

Making a difference in the real world? A meta-analysis of the quality of use-oriented research using the *Research Quality Plus* approach

Robert K. D. McLean ^{1,2*} and Kunal Sen ³

¹Policy and Evaluation Division, International Development Research Centre, Ottawa, Canada, ²Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa and ³Global Development Institute, University of Manchester, Manchester, UK

*Corresponding author. Email: rmclean@idrc.ca

Abstract

High-quality, use-oriented, and well-communicated research can improve social outcomes in low- and middle-income countries and, by doing so, accelerate development progress. We provide a meta-analysis of research supported by Canada's International Development Research Centre. We use a large and unique data set that comprises 170 research studies undertaken over the period 2010–2015. The research examined spans multiple disciplines of the social and natural sciences and was conducted across the globe, with the majority in Africa, Asia, Latin America, the Caribbean, and the Middle East. The evaluative framework we use—*Research Quality Plus*, *RQ+*—incorporates argumentation espoused in the *Leiden Manifesto*. As such, this article presents a case study of doing research evaluation differently and what the results can look like for research policy-makers. Our analysis suggests that contrary to conventional wisdom, there is no clear trade-off between the rigor and the utility of research and that research capacity-strengthening effort is positively correlated with the scientific merit of a project. We conclude that those located closest to a development challenge are generally those best positioned to innovate a solution. The results present novel evidence for those supporting, using, and doing research for development.

Key words: meta-evaluation; Southern research; capacity building; research quality; Research Quality Plus.

1. Introduction

High-quality research is an indispensable component of economic and social progress. In the Global South, this holds just as true. High-quality, use-oriented, and well-communicated applied natural and social science research can improve economic and social outcomes in Southern countries and, by doing so, accelerate development progress (DFID 2014). In the past several decades, there has been a significant increase in funding from bilateral and multilateral donor agencies to fund research about low- and middle-income countries. For example, the government aid agency of the UK will invest £390 million per year in research in 2017–2020 (DFID 2016). In the USA, the Global Development Lab of the United States Agency for International Development was created in 2014 to work specifically in science and innovation to tackle development challenges (USAID 2017). Philanthropists have become involved too. Take, for example, the *Grand Challenges* initiatives of the Bill and

Melinda Gates Foundation and their global propagation (BMGF 2017). At the same time, Southern granting councils are emerging and increasingly active in guiding the direction of scientific research in their own local contexts. For one example, 15 governments across Africa have made commitments to increase expenditure and coordination on science and research as a part of the Science Granting Council Initiative (CREST 2014; SGCI 2016).

Donors can have multiple objectives in funding research in Southern countries. These objectives include enhancing the quality of knowledge generation in the South, building capacity of Southern researchers and research institutions, and supporting research that generates evidence for policy and practice in Southern countries (Carden 2009). Yet, and despite the investment in research for development, there is limited knowledge of how effective the funding of research for development has been with respect to the multiple objectives that are expected of it.

Within the development sphere, but also well beyond it, researchers have extensively debated the best criteria for

determining the quality of natural, social, and behavioral science, and two general postulates have dominated this sphere:

1. First, that measuring the scientific merit of science is the domain of the scientist. Peer review has emerged and developed in line with this postulate, and over the past two decades, peer review has been increasingly supplemented by bibliometric measurement—a surrogate measure of the popularity of research among other researchers (Hicks et al. 2015).
2. Second, is that determining the scientific merit of research does not include assessment of the process and results of research that stretch beyond the realm of the researcher (e.g. capacity strengthening or impact). Broadly speaking, this is because these outcomes of research are seen to be a part of the social realm and beyond the direct system of science (Ofir et al. 2016).

Currently, this tradition of evaluating scientific quality is undergoing significant review and re-questioning. Concerns within the scientific community about the validity and reliability of bibliometric measurement are coupled with an increased desire from funders (public and private) for the demonstration of social impact of research investments (Hicks et al. 2015; Wilsdon et al. 2015; Holmes 2016). For example, the UK Government, in its review of the assessment of quality of research in UK higher education institutions, moved from a system that assessed only research outputs in the Research Assessment Exercise of 2008 to one that also incorporates the assessment of the impact of the research in the Research Excellence Framework of 2014 (Stern 2016). This debate is intertwined with the growth of a body of research that argues that the social value of science is not a matter of research publication and dissemination but, instead, a complex and iterative process of social interactions with research users, beneficiaries, and other intended and unintended stakeholders (Greenhalgh and Wieringa 2011; Bowen and Graham 2015; Nutley, Walter and Davies 2007; D'Este et al. 2018).

