

COMMENT

EQUALITY Too few people who are LGBTQ go into science, and too many leave **p.27**

HERITAGE How UNESCO has tried to broker peace through science and culture **p.29**

EVOLUTION Synthesizing many lines of evidence to trace the spine's start **p.31**



BIOLOGY Wet labs squeezed by scarce funds and bureaucracy in Italy **p.32**

ILLUSTRATION BY DAVID PARKINS



A better measure of research from the global south

Funders **Jean Lebel** and **Robert McLean** describe a new tool for judging the value and validity of science that attempts to improve lives.

In India, the world's leading producer of mangoes, up to 40% of the harvested fruit is destroyed in transit before delivery. This costs up to US\$1 billion in lost income each year, affecting the lives and livelihoods of millions of farmers, traders and consumers. So researchers from India, Sri Lanka and Canada developed a suite of nanomaterials that can be sprayed onto fruit on the tree, in packaging or in transit, to extend its life. They

trapped hydrophobic hexanal molecules (derived from plant waste) in a hydrophilic membrane so that they could be suspended in liquid for application to the fragile fruit.

In Egypt, more than 95% of women have experienced sexual harassment at least once, and most cases go unreported. So, in 2010, researchers at the Youth and Development Consultancy Institute in Cairo developed Harrassmap (<https://harrassmap.org/en>).

This online interactive resource enables people to report and map cases of sexual harassment. When it emerged that university campuses were hotspots, Cairo University implemented a policy to combat sexual harassment, the first of its kind in the Middle East. Other universities in Egypt are following suit.

Both projects help to solve pressing societal challenges. The researchers involved ►

► appreciate that the people who benefit from the projects are the ones who are best placed to judge the value and validity of the work. The research teams spent time developing their hypotheses and results with those who feel the effects. In each case, the research is robust and life-changing — exactly the combination that most people would say is the very purpose of science.

But both projects would score poorly if judged using only conventional approaches to evaluating research quality that prioritize the opinion of peers, the volume of papers published, and citations. That's a problem because it is endorsement from other scientists, not stakeholders, that drives career advancement for researchers in Egypt, Sri Lanka and India, as everywhere else.

Is the weakness in the science or in the way it is measured? Too often it is the latter, in our view. Dominant techniques of research evaluation take a narrow view of what constitutes quality, thus undervaluing unique solutions to unique problems. At Canada's International Development Research Centre (IDRC) in Ottawa, we fund just this sort of research: natural and social science that unearths fixes for the development challenges facing countries in the global south. The majority of the work we support is led by researchers from these countries.

So we at the IDRC developed a tool to evaluate the quality of research that is grounded in, and applicable to, the local experience. We used it to assess 170 studies and then did a meta-analysis of our evaluations. The results suggest that it is possible — and essential — to change how we assess applied and translational research.

TUNNEL VISION

The limitations of dominant research-evaluation approaches are well known^{1–5}. Peer review is by definition an opinion. Ways of measuring citations — both scholarly and social — tell us about the popularity of published research. They don't speak directly to its rigour, originality or usefulness. Such metrics tell us little or nothing about how to improve science and its stewardship. This is a challenge for researchers the world over.

The challenge is compounded for researchers in countries in the global south. For instance, the pressure to publish in high-impact journals is a steeper barrier because those journals are predominantly in English and biased towards publishing data from the United States and Western Europe⁶. With the exception of an emerging body of Chinese journals, local-language publications are broadly deemed lower tier — even those published in European-origin languages such as Spanish, Portuguese or French.

The metrics problem is further amplified for researchers who work on local challenges. Climate adaptation research is a case

in point. Countries in the global south are on the front lines of global warming, where context-appropriate adaptation strategies are crucial. These depend on highly localized data on complex factors such as weather patterns, biodiversity, community perspectives and political appetite. These data can be collected, curated, analysed and published by local researchers. In some cases, it is crucial that the work is done by them. They speak the necessary languages, understand customs and culture, are respected and trusted in communities and can thus access the traditional knowledge required to interpret historical change. This work helps to craft adaptations that make a real difference to people's lives. But it is also fundamental to high-level meta-research and analysis that is conducted later, far from the affected areas⁷.

