RQ+

Research Quality Plus

A Holistic Approach to Evaluating Research

Zenda Ofir
Independent Evaluator
Honorary Professor, University of Stellenbosch, South Africa

Thomas Schwandt
Professor, University Of Illinois, Urbana-Champaign, USA

Colleen Duggan
Senior Program Specialist, Policy and Evaluation Division, International Development Research Centre, Canada

Robert McLean
Senior Program Specialist, Policy and Evaluation Division, International Development Research Centre, Canada
Lead Evaluator, Canadian Institutes of Health Research, Canada

IDRC | CRDI
International Development Research Centre
Centre de recherches pour le développement international

Canada

Acknowledgements

The authors graciously acknowledge the critical review of this work received from colleagues around the world. Insight, guidance, and correction from the following peers has sharpened our thinking and expanded our perspective. Any errors, of fact or reason, belong to the authors alone.

Sarah Bowen
University of Alberta, Canada

Juliana Bravo
World University Service, Canada

Fred Carden
Using Evidence Inc, Canada

Fanie Cloete
University of Johannesburg, South Africa

Leandro Echt
Politics & Ideas, Argentina

Amy Etherington
IDRC, Canada

Osvaldo Feinstein
Universidad Complutense de Madrid, Spain/Argentina

Trisha Greenhalgh
University of Oxford, UK

Luize Guimaraes
Oxford Policy Management, Mozambique

Brent Herbert-Copley
Social Sciences and Humanities Research Council of Canada, Canada

Lise Korsten
University of Pretoria, South Africa

John Lavis
McMaster University, Canada

Enrique Mendizabal
OnThinkTanks.org , Peru

Heidi Monk
Artisanal Gold Council, Canada

Savi Mull
Global Development Network, India

Claire Nicklin
McKnight Foundation, Ecuador

Umezuruike Linus Opara
Stellenbosch University, South Africa

Andrew Shaw
Department for International Development, UK

Suneeta Singh
Amaltas, India

Matthew Smith
IDRC, Canada

Ian Thornton
UK Collaborative on Development Sciences, UK

Cecilia Tortajada
National University of Singapore, Singapore

Vanesa Weyrauch
Politcs & Ideas, Argentina

Tricia Wind
IDRC, Canada

Ofir, Schwandt, Duggan, McLean. 2016. Ottawa, Canada.

For inquiries please contact the IDRC Evaluation team.
evaluation@idrc.ca

www.idrc.ca

INTERNAL INFLUENCES
RESEARCH LEGITIMACY
RESEARCH INTEGRITY
RQ+
RESEARCH IMPORTANCE
POSITIONING FOR USE
RESEARCH CYCLE
EXTERNAL INFLUENCES

Introduction

Assessing the quality of research is hardly a new or novel idea. Researchers have long debated the best criteria and means for determining the scientific rigor and significance of empirical studies in the natural, social, and behavioral sciences. What is different is that the concern with research quality has taken on renewed meaning for academic institutions, governments, foundations, nonprofit agencies, and nongovernmental and intergovernmental organizations in light of the intersecting global interests in quality standards, performance measurement, accountability, evidence-based policy and practice, and value for money in research investments.

This set of global concerns has meant a more acute focus on the merits and shortcomings of traditional deliberative (e.g. peer review) and current analytic (e.g. bibliometric) means of evaluating the scientific merit of research (Feller and Stern, 2006; Wilsdon et al., 2015). In turn, it has encouraged a resurrection of interest in the importance of studying knowledge utilization and knowledge exchange processes, drawing specific attention to defining and assessing the impact or benefit of research to society. And, it has resulted in a more profound appreciation of the fact that science and its concern for generating empirically warranted descriptions and explanations of the social and natural worlds can no longer be considered a largely academic enterprise divorced from societal concerns about social goals.

Interest and debate about the quality and use of research, as well as the relationship of research undertakings to social goals and values, have been particularly intense in the field of internationally funded research on the development priorities of low-to-middle-income countries. It is in this arena, namely the arena of research for development (R4D),[1] that the approach to evaluating research quality discussed in this report was designed.

The International Development Research Centre (IDRC) in Ottawa, Canada, primarily funds and facilitates global South-based research for development.[2] While determining and enabling high quality research is at the core of its work, the Centre also has a strong evaluation function. It strives to evaluate its most significant investments using numerous centralized

1 This is not the only site where this debate is taking place. Another arena where vigorous interest in these issues is very evident is health care, e.g. Panel on Return on Investment in Health Research (2009) and Guthrie et al. (2013).

2 The mandate set out in the IDRC Act of 1970 directs the Centre "to initiate, encourage, support and conduct research into the problems of the developing regions of the world and into the means for applying and adapting scientific, technical and other knowledge to the economic and social advancement of those regions."

and decentralized evaluative mechanisms, structures, and platforms.[3] Any research assessment framework it employs has to accommodate the heterogeneity of its programs, both in terms of the diversity of *issues* IDRC addresses, such as technology for food security, global health, climate change and inclusive economic growth, and in terms of the types of *activities* it supports, such as research, capacity strengthening, promoting use of research for policy and practice, constituency and network building, and its partnering as a peer, mentor, or broker.

In 2012, IDRC launched an effort to develop a new approach to evaluating the quality of the research it funds. The result is presented hereafter as the "RQ+ Assessment Framework". This work was motivated by IDRC's desire to advance global research evaluation practice and, more pragmatically, by the need to bring a degree of standardization and transparency to the assessment of research quality – an important emphasis of its evaluative accountability exercise, its External Reviews.[4]

This report describes, in brief, the rationale behind the development of this new approach to evaluating research quality. It then discusses in some detail the assessment instrument itself, followed by a section on lessons learned from the first effort to implement it in the External Reviews. In conclusion, the report reflects on potential uses of the instrument as well as ideas for its further refinement.

What is discussed here is the inaugural version of the RQ+ Assessment Framework. It is currently undergoing revision and adaptation by IDRC for its own specific circumstances and uses. Readers – both funders of research and researchers engaged in evaluating research quality – are encouraged to treat this Framework as a dynamic, evolving tool that they can adapt and modify for their specific purposes.

We expect that this approach to evaluating research quality will be improved by the practitioners who use the Framework and through further examination by scholars concerned with issues of research quality and use.

3 Comprehensive details on the IDRC approach to evaluation are available at: idrc.ca/EN/Programs/Evaluation/Pages/default.aspx.

4 IDRC External Reviews have four foci: i) coherence, effectiveness, and appropriateness of program implementation strategies; ii) quality of the research supported by each program under review; iii) relevance and significance of program outcomes; and iv) issues for consideration by IDRC's Board of Governors and senior management.

PART 1
Rationale

Two sweeping considerations characterize the traditional picture of what is involved in assessing research quality. First, the design and evaluation of research have long been thought to be tasks uniquely within the province of scientists. Scientific values, such as transparency, objectivity, attention to empirical evidence, and intellectual honesty, guide the enterprise as a whole. In addition, specific criteria are spelled out for determining the rigor of research design and implementation, data collection, the reliability and internal and external validity of scientific claims, as well as the value of scientific theories in terms of their logical consistency, reproducibility, generativity, and logical falsifiability. Assessing the quality of scientific research in light of these values and criteria has long relied on peer review, a mechanism increasingly supplemented in the past decade or so by bibliometric and other scientometric analytic methods and, to a lesser extent, by reputational studies. This overall approach to evaluating research has largely governed the way scientists think, irrespective of the type of research under consideration – be it applied, basic, use-inspired, clinical, development, or experimental.

Second, it is generally held that determining the scientific merit of research does not include evaluating its use, uptake, and impact. In other words, research quality is traditionally judged on the basis of scientific values and criteria alone. To borrow a phrase, this is science-centric thinking – that is, the assessment of research is solely a scientific matter that takes place in isolation from the context of its use. (McNie *et al*., 2015). While research assessment increasingly examines the impact of research beyond the academy as well as return on investment (World Bank and Elsevier, 2014; DFID, 2014), these concerns are, strictly speaking, non-academic and non-scientific. As the argument goes, this is because they are located in the complex and value-laden world of policymaking and practice that lies outside the closed system of science.

This traditional picture of what assessing research quality entails is undergoing significant examination and change. Concerns emanating from the scientific community are directed at whether citation counts and journal impact factors are reliable and valid measures of research quality (Eyre-Walker and Stoletzki, 2013; University of Gothenburg, 2013). There is also worry that performance incentives in scholarly environments based on such measures can hamper the type of boundary-spanning and engaged scholarship that some regard as critical to advancing innovation. Fears that bibliometric measures to assess research are being misused or misinterpreted, and even damaging the system of research that they are designed to assess and improve, led to the publication of the *Leiden Manifesto for Research Metrics*, a set of ten principles to guide the responsible use of these quantitative measures in the respected scientific journal, *Nature* (Hicks *et al*., 2015).[5] This is not to say that the scientific community wishes to relinquish its longstanding investment in, and authority over, assessing the scientific merit of its investigations. However, there is concern that metrics traditionally used for judging the quality of research within research institutions may not be the metrics most useful for judging the potential societal benefit of research (Sarewitz, 2011).