As a result, there exists a global and cross-disciplinary re-questioning of whether the methods we use for research evaluation are those best suited for uncovering, measuring, comparing, and, by extension, achieving the potential value of scientific research. But, there is limited evidence of the usefulness of alternate methods of research evaluation.

In this article, we provide a meta-analysis of the quality of research supported by Canada's International Development Research Centre (IDRC), an organization with 48 years' experience funding research for the development priorities of Southern countries. This is a particular subset of the global research enterprise. For IDRC, research for development, or R4D, comprises research activities that aim to find solutions to growth, equity, justice, and efficiency challenges faced in Southern countries. The majority of this research is undertaken in Southern countries by Southern researchers, and it has spanned scientific disciplines from economics to neuroscience, and accepts multi- and transdisciplinary approaches common in fields such as agriculture or climate change. A detailed account of the historical experience of IDRC is available in the article by Muirhead and Harpelle (2010).

We conduct the meta-analysis using the Research Quality Plus (RQ+) approach. We present this analysis as a validation of the effectiveness of the RQ+ approach to research quality evaluation. RQ+ is a novel evaluation methodology that builds on the analytic assessment provided by bibliometrics/altmetrics and the deliberative results of peer review. Furthermore, it incorporates the majority of the theory-driven arguments espoused in the *Leiden Manifesto*

(Hicks et al. 2015) into a practical evaluative tool. For example, the RQ+ approach facilitates independent, expert review that is values-driven, inspired by systems thinking, accepting of quantitative and qualitative evidence, and systematic. At the same time, RQ+ moves beyond traditional measures of scientific research rigor, to capture the multiple objectives that underpin the greater potential of research for society, such as research uptake and use, capacity strengthening of researchers and/or research institutions, and the legitimacy of the research to local knowledge and demand.

In the following section of this manuscript, we provide an overview of the RQ+ approach and the RQ+ assessment framework as it was applied at IDRC—our data set's underpinning evaluative framework and eligibility criteria for independent study inclusion in the meta-analysis. In the third section, we provide a description of our methods to conduct the meta-analysis. In the fourth section of the article, we present the findings of our meta-analysis. In the final section, we offer some interpretation of the results and discuss their meaning. We argue that this exercise has offered a quantitatively powerful and a qualitatively rich evidence base to inform decision-making for a diverse range of actors in the research for development system. We are unaware of any data set of research for development quality with a similar explanatory value.

2. The RQ+ approach

The RQ+ approach emerged from a body of work undertaken at IDRC since 2012.¹ At the highest level, the RQ+ approach can be described as a stance for evaluating research quality that comprises three fundamental notions. These are introduced in detail below, but in brief are (1) accepting a multidimensional view of quality, (2) gathering contextual understanding, and (3) demanding judgment based on empirical evidence. The RQ+ approach was put into action at IDRC with a bespoke RQ+ assessment framework. A comprehensive description of the RQ+ assessment framework used at IDRC, the rationale for creating the RQ+ approach, as well as reflection on the first implementation of the approach is presented in the article by Ofir et al. (2016). Here we present a summary overview of the approach and the assessment framework, to position our meta-analysis. To our knowledge, the RQ+ approach has been used primarily for the assessment of research for development. We see no reason it would not apply, given appropriate tailoring, outside of this context.

2.1 Rationale and purpose for RQ+

At the heart of the operational model of Canada's IDRC is the financing of *research for development*. Simply put, this implies that IDRC-supported research aims for both scientific and societal impact, it is solutions-oriented and it occurs within a diversity of contexts. It is research that is intended to contribute to social and economic development progress in Southern countries. Although the synergies, challenges, and tensions of producing socially relevant and scientifically meritorious research are well described and debated in the academic literature, fewer practical contributions to how this research can be evaluated have been presented, and fewer still have been validated with systematic testing (Greenhalgh et al. 2016; Mendez 2012; Bornmann 2013; D'Este et al. 2018). Accordingly, the RQ+ approach was motivated by IDRC's desire to advance global research evaluation practice and, more

pragmatically, by the need to bring rigor to the assessment of the research it itself supports.