Does the current evaluation approach scrutinize and give equal recognition to the local researcher who focuses on specifics and the researcher who generalizes from afar? Does the current approach acknowledge that incentives are different for local and foreign researchers, and that those incentives affect research decisions? Are we adequately measuring and rewarding research that is locally grounded and globally relevant? In our view, the answer to all of these questions is no.

FROM NO TO YES

With the support and leadership of partners across the global south, the IDRC decided to try something different. The result is a practical tool that we call Research Quality Plus (RQ+)⁸.

The tool recognizes that scientific merit is necessary, but not sufficient. It acknowledges the crucial role of stakeholders and users in determining whether research is salient and legitimate. It focuses attention on how well scientists position their research for use, given the mounting understanding that uptake and influence begins during the research process, not only afterwards.

We think that the approach has merit beyond the development context. We hope that it can be tailored, tested and improved in a variety of disciplines and contexts, to suit the needs of other evaluators — funders such as ourselves, but also governments, think tanks, journals and universities, among others.

RQ+ has three tenets:

Identify contextual factors. There is much to learn from the environment in which research occurs. Instead of aiming to isolate research from how, where and why it was done, and by whom, evaluators should

examine these contexts to reach a claim about quality. For the IDRC, this included five issues: political, data, research environments, the maturity of the scientific field and the degree to which a project includes a focus on capacity strengthening. For another funder, journal or think tank, these might — or should — be different.

Articulate dimensions of quality. The underlying values and objectives of the research effort need to be made explicit. Evaluators weigh these dimensions of quality using a formula that fits the context and goals of the research. The dimensions that matter to the IDRC are: scientific integrity (a measure of methodological rigour), legitimacy (a measure of the fidelity of the research to context and objectives), importance (a measure of relevance and originality) and positioning for use (the extent to which research is timely, actionable and well communicated). (See Figure S1 in Supplementary Information.)

Use rubrics and evidence. Assessments must be systematic, comparable and based on qualitative and quantitative empirical evidence, not just on the opinion of the evaluator — no matter how expert they are. For the IDRC, this meant evaluators speaking to intended users, to others working in similar areas and to non-scientific beneficiary communities, as well as assessing research outputs and associated metrics.

ROAD TEST

The IDRC first used RQ+ in 2015. Independent specialists assessed 170 studies from 7 areas of research the centre had funded in the previous 5 years. For each area, three specialists rated projects using the three tenets described, looking at empirical data for each study: bibliometrics, interviews with stakeholders and IDRC reports on the work. The reviewers decided independently what data to collect and compare for each project, and held panel discussions to reach a consensus on the final ratings for each project.

This framework (see Figure S2 in Supplementary Information) encouraged a grounded, critical reflection on each project. And it helped systematic judgement to be applied across diverse contexts, disciplines and approaches to research. In exit interviews and follow-up discussions, the independent reviewers described the assessments as unlike any others they had done. They felt confident that the evaluation had been systematic, comprehensive and fair.

We learnt a lot from this process about the projects that the IDRC supports and how we could do better. For instance, we found that we need to prioritize gender across everything we fund, from climate modelling to the accessibility of justice, and not just in research projects that are

“Conventional evaluations were never this challenging, but neither were they so motivating.”



Women protest against sexual harassment in Cairo in 2013.

aimed specifically at women and girls. As enshrined in one of the United Nations Sustainable Development Goals (SDG5), gender equality is key for unlocking development potential, so it was a dimension examined by the reviewers.

They found, for example, that a programme using national data sets to examine the implications of taxation and food labelling should have disaggregated the data by gender to achieve more with the same investment. Reviewers also highlighted exemplars, such as the African Doctoral Dissertation Research Fellowship programme, which helps PhD students to complete theses at their home institutions, enabling greater uptake by female applicants who shoulder more family duties. The programme considers gender balance when selecting applicants, and in reviewing proposed research.