In addition, a growing body of research is finding that research utilization is not a matter of dissemination and translation but rather a complex, iterative, interpretive, interactive, and social process involving linkages and exchanges among researchers, policymakers, and practitioners (Nutley *et al*., 2007; Lavis et al., 2006; Greenhalgh and Wieringa, 2011; Bowen and Graham, 2015). Studies in the sociology of scientific knowledge production are demonstrating that the traditional view that science is largely an academic endeavor, divorced from society, is giving way to a variety of understandings about how the two are entangled. Science is now understood as a social

5 The ten principles of the Manifesto are: (1) quantitative evaluation should support qualitative, expert assessment; (2) measure performance against the research missions of the institution, group or researcher; (3) protect excellence in locally relevant research; (4) keep data collection and analytical processes open, transparent and simple; (5) allow those evaluated to verify data and analysis; (6) account for variation by field in publication and citation practices; (7) base assessment of individual researchers on a qualitative judgment of their portfolio; (8) avoid misplaced concreteness and false precision; (9) recognize the systemic effects of assessment and indicators; (10) scrutinize indicators regularly and update them.

process influenced by both individual and societal values and norms. A strong focus on the context of the application of science is accompanied by awareness of the fact that multiple actors are involved in the knowledge production process, and that this has consequences for the kinds of knowledge produced and how knowledge communication takes place (Nowotny *et al.,* 2003).

For example, the salience and legitimacy of research processes and products as perceived by audiences other than scientists are now considered important in understanding the production of scientific knowledge. Even "accurate" scientific information that is seen as irrelevant to the needs of particular decision makers or that ignores their perspectives and values is likely to be disregarded (Clark *et al*., 2006). Several scholars studying research evaluation in The Netherlands, put the case this way:

> *"Most current research is produced in a complex socio-economic context in which demands are made by a variety of social actors; ... [it] addresses complex questions (for example AIDS, global warming, migration, cultural identity); is often multi-, inter- and/or transdisciplinary and is conducted in a context in which experts with different backgrounds, knowledge and expertise operate and different demands and interests have to be negotiated.... This complexity requires a different approach to evaluation than the traditional peer review that mainly emphasizes scientific excellence and relies on publications in high impact journals for its primary indicators. [Q]uality in our approach is defined as a multidimensional concept which includes the expertise of stakeholders in different social domains" (Spaapen et al., 2007).*

Finally, in the field of international development research, it is widely acknowledged that research endeavors address "wicked problems"; that research undertakings are nested in socio-economic, political, and cultural circumstances; that multiple agents (including partnerships and networks) and their agendas are involved in the production of scientific knowledge; and that mechanisms that generate impacts of both interventions and of research itself are variable, unpredictable, and often contested. All of this is to say that it is generally understood that the assessment of research unfolds in highly context-dependent and often contested settings. As displayed in Box 1, the circumstances under which the research that IDRC funds unfolds, and the kinds of research it funds, provide a good example of the kind of context-dependency that must be considered a key influence in evaluating research quality.

The principal lesson drawn from the current discourse on assessing research quality is that researchers are exploring how an abiding interest in scientific values as the basis of assessing research quality might be effectively combined with considerations of user involvement in, and user-oriented criteria for, judging quality. Thus, the RQ+ Assessment Framework described in Part 2 of the paper is underpinned by the following two key issues:

1

It reflects the fact that scientific merit is a necessary but insufficient condition for judging research quality. Scientific expertise, while highly valued, is not the only expertise that matters. Assessment must also take into account whether research is perceived as salient and legitimate in the eyes of multiple stakeholders and potential users, and it must attend in a significant way to questions of the effectiveness of the research in terms of its relevance, use, and impact. The latter kind of assessment relies on the expertise of relevant stakeholders in judging the legitimacy and relevance of the research undertaking.

2

It acknowledges that the determination of research effectiveness is not solely in the hands of the researchers, or of the research funders or research project managers. Matters of research use, influence, and impact are not necessarily within their immediate sphere of control. These groups do, however, play a significant role in learning about and establishing, to a certain extent, those conditions that may enhance users' interest in, and use of, the research. It is therefore reasonable to hold researchers, research managers, and research funders responsible or accountable for the extent to which research is well-positioned for use.

These considerations are reflected in Figure 1, which highlights that research quality and research effectiveness are not strictly bounded concepts. In the minds of many stakeholders

BOX 1 The IDRC "context" for research

Attributes of IDRC-funded research

- Provides "use-inspired"[6] research that is policy/practice relevant, problem focused, solution oriented, or informed by national or local priorities.
- Produces multi-, inter- or transdisciplinary research, sometimes across disparate fields.
- Uses mixed methods primarily.
- Addresses complex and integrative problems.
- Displays sensitivity to, and respect for, local voices, knowledge, and contexts, specifically in the Global South.
- Displays sensitivity to risk for vulnerable individuals and societies, as well as fragile institutions, systems, and contexts.

Attributes of IDRC's supplementary support to researchers

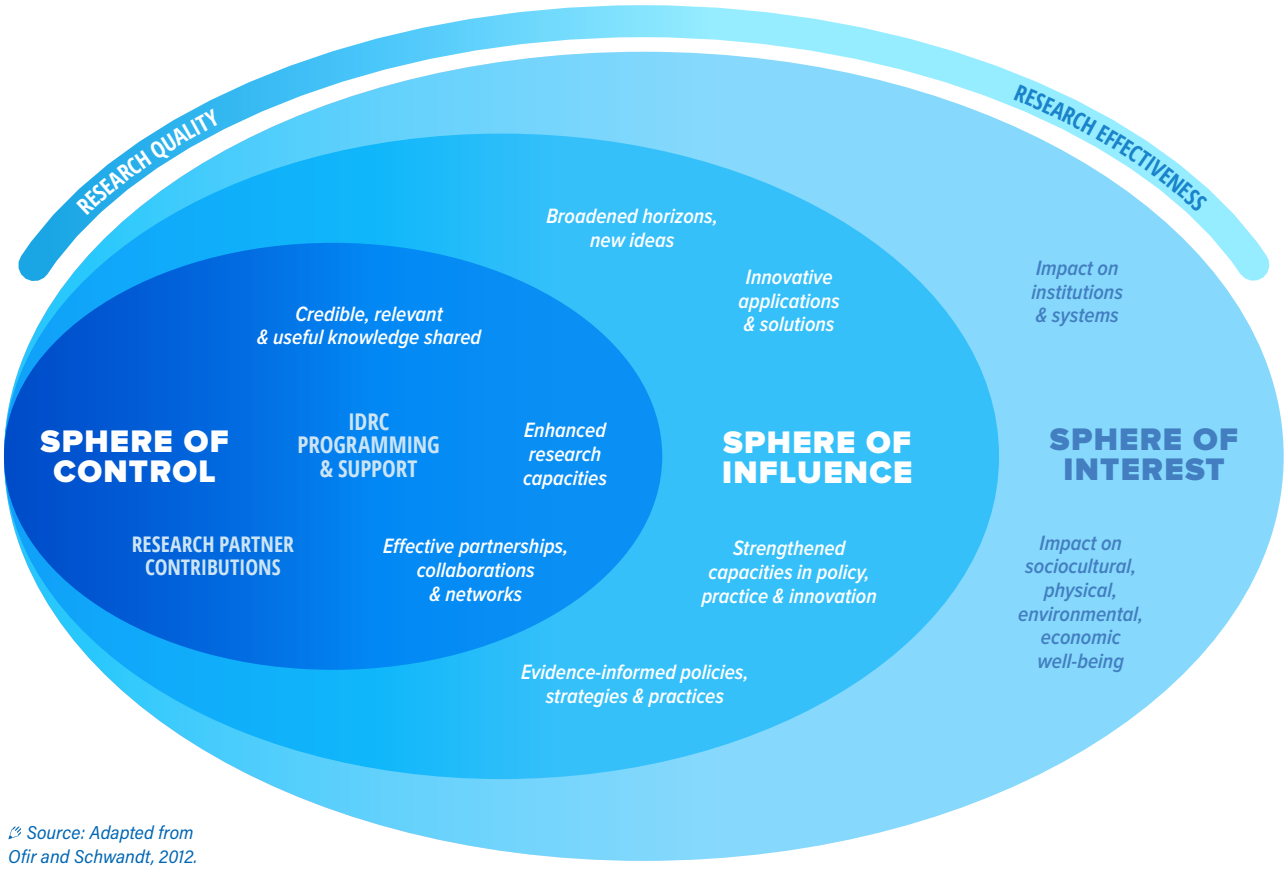
- Strengthens research capacities of individuals and institutions, through long-term investment.
- Takes risks by, for example, supporting entirely new fields of work, engaging with complex regional or global challenges, and supporting work in conflict-ridden or poverty-stricken environments.
- Encourages knowledge generation in the global South.
- Facilitates research networks, research to policy linkages, and access to resources.
- Builds constituencies and networks for change.
- Targets changes in policies, practices, institutional systems, and technologies.
- Partners as a mentor, advisor, peer, and broker.