To ground this motivation in the state of the art of research evaluation and the perspectives of IDRC's Southern research community (a group of researchers who are severely underrepresented in research quality and evaluation debates), two foundational studies were conducted. Mendez (2012) undertook a broad literature review of research evaluation frameworks, and Singh et al. (2013) sought to identify and document Southern perspectives of research quality.²

Mendez (2012) focused on what constitutes research excellence and on mechanisms to evaluate it. The literature reveals that there is no single definition, standard, or method for research excellence evaluation. Rather, there are many definitions for both research and excellence, there is no agreement on the quality dimensions that should be used to evaluate research, and there are large debates around the mechanisms used to evaluate research excellence (e.g. peer review and bibliometric analysis). This article does not answer questions about which definition or approach is better; instead, it presents the range of arguments and ideas found in the literature.

Singh et al. (2013) undertook an empirical enquiry into how Southern researchers view research excellence and how their experiences can inform the creation of a framework for the assessment of research excellence at IDRC. The study collected primary data through surveys and interviews, and although it did not draw a specific definition of research quality, it presented a novel and useful data set for RQ+ ideation.

As this body of work evolved, so too came a number of high-level calls for reform in the global research evaluation sphere, likely the most impactful of these was the Leiden Manifesto (Hicks et al. 2015). By citing malpractice in the use of metrics for research evaluation and forwarding 10 principles for improvement, the Leiden Manifesto aimed to contribute to advancing science and how it might interact more fluidly with society. This created a powerful backdrop for, and input to, the development of RQ+. As a result, RQ+ is positioned to address the systemic weaknesses in the research evaluation outlined in the Leiden Manifesto and presents one way for moving the principles of the Manifesto into practice.

In sum, IDRC's development of RQ+ stemmed from a number of influences. First, a practical desire to do better at evaluations of research quality at IDRC. Second, a body of research and reflection undertaken by IDRC and its research community from 2012 to 2015. Finally, the backdrop of a global movement calling for reform and improvement across the research evaluation enterprise.

2.2 The RQ+ approach holds three tenets

1. *Accept a multidimensional view of quality that is based on the values and objectives that drive a research agenda:* For IDRC, scientific rigor is a non-negotiable, but being interested in research for development, a complete picture of quality moves beyond this traditional measure of rigor to encapsulate research legitimacy, importance, and how the research is positioned for use. To another funder, government, think tank, journal, university, and so on, these quality dimensions may be very different. This is a good thing. As the Leiden Manifesto states, 'the best judgments about the quality of research should be taken by combining robust statistics with sensitivity to the aim and nature of the research that is evaluated' (Hicks et al. 2015).

2. *Research happens in a context, embrace and learn from this:* The predominant forms of research quality assessment aim to isolate research from its environment (e.g., blinded peer review). The RQ+ approach argues that this reductionist method of quality appraisal limits what we come to know about knowledge production processes and results. For instance, considering research not as isolated from but as a product of varying political, organizational, disciplinary, and/or data environments supports a systems-oriented assessment of quality. As the Leiden Manifesto states, '... (research evaluations) should take into account wider socio-economic and cultural contexts. Scientists have diverse research missions.' (Hicks et al. 2015).
3. *As with the research we conduct, judgments should be underpinned by empirical evidence. Not just opinion:* For example, go out and ask the intended users of a research project for their insights and balance these against the voice of the beneficiary community, expert researchers in the same field, and the bibliometrics. It is an unfortunate paradox of the sciences that the most utilized approach to research evaluation rests entirely on opinion. As the Leiden Manifesto states, 'decision-making about science must be based on high-quality processes that are informed by the highest quality data' (Hicks et al. 2015).

2.3 The RQ+ assessment framework

The practical manifestation of RQ+ at IDRC is found in the RQ+ assessment framework (IDRC 2017). The framework presents a tool for evaluating research quality in a systematic and transparent way. A postulate of the RQ+ approach is that research evaluation should be tailored to context, and so, it should be cautioned that what is presented hereafter is the framework as it is currently envisioned for IDRC, and how it was constructed and applied in the 2015 evaluations analyzed in this manuscript. Those interested in using the framework should begin with a comprehensive review of its components vis-à-vis their own research objectives, values, and environment.

The RQ+ assessment framework consists of three components: (1) research quality dimensions and subdimensions, (2) contextual factors, and (3) evaluative rubrics. These components are presented in turn hereafter.

2.3.1 Research quality dimensions and subdimensions

Ofir et al. (2016) describe a benefit of applying an evaluation framework that captured the essence of IDRC values as an increased confidence of the evaluators in the eventual utility of the results. In evaluator jargon, 'what mattered was measured'.

Technically, these values were categorized as research quality dimensions and subdimensions. The four principal quality dimensions in RQ+ as applied in this exercise were (1) research integrity, (2) research legitimacy, (3) research importance, and (4) positioning for use.