As a result, the IDRC has rolled out, among other things, a new data system to mine gender data and workshops for staff to share and see good work.

In our experience, conventional evaluations were never this challenging, but neither were they so motivating and useful.

THREE MYTHS BUSTED

To draw more-general lessons, the IDRC worked with an independent specialist to conduct a statistical meta-analysis using blinded data (see ref. 9 for a review). We

aggregated results from our 7 independent evaluations of 170 components from 130 discretely funded research projects in natural and social science, undertaken in Africa, Asia, Latin America, the Caribbean and the Middle East¹⁰. This revealed three things.

Southern-only research is high quality.

Research housed wholly in the global south proved scientifically robust, legitimate, important and well-positioned for use. Researchers in the region scored well across each of these criteria (higher, on average, than the northern and north-south-partnered research in our sample). In other words, those most closely linked to a particular problem seem to be well placed to develop a solution. (See Figure S3 in Supplementary Information.)

This finding challenges assumptions that researchers in the north automatically strengthen the capacity of partners in the south¹¹. There are many positive reasons to support north-south research partnerships, but the data suggest that we must be strategic to optimize their impact.

Capacity strengthening and excellence go hand in hand.

Too many funders assume that research efforts in which teams receive training and skills development inevitably produce poor-quality research. The meta-analysis found no such trade-off. In fact,

we found a significant positive correlation between scientific rigour and capacity strengthening.

This suggests that research requiring a focus on capacity strengthening need not be avoided out of a desire for excellence. Indeed, it implies that the two can go hand in hand.

Research can be both rigorous and useful.

In the fast-paced world of policy and practice, findings need to get to the right people at the right time, and in ways that they can use (see 'Co-producing climate adaptations in Peru'). We often hear of tension between sample saturation or trial recruitment and the decision-making cycle of policymakers or industry implementers. Happily, the meta-analysis found a strong positive correlation between how rigorous research is and how well it is positioned for use.

This finding builds the case for investing in scientific integrity, in even the most applied and translational programmes.

FOUR CONCERNS

We have four main concerns about RQ+ and how it can be refined and adapted for broader application.

First, bias is baked into our study. We used our own tool to examine research we had already supported. RQ+ focused our post-hoc evaluations on the values that ►



Farmers in Pampallacta, Peru, inspect harvested potatoes.

CASE STUDY

Co-producing climate adaptations in Peru

More than 500,000 people live in the Mantaro Valley in central Peru, where agriculture is the main source of income. The valley's small-scale farmers provide most of the vegetables and grains consumed in the capital, Lima, but are struggling to respond to the increasing frequency and intensity of extreme droughts, heavy rainfalls and frosts.

Using new and creative combinations of physical measurements and participatory engagement methods such as community mapping, the Geophysical Institute of Peru in Lima is providing a clearer picture of how the climate has changed in the region. This research is informing local policy and guiding adaptation actions. The project mapped hotspots across the region that were susceptible to climate change, and convened discussions with farmers and fishers about how they could adapt schedules and techniques to minimize its impact.

The team did not rush to publish the research in top-tier Western journals, partly because of the English-language barrier but largely because of the urgency of the problem. The research outputs needed to be immediately understandable and usable, so the team rapidly published its findings in working papers and reports (many of which were collected in a Spanish-language book^{13,14}). These were immediately accessible to those in local government who needed the evidence to steer the response. As such, predominant metrics do not capture the value of this work.

The RQ+ review shone a different light on this project and its achievements. It scored highly for integrity (including innovative blending of techniques for knowing the climate), for being legitimately grounded in local needs and knowledge, for addressing an urgent problem, and for focusing on uptake and action. **J.L. & R.M.**

► matter to our organization. The method examines our objectives and priorities, as we define them. Some would counter that it reifies them.