Source: Ofir and Schwandt, 2012.

they overlap and blend into each other. The three spheres in Figure 1 depict the research development process and the non-linear progression from production to use. The spheres show that the technical conceptualization and execution of the research, and the production of research outputs, are largely under the control of the researchers, funders, and program managers. The influence that the research might have on policy and practice is sometimes, but not always, in their sphere of influence. And while the longer term development impacts to which the research contributes may well be within their sphere of interest, these actors tend to have very little, if any, influence over how the research products are used and transformed, and when and how development impacts will emerge. Thus the fact that the uptake, use, influence, and impact of research depend on the interaction of multiple actors, agencies, and socio-political circumstances has direct implications for how research quality and effectiveness can be defined, tracked, and evaluated.

6 Stokes (1997) coined the term "use-inspired".

FIGURE 1 Research quality, research effectiveness, and spheres of control



Source: Adapted from Ofir and Schwandt, 2012.

RQ+ at a glance

RQ+ Assessment Framework

The RQ+ Assessment Framework provides a systems-informed approach to defining and evaluating the quality of research, and its positioning for use and impact. It allows tailoring to context, values, mandate and purpose, and can support planning, management and learning processes at any stage in the lifetime of a research project, program or grants portfolio.

RQ+

Research Quality Plus

A Holistic Approach to Evaluating Research

Framework Components

The RQ+ Assessment Framework consists of three main components:

1. KEY INFLUENCES

Constraining and enabling contextual influences - within or external to the research effort - most likely to affect research performance are identified.

The rating of the key influences using rubrics and a three point scale (e.g. low, medium, high) establishes a risk profile that is used to inform the quality assessment.

The key influences can be 1) constraining (negative) or 2) facilitating / enabling (positive)

Examples from IDRC experience:

1) Maturity of the research field
2) Research capacity strengthening
3) Risk in the data environment
4) Risk in the research environment
5) Risk in the political environment

2. DIMENSIONS & SUBDIMENSIONS

The four dimensions and their subdimensions encapsulate the quality assessment criteria.

Tailored for IDRC:

1. ***Research Integrity***
2. ***Research Legitimacy***
   - *2.1 Addressing potentially negative consequences*
   - *2.2 Gender-responsiveness*
   - *2.3 Inclusiveness*
   - *2.4 Engagement with local knowledge*
3. ***Research Importance***
   - *3.1 Originality*
   - *3.2 Relevance*
4. ***Positioning for Use***
   - *4.1 Knowledge accessibility & sharing*
   - *4.2 Timeliness and actionability*

3. EVALUATIVE RUBRICS

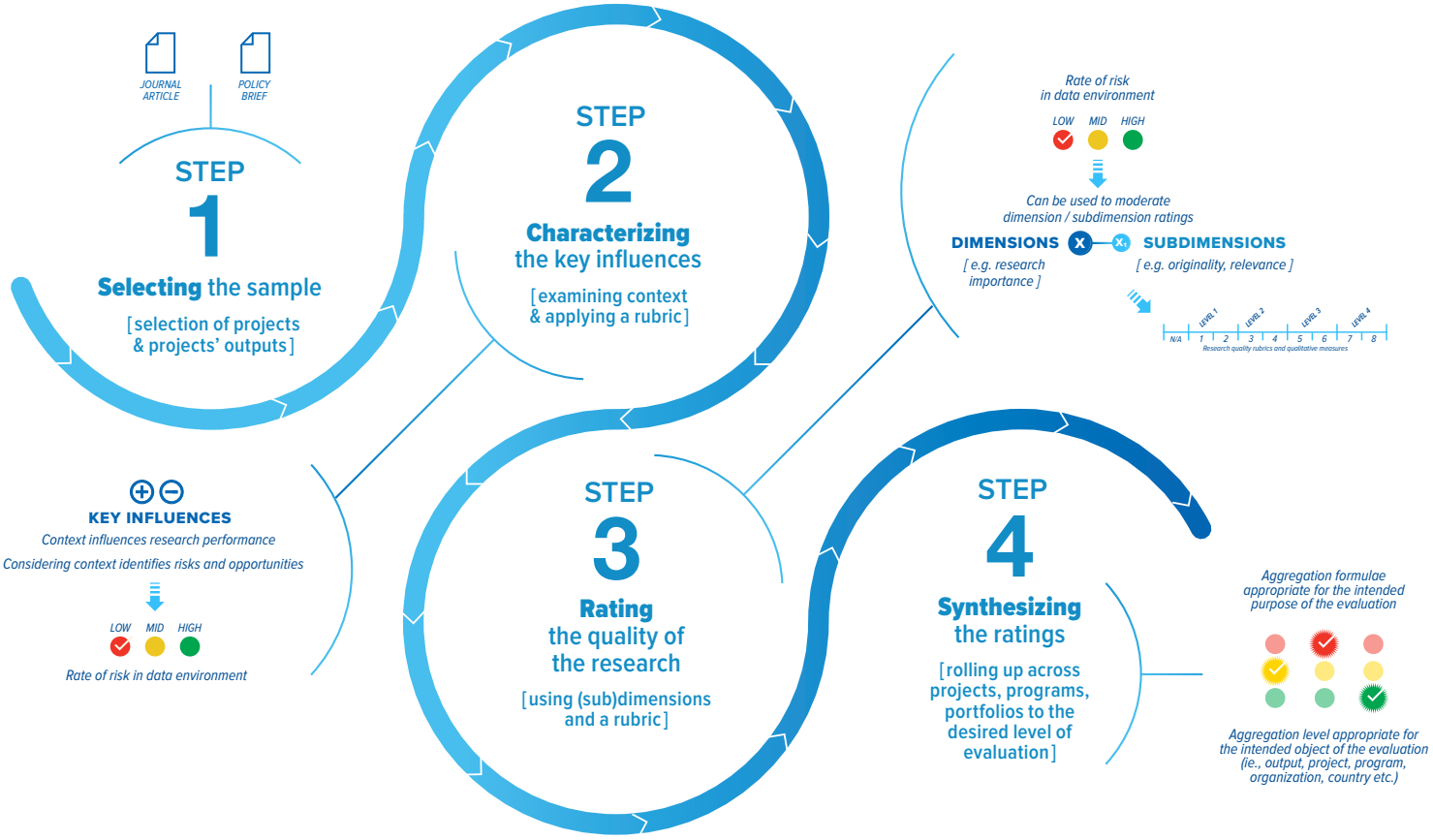
Performance is characterized using customizable research quality rubrics.

Characterization of each key influence, dimension and subdimension is done using tailored rubrics that combine quantitative and qualitative measures.

Ratings on an 8 point scale show four levels of performance (or progress). This is an example. Scales should be created to fit a purpose or intention.

Framework Roadmap

The components of the RQ+ Framework form a dynamic relationship



The application outlined here describes how the RQ+ Assessment Framework was applied by IDRC in summative External Program Reviews.

Authors: Zenda Ofir / Thomas Schwandt / Colleen Duggan / Robert McLean

IDRC | CRDI
International Development Research Centre
Centre de recherches pour le développement international

PART 2
The RQ+ Assessment Framework

The Framework Components

The RQ+ Assessment Framework consists of the three main components. They include: i) key influences; ii) research quality dimensions and subdimensions; and, iii) evaluative rubrics. The discussion that follows addresses selected aspects of these components. The components in their entirety, including detailed descriptions of the key influences and each of the quality dimensions, can be found on the IDRC Evaluation webpage, www.idrc.ca. The previous page's 'RQ+ at a glance' infographic, provides a quick reference summary.

The key influences

This component highlights those influences – either within the research endeavor or in the external environment – most likely to affect the quality of the research. Such influences cannot be fully predicted if the assessment is ex ante, but this sensitivity to context is one of the most novel aspects of the Framework. The key influences are meant to help evaluators, managers, funders, and others to make meaningful and systematic considerations of the enabling or constraining factors of the research and the risk profile of the project, program, or portfolio, and to incorporate these to the extent possible into their assessments.

IDRC identified five main influences on research quality. [7]

1

MATURITY OF THE RESEARCH FIELD— The extent to which well-established theoretical and conceptual frameworks exist and from which well-defined hypotheses have been developed and subjected to testing, as well as a substantial body of conceptual and empirical research in the research field.