Research integrity considered 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. Reviewers placed specific emphasis on the research design, methodological rigor, literature review, and the relationship between evidence gathered and conclusions reached and/or claims made, in their scoring.

Research legitimacy considered 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 (Cash et al. 2003). This dimension captured legitimacy in terms of 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. The research legitimacy dimension had four subdimensions: (1) addressing negative consequences, that is, the potentially negative consequences and outcomes for populations; (2) gender responsiveness, that is, how responsive to gender concerns is the project; (3) inclusiveness, that is, whether the project is inclusive of vulnerable populations; and (4) engagement with local knowledge, that is, whether local context and engagement has been a focus of the project.

Research importance considered 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. It had two subdimensions: (1) originality of the research and (2) the relevance of the research.

Positioning for use considered 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 was enhanced. The incorporation of this dimension in the RQ+ framework was guided by the understanding that the uptake of research is inherently a political process and that 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, and, therefore, requires strategies to integrate potential users into the research process itself wherever this is feasible and desirable.

Figure 1 presents a visual representation of the multidimensional nature of research quality expressed in the RQ+ approach (it includes the dimension of research integrity and all subdimensions).

2.3.2 Contextual factors

Contextual factors—either within the research endeavor or in the external environment—are issues that hold the potential to affect (positively or negatively) the quality of the research. The RQ+ framework identifies five main contextual factors.

The first is *the maturity of the research field*, which is 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.

The second factor is *research capacity strengthening*, which is 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.

The third factor is *risk in the research environment*, which is 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.



Figure 1. Research quality as multidimensional.

Source: McLean and Feinstein (2016).

The fourth factor is *risk in the political environment*, which is 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.

The final factor is *risk in the data environment*, which is 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.

Figure 2 presents an illustration of research quality as a context-bounded and dynamic concept.

2.3.3 Evaluative rubrics

The final component of RQ+, *the evaluative rubrics*, sets judgement criteria for reviewers, clarifying how performance should be measured for each dimension and subdimension of research quality and each contextual factor. The rubrics were a feature that facilitated the blending of qualitative and quantitative evidence into a single evaluative assessment (Ofir et al. 2016). The standardized rubrics facilitated the systematic approach to evaluation judgement that allowed for the meta-analysis that follows in this manuscript.

In terms of research quality dimensions and subdimensions, the rubrics used graduated levels of achievement. Each subdimension for research legitimacy, research importance, and positioning for use and the principal dimension of research integrity was scored from 1 to 8, with scores of 1 or 2 indicating unacceptable levels of achievement, scores of 3 and 4 less than acceptable, scores of 5 and 6 acceptable to good, and scores of 7 and 8 very good achievement. Once scores were arrived for the subdimensions of research legitimacy, research importance, and positioning for use, they were aggregated to arrive at an overall score for the relevant dimension.

For contextual factors, reviewers made ratings using a three-point rubric. For the three contextual factors related to risk in the



Figure 2. The dynamism of research quality.
Source: Ofir et al. (2016).

political, research, and data environment, and for the contextual factor related to maturity of the research field, projects scored 1 when exhibiting low risk or high maturity in the field, 2 for medium risk and maturity, and 3 for high risk or low maturity accordingly. Projects where research capacity strengthening was of low focus scored 1, projects scored 2 when of medium focus, and projects scored 3 when a high focus was placed on capacity strengthening.

In Figure 3, we provide an example of the rubric for the RQ+ subdimension: engagement with local knowledge.

Figure 4 outlines a complete picture of the RQ+ assessment framework.

3. Methods

The methods section of the article is presented in two parts. First, we outline the process we undertook to select studies and aggregate data to conduct the meta-analysis. Second, we present our overarching approach to statistical analysis.

3.1 Meta-analysis and sample overview

Meta-analysis is a technique that collates the results of multiple scientific studies into a single record; statistical methods are then

applied to the analysis of the amalgamated data set, doing so to increase the point precision and generalizability of results (Liu 2015; Gurevitch et al. 2018).

In 2015, seven external evaluations of IDRC supported research—which had embedded the RQ+ approach—were completed. The RQ+ data from these seven evaluations comprise the metadata we analyze and present in this article. The systematic use of the RQ+ approach allowed valid quantitative aggregation.

