions of impact and allows different types of researchers, who excel at different categories, to receive high overall scores.

**Cons:** The interpretation of scores remains subjective. For example, a lower score doesn’t necessarily indicate a less worthy opportunity, and a higher score may not mean an inherently better project or researcher. For example, on a “scale” axis, local work impact will score lower than global impact work, despite its importance and relevance to the local community. Researchers and institutions must develop an agreed-upon process for tracking, reporting, and interpreting Impact Compass scores.

**Narratives**

To avoid reducing diverse research products to numbers alone, researchers are often asked to draw out the connections between their work and its outcomes and impacts qualitatively,
through written stories, reports, or case studies. The narratives can be assessed by a panel of experts, included as part of academic CVs, or converted to scores using a standardized rubric.

Narratives about research impact are common to researchers applying for various kinds of funding. For example, the U.S. National Science Foundation (NSF) requires proposals to include a “broader impact” statement that discusses the societal relevance of the proposed activities and the involvement of underrepresented groups [10]. Because the NSF is so influential among U.S. funders, organizations like the American Anthropological Association recommend using the NSF’s Broader Impact guidelines as a starting place for creating university standards for assessing research impact [11].

**Pros:** Narratives incentivize researchers to articulate how and why their work is relevant, impactful, or important [12]. This allows researchers to communicate context that is not always legible in quantitative metrics. Narratives can provide compelling cases for difficult-to-prove impacts, including lagging impacts. Panels reviewing narratives can assess social/public impact holistically using any number of predetermined criteria and goals. Unlike many quantitative metrics, the narrative format is not discipline-specific.

**Cons:** Reviewing narratives is time intensive and expensive. Review processes may lack transparency and clear criteria. Interpretation of impact is subject to reviewers’ biases for specific topics, credentials, methodologies, geographies, etc. And in the narrative format, researchers are incentivized to “reach” for causation when societal impact is difficult to determine or the impact of the work has not yet been realized.

**BEST PRACTICES**

In developing metrics for societal impact, institutions can hew to the ethical and effective use of such metrics by drawing on the following guidelines [13–17].

1. Use “baskets of indicators” [13]. Evaluation should include a suite of quantitative and qualitative indicators that are designed to measure different aspects of research impact.

2. Adopt a “narrative with numbers” approach. Quantitative measurements should complement rather than replace qualitative data and stories.

3. Limit the use of journal impact factors, which measure journal prestige rather than research impact or quality [14].

4. Recognize and correct systemic and personal biases embedded in assessment processes [15]. For example, is local, qualitative work consistently being overlooked in favor of global, quantitative research?

5. Measure performance or merit against the mission statement of the research group, unit, or institution [16].

6. Encourage researchers to develop individual professional development plans and use these plans as a basis for annual evaluation. Collectively develop and agree upon the suite of metrics that will be used to measure research impact.

7. Evaluate and update indicators regularly [16].
U.K. Research Excellence Framework (REF2021)

The United Kingdom’s Research Excellence Framework (REF) is a national system used to assess research quality in higher education institutions in order to distribute educational funding in the U.K. The first version of REF was created in 2014 by the U.K.’s main funding bodies. The goal was to create a shared policy vision for the “continuation of a world-class, dynamic and responsive research base across the full spectrum within UK higher education” [18]. The REF evaluation process is designed to improve accountability for public investments in research, to provide benchmarking information for institutions, and to inform the allocation of funding across competing institutions and programs [18].

REF evaluation is focused on three elements: “Originality, significance and rigour” of the research outputs compared to international quality standards (60%); “Reach and significance” of research impacts on the greater society, culture, economy, policy, health, or environment (25%); and “Vitality and sustainability” of the environment in which the research takes place (15%) [19]. REF2014 was the first national evaluation to include societal impact criteria in the allocation of research funding [20]. In this scheme, universities submit case studies of exemplary research that are ranked by discipline-specific sub-panels of senior academics, international members, and research end users.

Research impact, as defined by REF, includes “an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia” [19]. Impacts can apply to any geographic unit—whether institutional, local, regional, national, or international—in which a “beneficiary, community, constituency, organisation or individual’s... activity, attitude, awareness, behaviour, capacity, opportunity, performance, policy, practice, process or understanding” is changed [19]. Institutions chose cases for submission that best illustrate areas where research made a “distinct and material contribution to the impact taking place, such that the impact would not have occurred or would have been significantly reduced without the contribution of that research” [19].

Submitted case studies must explain how (or through what means) the research led to or contributed to the presumed impact, and they must include sources of information to corroborate these claims, including documentation of policy impact, references to discussion of an academic or their work “in the records of meetings, conferences, seminars, working groups and other interchanges” [20].

Under the impact category, research institutions submit references—people or groups that can provide testimonials to corroborate the stated impact of each case study [19]. The relationship between research and impact need not be direct or linear (for example, in the case of co-produced research).
The review panel then rates the submitted works on the following scale, with scores helping to determine the distribution of the next year’s institutional support [21].

- **Four star**  Outstanding impacts in terms of their reach and significance.
- **Three star**  Very considerable impacts in terms of their reach and significance.
- **Two star**  Considerable impacts in terms of their reach and significance.
- **One star**  Recognised but modest impacts in terms of their reach and significance.
- **Unclassified**  The impact is of little or no reach and significance; or the impact was not eligible; or the impact was not underpinned by excellent research produced by the submitted unit.

The REF framework was updated in 2021, though it continues its reliance on panel review and narrative case studies as the basis for measuring research impact. Critiques of the REF process note that it is labor intensive, expensive, and takes away from other valued activities. Some have raised concerns that implementation of the REF has decreased the authenticity and novelty of research [22–24].

The data included in REF assessments provide unique opportunities to conduct secondary analysis of impact case studies. A recent analysis of REF projects rated highly for broader impact identified the importance of additional data interpretation, customized knowledge products, and investments in boundary-spanning or “social learning” as key components of high-ranking cases [25].
CONCLUDING THOUGHTS

Traditional metrics for assessing research impact, allocating funding, and evaluating promotion often discourage or disincentivize engaged scholarship. Because engagement takes time, impact is difficult to measure using traditional metrics such as publication count or grant income. To build a culture of engagement, then, we must rethink how we measure research impact that aligns scholarly and societal metrics.

Embracing both qualitative and quantitative approaches will strengthen the culture of engagement. Alternative and detailed metrics, while not immune to biases, can be collated for comprehensiveness and the inclusion of different perspectives that better capture impact and highlight engagement and outreach activities. Though the use of these metrics is resource-intensive, their development and institutionalization are essential for encouraging researchers to take part in engaged scholarship.

Institutions may further foster a culture of engagement by allowing researchers to articulate their own theories of change (see glossary) and choose the metrics that best fit assessment of their research. Any metric will, by design, incentivize activities and reporting that align with the metric. In other words, all metrics are biased, but some are useful, especially when coupled with narratives or information that helps situate the research in a broader societal context.

REFERENCES