2

RESEARCH CAPACITY STRENGTHENING— The extent to which the research endeavor or project focuses on strengthening research capacities through providing financial and technical support to enhance capacities to identify and analyze development challenges, and to conceive, conduct, manage, and communicate research that can address these challenges.

3

RISK IN THE RESEARCH ENVIRONMENT— The extent to which the organizational context in which the research team works is supportive of the research, where "supportive" refers, for example, to institutional priorities, incentives, and infrastructure.

4

RISK IN THE POLITICAL ENVIRONMENT— The extent of external risk related to the range of potential adverse factors that could arise as a result of political and governance challenges, and that could affect the conduct of the research or its positioning for use. These range from electoral uncertainty and policy instability to more fundamental political destabilization, violent conflict, or humanitarian crises.

5

RISK IN THE DATA ENVIRONMENT— The extent to which instrumentation and measures for data collection and analysis are widely agreed upon and available, and the research environment is data rich or data poor.

7 Many influences were identified during the development process in IDRC, but these were eventually narrowed to five. In other organizations these factors might be very different.

BOX 2 Examples of key influences, with their rubrics

Maturity of the research field

☐ **(1) ESTABLISHED FIELD**
Well-established and recognized theoretical and conceptual frameworks, a substantial body of conceptual and empirical research, discernable outlets (journals, conferences, curricula) and the presence of a vibrant corps of experienced researchers all characterize the field.

☐ **(2) EMERGING FIELD**
Recognized by members and non-members, with a discernable body of work , theory, and practice, discernable outlets, and a modest body of active researchers who easily associate with the field, and recognize each other.

☐ **(3) NEW FIELD**
The field of research has a very limited theoretical or empirical knowledge base that is still debated or rapidly changing, is not widely recognized, has no dedicated journals or academic programs, and with only a few active researchers seeking recognition.

Research capacity strengthening

☐ **(1) LOW FOCUS**
Research capacity strengthening is inexistent or is a low priority in this project.

☐ **(2) MEDIUM FOCUS**

☐ **(3) STRONG FOCUS**
Research capacity strengthening is an important priority in this project alongside other equally important priorities and intentions.

Risk in the data environment

☐ **(1) LOW RISK**
Instrumentation and measures for data collection and analysis are widely agreed upon and available; the data environment is well developed, stable, and data rich.

☐ **(2) MEDIUM RISK**

☐ **(3) HIGH RISK**
Instrumentation and measures for data collection and analysis are not available; the research activities are conducted in severely underdeveloped, unstable, and/or data-poor environments.

Risk in the research environment

☐ **(1) LOW RISK**
Research environment – institutional priorities, incentives, facilities, etc. – is established and supportive.

☐ **(2) MEDIUM RISK**

☐ **(3) HIGH RISK**
Research environment is weak or largely under-developed, and not supportive.

Risk in the political environment

☐ **(1) LOW RISK**
Stable political environment with established governance practices, no conflict, etc.

☐ **(2) MEDIUM RISK**

☐ **(3) HIGH RISK**
Very unstable or volatile political environment with weak governance practices, conflict, etc.

The research quality dimensions

There are four principal quality dimensions in the RQ+ Assessment Framework. In our view, these are not discrete. The dimensions are closely interrelated. We have identified and demarcated these dimensions out of a desire for deeper study of the particular issues they represent.

1

RESEARCH INTEGRITY— Considers the technical quality, appropriateness and rigor of the design and execution of the research as judged in terms of commonly accepted standards for such work and specific methods, and as reflected in research project documents and in selected research outputs. Specified emphases include the research design, methodological rigor, literature review, systematic work, and the relationship between evidence gathered and conclusions reached and/or claims made. Peer reviewed and non-peer reviewed outputs undergo different assessment processes using different criteria.

2

RESEARCH LEGITIMACY— Considers the extent to which research results have been produced by a process that took account of the concerns and insights of relevant stakeholders, and was deemed procedurally fair and based on the values, concerns and perspectives of that audience. Legitimacy deals primarily with who participated and who did not; the process for making choices; how information was produced, vetted and disseminated; how well knowledge was localized, and if it respected local traditions and knowledge systems. This dimension also includes a subdimension that asks the assessor to consider the potentially negative consequences and outcomes for populations affected by the research, gender-responsiveness, inclusiveness of vulnerable populations, and engagement with local knowledge.

3

RESEARCH IMPORTANCE— Considers the importance and value to key intended users of the knowledge and understanding generated by the research, in terms of the perceived relevance of research processes and products to the needs and priorities of potential users, and the contribution of the research to theory and/or practice. Subdimensions include the originality and relevance of the research.

4

POSITIONING FOR USE— Considers the extent to which the research process has been managed, and research products/outputs prepared in such a way that the probability of use, influence and impact is enhanced. The uptake of research is inherently a political process. Preparing for it therefore requires attention to user contexts, accessibility of products, and 'fit for purpose' engagement and dissemination strategies. It also requires careful consideration of relationships to establish before and/or during the research process, and the best platforms for making research outputs available to given targeted audiences and users. Positioning for use calls for strategies to integrate potential users into the research process itself wherever this is feasible and desirable. Subdimensions include knowledge accessibility and sharing, actionability, and timeliness.

The first three dimensions – Research Integrity, Legitimacy and Importance – are the core *quality* features typically found in more or less developed forms in most research quality assessment frameworks. The fourth dimension – Positioning for Use – is less typical and is the *plus* (RQ+) feature of the framework. During the Framework development process, IDRC and its research partners determined that it would be reasonable to hold themselves accountable for taking steps to increase the likelihood that the research would be used – in other words, for positioning the research findings for influence and eventual impact.[10]

The evaluative rubrics

Characterizations for each key influence and performance levels for quality dimensions and subdimensions are based on customizable assessment rubrics that make use of both qualitative and quantitative measures. For the key influences (See Box 2), the rubrics set out descriptive language that is meant to guide the assessor in identifying particular characteristics or features that might be present in varying degrees in a particular risk or factor found in the context.

8 For the full set of rubrics see: IDRC, 2014.

9 As done for the IDRC 2014/2015 External Program Reviews.

10 The decision to include this dimension was based on the scope of IDRC's mission and mandate, and was also linked to the research programming time period that the Framework was meant to assess (a portfolio of projects spanning three to five years). Depending on their values and purpose, other funders or research organizations might deem this dimension to be unrealistic, or might try to track quality further out into the spheres of influence and interest.

For the quality dimensions and their subdimensions, the rubrics set out criteria for judgment, clarifying how performance is measured. Using graduated levels of achievement, the rubrics explain how good is "good" or how poor is "unacceptable". Box 2 provides selected examples of some (not all) of the rubrics for the quality dimensions.[8]

Synthesis of the rubric ratings in different permutations allows for performance to be classified into four levels (from "Unacceptable" to "Very Good") - as a snapshot of performance,[9] or to follow progress made over time when research capacities are being strengthened and/or projects or programs are on a trajectory towards research excellence.

BOX 3 Examples of the quality dimension and subdimension rubrics

Dimension 1.0: Research Integrity

UNACCEPTABLE	LESS THAN ACCEPTABLE	ACCEPTABLE TO GOOD	VERY GOOD
1 2	3 4	5 6	7 8
The research has little to no scientific merit. The defensibility of the approach is questionable. There are severe lapses in methodological rigor of literature review, data collection and data analysis.	There is evidence of efforts to meet methodological standards but the efforts do not fully succeed. There are major shortcomings in the justification for the choice of research design and methods.	Accepted methodological standards in the design and execution of the research are met.	The scientific merit is without question. There is evidence of exceptional thoroughness in the research design and all phases of research execution. The project could serve as an exemplar of what it means to achieve this criterion.

Dimension 2: Research Legitimacy; Subdimension 2.4: Engagement with Local Knowledge

NOT APPLICABLE	UNACCEPTABLE	LESS THAN ACCEPTABLE	ACCEPTABLE TO GOOD	VERY GOOD
	1 2	3 4	5 6	7 8
The nature of the research is such that local knowledge and engagement do not need to be taken into account.	Engagement with local contexts has been neglected during the research process. Several major weaknesses can be found, related to how research needs and questions were identified, local communities or populations engaged, local contexts and knowledge systems considered, and local benefits from the research process assured.	Local contexts and engagement have been considered during the research process, but some weaknesses remain related to how research needs and questions were identified, local communities or populations engaged, local contexts and knowledge systems considered, and/or local benefits from the research process assured.	Local context and engagement have been a focus in the research process. Few, if any, minor weaknesses remain related to how research needs and questions were identified, local communities or populations engaged, local contexts and knowledge systems considered, or local benefits from the research process assured.	Local context and engagement have been a clear and systematic focus in the research process. Research needs and questions were appropriately identified, local communities or populations engaged, local contexts and knowledge systems considered and respected, and local benefits from the research process assured.