Each assessment of quality made in each of these seven evaluations was derived by a team of three independent subject matter experts and reported publically in formal evaluation reports (these are available in IDRC Digital Library 2017). To arrive at the scores for the RQ+ rubric, for each project, the experts conducted desk-based reviews of project documentation (including research outputs and publications) and conducted interviews of the project staff responsible for administering the projects, researchers involved in the project, and key research users (such as policymakers in Southern countries and senior staff in bilateral and multilateral development agencies). The RQ+ approach aimed to increase validity and accuracy by requiring reviewers to go beyond an assessment of the project output (e.g. publication) to collect and triangulate data from various primary and secondary sources. To facilitate a neutral

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.

Figure 3. Example of the evaluative rubric for engagement with local knowledge.

Source: Ofir et al. (2016).

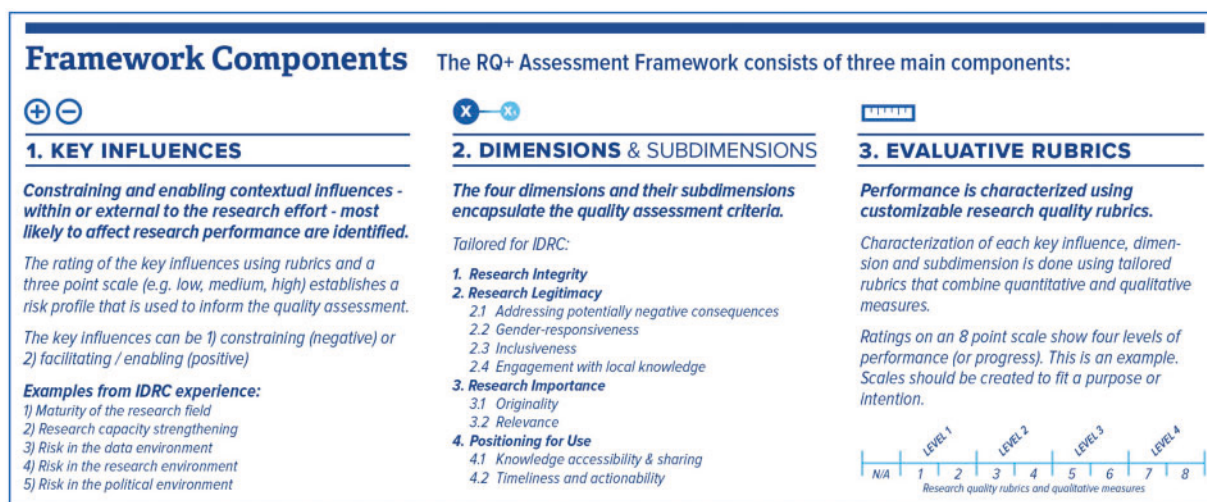


Figure 4. The components of the IDRC-tailored RQ+ assessment framework.

Source: Ofir et al. (2016).

and independent review, the external review team selected and implemented the approach to collecting and synthesizing these data on their own terms. Processes used across the seven evaluations were not entirely similar. In some cases, surveys of research-user groups were used, and in others, in-depth interviews with beneficiaries.³

The aggregate metadata includes 170 components from 130 discretely funded research projects funded by IDRC between 2010 and 2015. The areas of the research ranged from climate change, water, and health, to governance, justice, and economics. The research happened around the world; the majority in Africa, Asia, the Caribbean, Latin America, and the Middle East. The types of institutions that were involved in the research were universities, research institutes, government agencies, and nongovernmental organizations.

Using IDRC records, we cross-tabulated four demographic variables (project financial size, region, multiple funders or not, and institution type), project by project, into this data set.

We are unaware of any data set tracking research for development that matches this magnitude, depth, and breadth.

3.2 Statistical analysis

We first analyzed the data using summary statistics—mean, standard deviation, and minimum and maximum values of each RQ+ dimension/subdimension score for the 170 components.⁴ We next conducted one-way analysis of variance (ANOVA) tests for different categorizations of the grants—by region, by recipient institution, and by broad region—to assess whether there are significant differences in the means of RQ+ dimensions across the various categorizations.⁵ We conducted omnibus *F* tests where the null hypothesis of no difference between the means of the population subsamples was tested across each data categorization. If the null hypothesis is rejected, then we can infer that at least one of the population subsamples is different from the other means. However, the *F* test cannot tell us which mean is different from the others. To find out which means are different, we used a multicomparison method—the Tukey *t*-test—that allows us to test which mean of a specific RQ+ dimensions for a particular population subsample is different from the means of the same RQ+ dimension for the other population subsamples. The test compares the difference between