Second, this tool, much like all others, could have a distorting effect. For instance, by asking reviewers to examine integrity and legitimacy — issues that we identify as fundamental to our success — we turned their attention away from other factors, such as productivity (volume of publications and outputs) and cost-efficiency.

Third, there is the risk that RQ+ results become isolated if they are not comparable with the prevailing measures of research

quality used by the global research enterprise. Is RQ+ just another demanding hurdle for researchers in the global south? That's a question we are still working to answer.

Fourth, RQ+ costs more and takes longer than asking two or three peers to offer their opinions. Our hunch is that it takes almost twice as much time and money, largely because it requires empirical data collection by the evaluators. For us, that is time and money well spent: the results help us to hone our approach to funding and engagement.

These concerns will guide our efforts to improve RQ+, as will input from our peers and partners.

MORE LIKE THIS

What next? If the trillions of dollars being invested in research globally each year¹² are to make a difference, we must do better than crude quantification of citations, as the Leiden Manifesto¹ and the San Francisco Declaration on Research Assessment² have made clear.

We believe RQ+ presents a practical solution. The approach and findings of our meta-analysis now need replication in other contexts. At IDRC, we are planning another retrospective assessment in 2020. We are excited by what progress and shifts it might uncover. We are already looking at ways we can use RQ+ for grant selection, monitoring the progress of individual projects, and communicating our organizational objectives to funding partners and applicants.

Similarly, we encourage other funders and institutions to improve their evaluations in three ways: consider research in context; accept a multidimensional view of quality; and be systematic and empirical about evidence collection and appraisal. It's time science turned its greatest strengths on itself — experiment, appraise, debate and then improve. ■

Jean Lebel is president and **Robert McLean** is senior programme specialist at the International Development Research Centre, Ottawa, Canada.
e-mails: jlebel@idrc.ca; rmclean@idrc.ca

- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. & Rafols I. *Nature* **520**, 429–431 (2015).
- Annual Meeting of the American Society for Cell Biology. *San Francisco Declaration on Research Assessment* (ASCB, 2012).
- Canadian Institutes of Health Research. *Evaluation of CIHR's Knowledge Translation Funding Program* (CIHR, 2013).
- Wilsdon, J. et al. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management* (Higher Education Funding Council of England, 2015).
- Holmes, B. *The Rise of the Impact Agenda*. Fuse International Conference on Knowledge Exchange in Public Health, Newcastle, UK (27 April 2016); available at <https://go.nature.com/2yujm39>
- Amano, T., González-Varo, J. P. & Sutherland, W. J. *PLoS Biol.* **14**, e2000933 (2016).
- Amano, T. & Sutherland, W. J. *Proc. R. Soc. B* **280**, 20122649 (2013).
- Ofir, Z., Schwandt, T., Duggan, C. & McLean, R. *Research Quality Plus (RQ+): A Holistic Approach to Evaluating Research* (IDRC, 2016).
- Gurevitch, J., Koricheva, J., Nakagawa, S. & Stewart, G. *Nature* **555**, 175–182 (2018).
- McLean, R. & Sen, K. *Making a Difference in the Real World? A Meta-Analysis of Research for Development* (IDRC, 2018).
- Bradley, M. in *Putting Knowledge to Work* (ed. Mougeot, L. J. A.) Ch. 2 (Practical Action, 2017).
- '2017 Global R&D Funding Forecast' *R&D Mag.* (Suppl.) Winter 2017, 3–6 (2017).
- Instituto Geofísico del Perú. *Eventos meteorológicos extremos (sequías, heladas y lluvias intensas) en el valle del Mantaro*. Vol. 1 (IGP, 2012).
- Instituto Geofísico del Perú. *Manejo de riesgos de desastres ante eventos meteorológicos extremos en el valle del Mantaro*. Vol. 2 (IGP, 2012).

Supplementary information accompanies this article: see go.nature.com/2ja2dfr.