Dimension 3: Research Importance; Subdimension 3.2: Relevance

UNACCEPTABLE	LESS THAN ACCEPTABLE	ACCEPTABLE TO GOOD	VERY GOOD
1 2	3 4	5 6	7 8
There is little or no evidence that the research might contribute to a local priority, a key development policy or strategy, or an emerging area that might demand solutions in the foreseeable future. Needs assessments and justification for the work are absent or unconvincing.	There is some evidence that the research might contribute to a local priority, a key development policy or an emerging area that might demand solutions in the foreseeable future. A focus on this area of work at this time appears sufficiently justified.	There is good evidence that the research might contribute to an important local priority, a key development policy or strategy, or an emerging area of some significance that might demand solutions in the near future. A focus on this area of work at this time has been well justified.	There is good evidence that the research is already recognized as having the potential to address a critical local priority, a key development policy or strategy, or an important emerging area that is highly likely to demand solutions in the near future. A focus on this area of work at this time puts the researchers at the cutting edge of an active and/or important field of work.

Dimension 4: Positioning for Use; Subdimension 4.2: Timeliness and Actionability

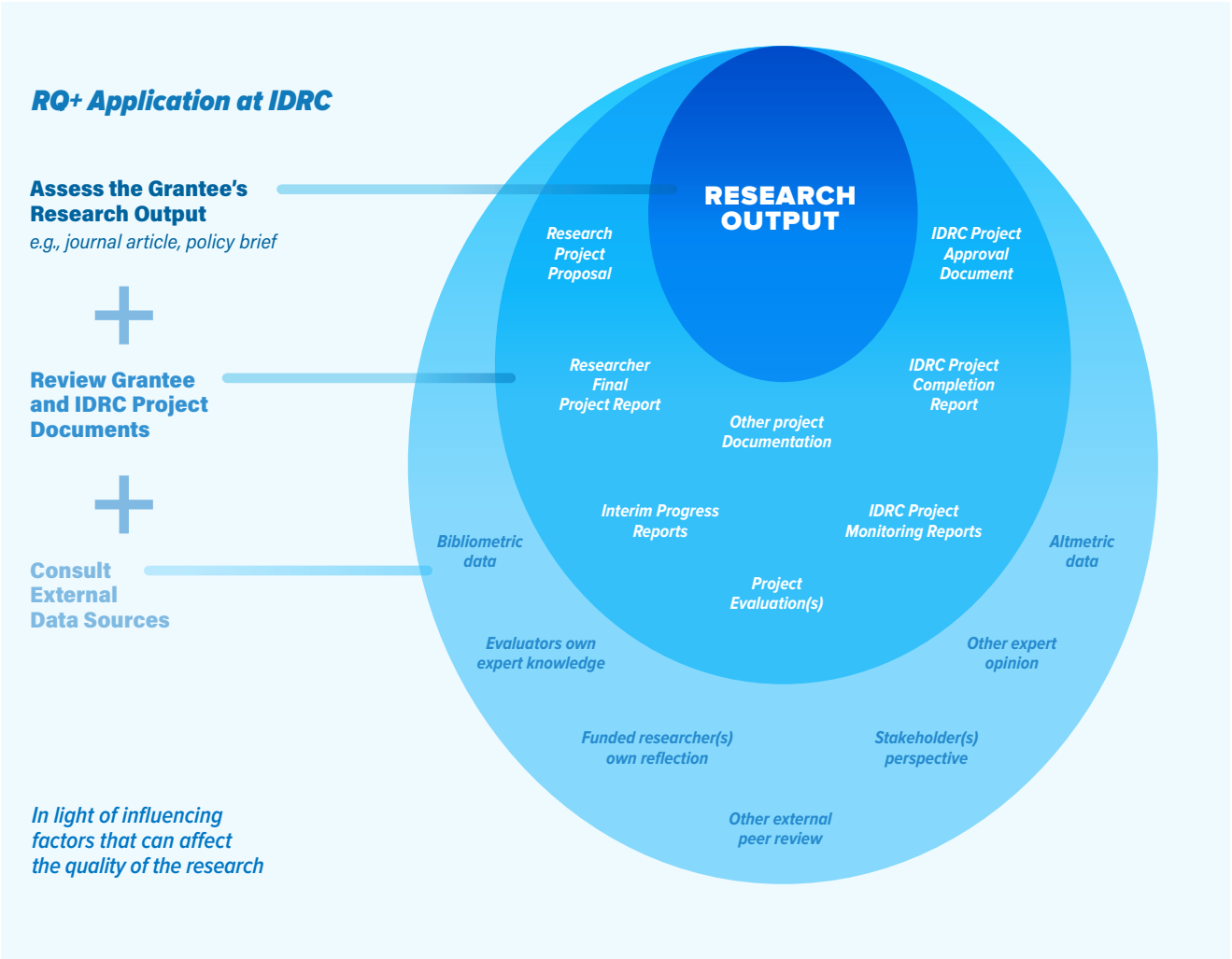
UNACCEPTABLE	LESS THAN ACCEPTABLE	ACCEPTABLE TO GOOD	VERY GOOD
1 2	3 4	5 6	7 8
There is little or no evidence that any analysis of relevant user environment was undertaken and that institutional, political, social, or economic contingences were considered.	There is evidence that some analysis of the user setting was under undertaken; however, consideration of is incomplete and, furthermore, the analysis is not accompanied by discussion of actual strategies or plans to move the knowledge to policy or practice.	There is evidence that the user environment and major contingencies have been examined and reflected upon and connected to strategies and plans for moving the research into policy or practice in a timely manner.	The analysis of the user environment and contingencies is exceptionally thorough and well-documented or articulated. There is evidence of careful prospective appraisal of the likelihood of success of strategies designed to address contingencies.

Using the Framework

Outputs are the unit of analysis of many research assessment frameworks that depend heavily on peer review. **The RQ+ Assessment Framework is based on the premise that a credible, balanced, and comprehensive assessment of research quality requires consideration of elements beyond the research outputs, to include important aspects of the context in which the research has been conducted, and the manner in which it has been managed.** The application of the framework therefore straddles contextual analysis, and output and research project assessment.

As in most research assessment processes, expert peer review plays an important role when applying the Framework. External experts from relevant fields of research are required to arrive at judgments on, for example, the extent to which each subdimension criteria has been met. They also gather data and evidence to better characterize the research context, and to triangulate and deepen their findings based on multiple methods and sources. During the IDRC's application of the Framework in the 2014/2015 summative External Reviews, project and program files were reviewed, and surveys and interviews conducted among key informants (see Figure 3). [11]

FIGURE 3 The data collection required to operationalize the RQ+ Assessment Framework at IDRC



11 Key informants could be experts in the field with knowledge of the context in which the work is being done, intended users, primary stakeholders such as project and program staff) and, ideally, where this is feasible and warranted, the research participants.

A stepwise approach to operationalizing RQ+

Using a tailor-made assessment instrument, IDRC operationalized the RQ+ Assessment Framework in four main steps, shown in Figure 4. They are briefly described below. [12]

FIGURE 4 Steps in the application of the RQ+ Assessment Framework

STEP 1

Selecting the Sample

The sampling strategy for a program portfolio is a two-tier process: i) selection of a sample of research projects in the portfolio that meet certain criteria; and from this project sample, ii) selection of a set of research outputs or products that serve as the focus for the assessment. Sampling in both cases can be purposeful or random. For a sample to be representative of a program portfolio, the reviewers might cluster and map projects according to selected attributes such as thematic focus, size of grant, support period, type of outputs, (potential) user audience or interest, research location, location of intended or actual outcomes, and the type of outcomes. They follow the same type of process when selecting the outputs.

STEP 2

Characterizing the selected projects based on key influencing factors

During Step 2, the reviewers use project documentation, their own expertise and discussions with key informants to characterize each of the selected projects according to the key influences most likely to affect the research quality. Reviewers characterize each key influence using a rubric (Box 2). Cluster diagrams or similar visual aids can be used to map and compare influences within and across projects and program portfolios, and to consider their effect on progress made along expected trajectories towards research excellence. These types of data visualizations also provide evaluators and program managers with the potential to systematically map levels and types of risk and consider strategies that can inform both the management and evaluation of portfolios.

12 More detailed information can be found in IDRC, 2014.

STEP 3

Rating the quality of the research

Step 3 in the assessment process involves rating the quality of outputs in the project sample using the research quality dimensions and subdimensions included in the Framework. Each subdimension is rated on an eight-point scale (examples in Box 3). The assessment rubrics were developed cognizant of IDRC's values and understanding of each of the quality dimensions and subdimensions. The criteria and standards for the assessment and the evidence base for their synthesis into overall evaluative judgments are both clarified and made more transparent by the rubrics.

The reviewers have to decide, or be instructed about, the level of consistency and standardization of criteria and the extent to which different weights are to be allocated to different dimensions (and subdimensions) across projects and programs. This is essential, and as discussed below, such decisions can be challenging

STEP 4

Synthesizing the ratings

In the final step of the assessment process, the reviewers aggregate the ratings given for each sub-dimension to reflect the ratings per dimension for each output. This is then rolled up, first to the project level, then to the program portfolio level. With appropriate formulas, ratings obtained from a set of projects can be aggregated up to program level per dimension, subdimension or set of influencing factors to highlight strengths or weaknesses in research quality dimensions, or in specific types of projects, or under specific contextual circumstances (as identified through analysis of the influencing factors). It is feasible for such roll-up to be done across different program portfolios or up to the level of the organization as a whole. [13]

Overall ratings of a program portfolio of projects can be recorded in a table. The eight point scale helps to classify performance in each dimension, with ratings moving from the lowest level (unacceptable) to the highest level (very good). Where quantitative ratings may be perceived as too provocative, color coding ("traffic lights") may serve as a substitute. In fact, qualitative data synthesis may be just as, if not more, important than quantitative ratings synthesis depending on the learning objectives of the assessment. Where this is the case, the qualitative reasoning underpinning the rubric guided ratings is the key data to synthesize and examine. In qualitative or quantitative synthesis we note aggregation efforts can be viewed as a composite. (In quantitative synthesis this means not necessarily as a mean, median, or mode.) Weighting of the composition is dependent on the context and objectives of the research projects, research program portfolio, or organization.

For management purposes, the roll-up of the key influencing factors and research quality dimensions and subdimensions can be used to identify patterns or potential correlations in the data. For example, a project cluster identified as high risk according to the influencing factor analysis can be compared with the rating scores or levels of research quality to determine if any specific factor or combination of factors influenced specific dimensions of research quality.

13 These levels of aggregation have not yet been tested by IDRC.

FIGURE 5 A fictional synthesis to identify project & program trends

Research Output		Fictional peer reviewed Journal Article PROJECT A	Fictional Book PROJECT B	Fictional Policy Brief PROJECT C	Fictional Government Report PROJECT D	Fictional Working Paper PROJECT E	Fictional Book PROJECT B	Fictional Policy Brief PROJECT C	Fictional peer reviewed Journal Article PROJECT A	Fictional peer reviewed Journal Article PROJECT A	Fictional peer reviewed Journal Article PROJECT A	Roll up per dimension or influencing factor
Stage 2: Characterizing the Influencing Factors of Projects Chosen for Review												
Key influence 1	Maturity of the Research Field	Emerging	Established	Established	Established	Established	Established	NA	Established	NA	Established	
Key influence 2	Researcher Capacity Strenghtening	Medium Focus	Low Focus	Medium Focus	Low Focus	Strong Focus	Strong Focus	NA	Low Focus	NA	Low Focus	
Key influence 3	Risk in the Data Environment	Medium	Low	High	Low	Medium	Low	NA	NA	NA	Low	
Key influence 4	Risk in the Research Environment	Medium	Medium	Medium	Low	Medium	High	NA	High	NA	Medium	
Key influence 5	Risk in the Political Environment	High	Medium	High	Low	Low	High	NA	High	Medium	High	
Stage 3: Rating the Research Quality Plus of the Outputs												
Research Integrity	Research Integrity	4	7	4	3	6	6	1	1	1	1	
Research Legitimacy	Addressing potentially negative consequences	1	6	3	3	1	1	1	1	1	1	
	Gender responsiveness	1	6	3	3	1	1	1	1	1	1	
	Inclusiveness	1	6	3	3	1	1	1	1	1	1	
	Engagement with local knowledge	3	8	1	3	1	2	1	6	1	7	
Research Importance	Originality	7	6	4	2	4	4	6	6	7	7	
	Relevance	7	6	7	6	7	8	6	7	7	7	
Positioning for Use	Knowledge accessibility and sharing	8	5	4	2	6	6	4	7	6	7	
	Timeliness and Actionability	8	7	6	3	6	6	6	7	7	8	
Roll up per Output												

Some Reflection on Trade-offs in the Framework's Design

The design of any research evaluation framework inevitably involves trade-offs. There is no silver bullet, and in developing the RQ+ Assessment Framework, several choices were made as a result of stakeholder feedback received during the design and initial testing phases.

We recognized that the comprehensiveness of the framework and the systematic, detailed work demanded in the assessment process would come at the expense of ease of use and desirable timeframes. However, compared with conventional peer review, the approach required to judge the research quality dimensions would enable greater clarity and richer, more nuanced and, hence, more accurate and transparent assessments, primarily through the use of rubrics and multiple sources of evidence that require the systematic integration and interpretation of qualitative and quantitative data.

We also regarded the flexibility built into the framework through the potential for customization of the subdimensions and rubrics as imperative for the reasons explained in the design rationale. Yet, this can complicate applications where strict standardization of measures is required, as well as the aggregation and comparison of ratings by different organizations and initiatives.

A third and related trade-off called for allowing the flexibility to give different weights to each of the quality dimensions and subdimensions (and key influences) at the behest of efforts to standardize across organizations. The extent to which weighting is done largely depends on the values of the organization applying the Framework. For an example, during the design phase at the IDRC, some program staff and reviewers asked to test the Framework queried whether the Research Integrity dimension should be given more weight out of concern that "classic scientific merit" would otherwise be devalued. Given IDRC's corporate policy on open access publishing, there is also an argument to be made to give higher ratings in the Positioning for Use dimension to projects and programs that provide for open access publishing. Yet this might sit uncomfortably with other programs, and cause disagreement with funding partners who regard research publication in prestigious peer-reviewed journals as the primary indicator of quality, even if those journals do not have open access policies.

The reviewers participating in the first application of the RQ+ Assessment Framework confirmed the merit of several of these decisions, but also brought light to the complications these trade-offs created in evaluation. The following section focuses on the lessons and experience of the application of the Framework at the IDRC.

PART 3

Lessons from the implementation of the RQ+ Assessment Framework by IDRC

The Context of First Application

External evaluation is a cornerstone of IDRC's overall evaluation system. External Program Reviews constitute one of the Centre's highest level accountability processes. At the time of this publication, they were structured as summative evaluations of each IDRC program, and conducted once every five years by a team of three independent and external subject area experts, managed by IDRC's Policy and Evaluation Division. The teams of reviewers were comprised of seasoned researchers and evaluators with extensive academic peer review experience. The Reviews have two main purposes. First, they support IDRC's accountability for program delivery and, accordingly, the IDRC Board of Governors is their primary user. Second, they are intended to provide new knowledge to program managers, thus assisting their management of grant portfolios.

External Reviews are comprehensive, summative evaluations that consider program implementation, performance, and outcomes. They address four issues: i) the coherence, effectiveness, and appropriateness of program implementation strategies; ii) *the quality of the research supported by each program under review*; iii) the relevance and significance of program outcomes; and iv) issues for consideration by IDRC's Board of Governors and senior management. Because research quality was one of these four foci, the External Reviews provided a unique opportunity to apply the RQ+ Assessment Framework and bring about a degree of standardization of judgment across program assessments.

In preparation for its application in the Reviews, the Framework was first pre-tested through a series of simulations based on programs already reviewed in 2009 and 2010. The simulations aimed to surface serious weaknesses before they could affect the set of Reviews that was about to be launched. During a three-month period in early 2014, four former external reviewers familiar with IDRC, the External Review process and the selected programs[14] applied the Framework and documented their experiences across a wide range of issues. These included the extent to which the framework enabled fair assessment, its suitability for different types and purposes of research, and the feasibility of its use for program-level assessment within the parameters and timeline of an External Review process. The simulation experiences were shared in a workshop with a group of IDRC middle and senior program managers. Based on lessons learned, the Framework was refined and applied in seven IDRC External Reviews.

The remainder of this paper will present IDRC's experience in addressing this issue during the most recent round of External Reviews, conducted between June 2014 and September 2015. The following subsections outline lessons learned in terms of successes, challenges, and potential future directions.

Reasons to be Optimistic

Embedded values become explicit

The RQ+ Assessment Framework was created through a reflective process within IDRC. This process aimed to produce assessment criteria which could encapsulate the organization's values and shared understanding of research quality. This meant the Framework itself guided the external reviewers towards issues that were pertinent, cherished, and specific to IDRC. This Review process, using an instrument that captured the essence of IDRC values, resulted in assessments that primary stakeholders perceived as very useful. The reviewers also benefitted. They felt comfortable making judgments on criteria which were identified as important to the main users, and which IDRC grant recipients were reasonably expected to pursue in their research.

14 One of the reviewers was also one of the designers of the RQ+ Framework.

The reviewers noted during their debriefings that the transferability of this aspect of the RQ+ approach held significant potential for application by other research entities. The overall structure and process of implementing the approach could be standard, yet the subdimensions and influencing factors have the potential to change to a lesser or larger extent to reflect the values or learning agenda of another organization. The implementation of the Framework could also have secondary benefits, such as bringing about a common understanding of values within and across organizations, and promoting desired goals in research management and administration.

The systematic nature of the approach strengthened evaluation processes and results

Here we note both process and product optimism. In terms of *process*, the RQ+ Assessment Framework provided detailed and clear direction to the reviewers. The extensive use of rubrics clarified the basis for the assessments and brought greater consistency to the judgments than typically experienced during conventional peer review processes. The reviewers perceived this as a valuable advantage, despite the significant level of analysis required to use the rubrics across each subdimension and influencing factor. Many reviewers suggested that this "guided process" actually simplified the assessment.

The Framework was also seen as strengthening the *result* of the External Review. It was given credit for helping to address (albeit not fully) a particular challenge in research evaluation, namely the ability to compare very different types of research. As articulated by one reviewer, the Framework gave "a sense of comfort when comparing the results of a randomized control trial with the results of an advocacy-driven or participatory action research project." Reviewers found this to be a novel and useful contribution. The reasons cited for this advantage included the integration of multiple data sources, the consistent set of dimensions and subdimensions on which all the research was evaluated, and the explicit recognition of the need to consider contextual factors.

The reviewers also expressed their appreciation for the fact that the systematic and detailed process of applying the Framework not only increased their confidence in the quality of their review, but also improved the chance of replication. This is an exciting observation, given the ongoing debates and doubts about the replicability of peer reviews.

Going beyond the research output enables precision, accuracy and richness

The External Review experience demonstrated that the RQ+ Assessment Framework facilitates precise and rich analysis as a result of the comprehensive nature of the approach. This advantage is exemplified in, what reviewers coined, the push to "go beyond the research output." While more complex, it is an improvement over other commonly used peer review systems that are restricted to the object under review. For example, a research paper submitted to a journal is assessed only on the merit of the paper. Likewise, in the case of a research grant application, it is likely that only a research proposal and possibly the track record of the applicant or applicant team will be assessed. In both of these examples, peer reviewers are required to draw predominately, and in most cases only, on their acquired experience and knowledge. It is an unfortunate irony that research evaluation can include so little empirical observation, ie. data collection, validation, and analysis.[15]

To address this, the implementation of the RQ+ Framework requires such peer assessment, but also demands that data be collected from the environment in which the research has occurred. Qualitative interviews with the researchers who conducted the project, research managers, and actual or potential research users (policymakers and practitioners)[16] added richness and resultant accuracy to the assessment at both project and program levels. Other notable methods used by the reviewers to "go beyond the output" included scans of bibliometrics and altmetrics and tailored surveys of "highly influential actors" in the particular field of research.

15 We note reviewer experience could also be categorized into RQ+ dimensions or rated as, e.g. specialist researchers, knowledge users, or researchers from complimentary disciplines. The IDRC external reviews described here did not take this step.

16 Research participants were not included in the External Reviews, but where time and resources allow this will be a valuable addition, in line with the spirit of the Framework.

Challenges and Pitfalls

The application of the RQ+ Assessment Framework at IDRC also brought some notable challenges to light.

The assessment is time-consuming, especially when robust triangulation is an imperative

Implementation of the RQ+ approach requires synthesis of qualitative and quantitative data from multiple sources and methods. Extensive consultation with internal and external stakeholders proved to be essential to filling gaps and allowing for sufficient triangulation. The quality and accessibility of sources of project and program data and information are therefore critical factors in the feasibility and value of the approach. Furthermore, where monitoring systems are set up without reference to the Framework components, data collection may require significant time and resources. Reviewers were of the view that the trade-off between comprehensiveness, ease-of-use, and reliable results versus the additional time needed was worthwhile. However, when applying the Framework, careful planning should go into determining a justifiable degree of effort and time.

Rigorous and credible sampling is critical

The IDRC External Reviews required selecting a sample of research projects from each program portfolio. The sampling strategy is crucial for the credibility of the whole effort, and can require a significant level of mindfulness to execute. It must therefore be carefully and thoughtfully conducted in order to ensure fair representation of the research program under review.

Sampling in the IDRC application was done using a purposeful approach, implemented in an iterative manner. In this case, criteria for research project selection were shared and discussed with research program managers before a final sample was drawn. While this might not be the best approach under all circumstances, it worked well for the context at hand. Random sampling may be a completely appropriate and preferable approach for project selection in another context.

Sampling from large and complex program portfolios can be technically and politically challenging, and, for best results, must be perceived by primary stakeholders and users as credible and legitimate. Considering from the start how to ensure a purposefully or randomly selected sample that is credible and acceptable to all primary stakeholders, or endorsed by the final authority, will streamline discussions and manage expectations. In the IDRC context, consultation with program staff was of great value. Of course, a view to mitigating potential biases should be kept in mind. If the purpose of the assessment demands a sampling process at arm's length from program staff, other stakeholders might be engaged in helping to define the sample.

Quantification after blending quantitative and qualitative data can appear to give simplistic results

At the micro level, the RQ+ approach asks reviewers to assess research projects using both qualitative and quantitative data. Rubrics were considered helpful to bring about more precision in judgment, including by blending the two types of evidence. However, this process became problematic when results were expressed in numerical values (e.g. the rubric ratings). In a sense, without reference to the precise wording of the rubrics, they were perceived as not appropriately capturing the rigor and depth and, hence, the true value and spirit of the assessment. Some reviewers tried to mitigate this perception by using color coding instead of quantitative ratings. The challenge was further compounded by subdimensions that were "not applicable" in certain programs.

At the macro level, data comparison and aggregation presented two challenges: i) understanding the relative values of scores between (sub) dimensions and deciding how these should be weighted and valued, and ii) working with the uncertainties created when following rubric aggregation to the program level, as the Framework guides the synthesis of project assessments into a program-level assessment based on numerical rubric based ratings. The value of a rubric in establishing a program-wide average or composite assessment for influencing factors or subdimensions at an overall program level can be – and was – seen by reviewers and program staff in both positive and negative terms.

On the positive side, the rubric ratings were seen to draw a well-defined and straightforward representation of research quality for the different subdimensions of interest. For an evaluation user interested in accountability and summative results, this might be a desirable and useful finding. On the negative side, the rubric was deemed to oversimplify the complexity of the assessment. For a research program

manager desiring details on *why or how* research quality could be improved, the underlying data behind the rubric rating might be more interesting.

It is worth noting that both reviewers and program managers undergoing review articulated these tensions regarding the rubric.

Using the RQ+ Assessment Framework for cross-program comparison can have undesirable, unintended consequences

Directly following from the previous point is the issue of using rubric aggregation as a mechanism for cross-program comparison. For example, if the ultimate user of the assessment is the senior management or board of an organization, such a comparison would be helpful in a strategic, organization-wide assessment of research performance. This can be especially useful if the larger performance discussion is underpinned by the blended qualitative/quantitative nuance that rubrics offer. However, there is a risk that comparison based on the simplistic interpretation of numerical scores might trigger competition among programs and, unwittingly, light the torch of a "program olympics". This could cause unwarranted and undesirable anxiety among program staff, and even a "race to the bottom" in performance measures.[17] Such anxiety runs counter to the IDRC's fundamental belief that monitoring and evaluation present useful and constructive methods of self- and external assessment that enable learning, accountability, improvement, and achievement of desirable results.

17 This is a risk, although such a dramatic scenario did not play out in the IDRC External Review process.

PART 4

The Potential of the RQ+ Assessment Framework

The Multiple Purposes of the Framework

The first application of the RQ+ Assessment Framework in the IDRC's External Reviews demonstrated that it can be successfully applied for summative evaluations in which research quality is a focus. However, this experience also highlighted alternative applications of the Framework across the research process.

It can facilitate match-making and coordinated planning between funding or research partners

By making explicit the priorities, values, and principles upon which judgments are made, the framework can be used to create a shared understanding among partners engaging in joint funding or assessment initiatives. This is exemplified in the Research Legitimacy dimension of the framework and its associated rubrics. IDRC funds research for development in the global South. The application of the Framework in this case therefore expressly recognized the importance of emphasizing local knowledge, inclusiveness and gender-responsiveness, and the need for attention to the potentially negative consequences of research. Partners may or may not have the same underlying values and priorities. Discussions on how the Framework will be used – and who will use it – can thus open opportunities for determining how differences can (or cannot) be accommodated. For example, the Research Integrity dimension incorporates traditional scientific merit considerations which typically occupy an important place in conceptions of research quality. During the first application of the Framework, some reviewers queried whether Research Integrity should be further elaborated or more heavily weighted.[18] They also discussed the prospect of incorporating a research efficiency or "value for money" dimension into the Framework.[19] These are all legitimate considerations that can and should form the basis of any conversation about "what matters" in research performance and evaluation.

As noted, one of the merits of the Framework is its flexibility. It can be adapted to different research contexts as well as management needs. Not every dimension of research quality needs to be assessed in every situation, and new dimensions can be developed to reflect particular organizational mandates. Yet if so desired, the Framework allows for standardization that can facilitate consistency and comparison across projects, programs, and organizations.

It can inform funding decisions and strengthen ongoing research projects

The RQ+ Assessment Framework can inform the development and assessment of research proposals submitted in response to competitive or targeted calls. Program officers and peer reviewers can also use the framework for consistent and systematic assessment of desired criteria at the beginning, or over the lifetime of a project or program. Research project managers can use it to ensure that research protocols under their purview include the dimensions that are important to them, and that the research team has considered factors that may influence potential success and has taken steps to ensure eventual knowledge translation.

Moreover, if the framework has been used concurrently to build a base of monitoring data on research quality, the strengths and weaknesses of a research program can be buffered as decisions are made about what to fund or how to structure proposed research protocols.

It can facilitate meta-evaluation and meta-analysis

The systematic and detailed approach to data collection and analysis demanded by the RQ+ approach may also facilitate synthesis and meta-evaluation. For example, evaluation results across programs can be rolled up to get an organizational perspective on research quality, or on specific dimensions of

18 In the event that weightings are applied to the different dimensions of the framework. For its use in External Reviews, IDRC chose not to assign weightings.

19 IDRC recognizes that having such a dimension could add real value to the framework, but was dissatisfied with the metrics currently in use and ultimately, did not consider them to be a good fit for its organisational mandate or business model.

research performance. Soon after publication of this paper, in early 2016, the IDRC will be in the planning phases of this type of meta-evaluation through review of its External Reviews. As another example, a research funder using the Framework to assess incoming applications might find value in periodic stock-taking (via aggregation of individual research application assessments) of the areas in which its community is showing strengths and weaknesses, especially if the organization has set strengthening of research capacity as priority. An academic journal using a tailored variation of the Framework to structure the peer review of its manuscripts could roll up yearly reviewer assessments to increase understanding of the strengths and weaknesses of the body of researchers or scientific discipline it serves to advance.

Finally, approaches to research evaluation that standardize the assessment criteria for an organization, program, journal or field – as the RQ+ Assessment Framework can – have the potential to facilitate meta-learning and knowledge synthesis through analysis of themes and crosscutting issues of interest. For example, an issue of interest - say gender - could be aggregated and compared across research disciplines, research programs, or even year over year.

It can facilitate understanding of how research impact is achieved

The comprehensive and systematic nature of the RQ+ approach has the potential to amass information which may prove useful in the study of research impact. Research is often funded with the intention of producing long-term impact for the public good. Research can also lead to many potential benefits – or costs – that may occur well after the research process has been concluded. Although the Framework focuses only on that which is within the primary stakeholders' sphere of control, the information gathered during assessments provides a comprehensive snapshot of the status of the project, program, or organization at a precise moment in time. Retrospectively combining such information with the tracing of project or program impacts (e.g. a policy change, lives saved or improved, a notable improvement in practice) may deepen understanding of the research process and contextual factors that generated these positive – or negative – real-world results.

Such insights have the potential to improve the stewarding of research and researchers towards long-term positive change. Questions in the case of IDRC could include, for example: Do the key influencing factors used in the Framework actually have influence (and what are the implications)? Do the subdimensions of research quality actually lead to, or enhance the chance of development outcomes? Does the IDRC approach to research for development actually spur research for development and real-world impact?

Building on Learning for Improving RQ+ use in the Future

Encouraged by the learning and results of applying the RQ+ Assessment Framework in its External Reviews, IDRC is considering its potential use – not only as a mechanism for evaluation, but as a framework for enhancing research management. As a research management tool, the Framework can support project and program delivery, and help ingrain the concepts reflected in the Framework more deeply into IDRC programming and the research that it funds.

References

Bowen, S. & Graham, I.D. 2015. "Backwards design or looking sideways? Knowledge translation in the real world: Comment on 'A call for a backward design to knowledge translation'." *In J Health Policy Manag.* (2015). 4(8):545-547. DOI:10.15171/ijhpm.2015.71.

Clark, W.C., Mitchell, R.B. & Cash, D.W. 2006. "Evaluating the Influence of Global Environmental Assessments." In Mitchell, R.B., Clark, W.C.., Cash, D.W & Dickson, N.M. ed., *Global Environmental Assessments: Information and Influence.* Cambridge, MA: MIT Press.

UK Department for International Development. 2014. *What is the Evidence on the Impact of Research for Development?* London, U.K.

Eyre-Walker, A. & Stoletzki. 2013. "The Assessment of Science: The Relative Merits of Post-Publication Review, the Impact Factor, and the Number of Citations". *PLOS Biology,* October 8, 2013, DOI: 10.1371/journal.pbio.1001675.

Feller, I. & Stern, P.C., ed. 2006. *A Strategy for Assessing Science: Behavioral and Social Research on Aging.* National Academy of Sciences, Washington, DC.

Greenhalgh T. & Wieringa, S. 2011. "Is it time to drop the 'knowledge translation' metaphor? A critical literature review." *J.r. Soc. Med:* 104:501-9. DOI: 10.1258/jrsm.2011.110285.

Guthrie, S., Wamae, W., Diepeveen, S., Wooding, S. & Grant, J. 2013. *Measuring Research: A Guide to Research Evaluation Frameworks and tools.* Prepared for the Association of American Medical Colleges. RAND Corporation, Santa Monica, CA, USA.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. & Rafols I. 2015. "The Leiden Manifesto for Research Metrics." *Nature,* April 23, 2015: 520: 429-431.

International Development Research Centre. 2014. *Towards Research Excellence for Development: The Research Quality Plus Assessment Instrument.* Ottawa, Canada. Available at idrc.ca/EN/Documents/Research-Quality-Plus-Assessment-Instrument.pdf.

Lavis, J., Lomas, J., Hamid, M. & Sewankambo, N.K. 2006. "Assessing country-level efforts to link research to action." *Bulletin of the World Health Organization:* 84:620-628.

McNie, E.C., Parris, A. & Sarewitz, D. 2015. "A Typology for Assessing the Role of Users in Scientific Research: Discussion Paper." Paper prepared by the Consortium for Science, Policy and Outcomes at Arizona State University. Available at cspo.org/wp-content/uploads/2015/02/Typology-Workshop-Paper.pdf.

Nowotny, H., Scott, P. & Gibbons, M. 2003. "Introduction. 'Mode 2' Revisited: The New Production of Knowledge." *Minerva* (2003) 41: 179-194.

Nutley, S., Walter, I. & Davies, H.T.O. 2007. *Using Evidence.* The Policy Press, Bristol, UK.

Ofir, Z. & Schwandt, T.S. 2012. *Understanding Research Excellence at IDRC: Final Report* (December 2012). Available at idrc.ca/EN/Documents/RE-study-Understanding-RE-at-IDRC-full-report.pdf.

Panel on Return on Investment in Health Research. 2009. *Making an Impact: A Preferred Framework and Indicators to Measure Returns on Investment in Health Research.* Canadian Academy of Health Sciences, Ottawa, Canada.

Spaapen, J. Dijstelbloem, H. & Wamelink, F. 2007. *Evaluating Research in Context* 2nd, ed. Netherlands: Consultative Committee of Sector Councils for Research and Development, p. 12. Available at qs.univie.ac.at/fileadmin/user_upload/qualitaetssicherung/PDF/Weitere_Aktivit%C3%A4ten/Eric.pdfp.

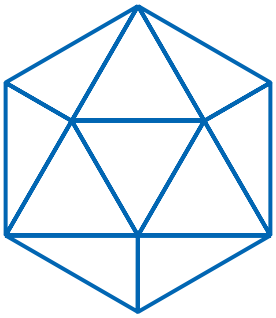
Sarewitz, D. 2011. "Institutional Ecology and the Social Outcomes of Scientific Research". *In* Fealing, K.H., Lane, J., Marburger, J.H. & Shipp, S., ed., *The Science of Science Policy,* 337-348. Stanford University Press, Stanford, CA, USA.

Stokes, D. 1997. *Pasteur's Quadrant: Basic Science and Technological Innovation.* Washington, DC: Brookings Institution Press.

Wilsdon, J. et al., 2015. The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. Higher Education Funding Council of England, Bristol, UK.

University of Gothenburg. 2013. "The Complex Role of Citations as a Measure of Scientific Quality," Science Daily, October 14, 2013.

World Bank and Elsevier. 2014. A Decade of Development in Sub-Saharan African Science, Technology, Engineering & Mathematics Research.



IDRC | CRDI
International Development Research Centre
Centre de recherches pour le développement international
Canada