Table 1. Results of RQ+ analysis for the entire sample

RQ+ components	Number of observations	Mean	Standard deviation	Minimum	Maximum
RQ+ contextual factors					
Maturity of the research field	170	1.78	0.68	1	3
Research capacity strengthening	166	2.14	0.81	1	3
Risk in the data environment	170	1.78	0.72	1	3
Risk in the research environment	169	1.70	0.70	1	3
Risk in the political environment	169	1.71	0.77	1	3
RQ+ dimensions					
1. Research integrity	169	5.81	1.70	1	8
2. Research legitimacy	63	5.67	1.58	1	8
2.1 Addressing negative consequences	76	5.37	1.92	1	8
2.2 Gender responsiveness	125	4.81	2.17	1	8
2.3 Inclusiveness	124	5.59	2.06	1	8
2.4 Engagement with local knowledge	148	6.29	1.55	1	8
3. Research importance	165	6.35	1.32	1	8
3.1 Originality	165	5.98	1.60	1	8
3.2 Relevance	165	6.71	1.35	1	8
4. Positioning for use	157	5.77	1.49	1	8
4.1 Knowledge accessibility and sharing	160	5.94	1.57	1	8
4.2 Timeliness and actionability	165	5.65	1.71	1	8

each pair of means with appropriate adjustment for the multiple testing.

Finally, we calculated correlation coefficients across and between contextual factors in the RQ+ framework and RQ+ dimension/subdimension scores to assess the relationship within and between contextual factors and research quality. We used nonparametric Spearman correlations due to the ordinal nature of the data. The level of significance was set at 5%. The analysis was undertaken using STATA version 14.0.

4. Results

We begin with an examination of the key influences on the research in the 170 cases. We find that there was a strong focus on research capacity strengthening, with the highest score among the five key influences (a mean of 2.14; [Table 1](#)). For the other key influences, most projects were in established or emerging fields, or low to medium risk.

Turning to the RQ+ dimensions, the highest level of achievement was observed for research importance, with an average of 6.71, suggesting the average project in the sample was judged very good in this dimension. In contrast, the average scores for research integrity, research legitimacy, and positioning for use were 5.81, 5.67, and 5.77, respectively. Within the research legitimacy dimension, gender responsiveness has the lowest level of achievement, with a mean of 4.81, and engagement with local knowledge has the highest level of achievement, with a mean of 6.29.⁶ Within the research importance dimension, relevance has a significantly higher score of 6.71, as compared with the originality subdimension 5.98. Within the positioning for use dimension, there is little difference in the level of achievements of the two subdimensions—knowledge accessibility and sharing and timeliness and actionability with scores of 5.94 and 5.65, respectively.

When we disaggregate the RQ+ dimensions by regions, we find that the highest levels of achievement are in Latin America, whereas the lowest levels of achievement are in sub-Saharan Africa for

research legitimacy and research importance and in Asia for research integrity and positioning for use ([Figure 5](#)).

Disaggregating the RQ+ dimensions by recipient institution type, we find the average score for research integrity highest for research institutions; for research legitimacy, it was highest for Non-Government Organizations (NGOs)/International Non-Government Organizations (INGOs); and for research importance, it was the highest for research institutions ([Figure 6](#)). For the positioning for use dimension, we find the highest score for the combination of multiple types of organization working together.

Categorizing the grants by the broad region where the researchers are located (South, North, and both regions), we find that Southern projects have the highest scores in all RQ+ main dimensions ([Figure 7](#)).

We next present results of the ANOVA tests. We begin with conducting ANOVA tests on the means of RQ+ dimensions by region. We find that the null of no difference in means across regions for research integrity and research importance can be rejected, but not for research legitimacy and positioning for use ([Table 2](#)). However, when we do pairwise comparison of means, we find that the *t*-ratio on difference in means for Asia as compared with Latin America is significant in the case of research integrity (with the mean for Asia lower than the mean for Latin America), and the *t*-ratio for the difference in means for sub-Saharan Africa as compared with Latin America is significant for research importance (again, with the mean for sub-Saharan Africa lower). No other *t*-ratios on difference in regional means by RQ+ dimension are significant at conventional levels of significance.

Conducting ANOVA tests on the means of RQ+ dimensions by recipient institution type, we find that the null of no difference in means across regions for research integrity can be rejected, but not for research legitimacy, research importance, and positioning for use ([Table 3](#)). The only *t*-ratios for difference in means by RQ dimension that are significant are for NGOs versus research institutions (the mean for NGOs is lower) and for multiple recipients versus research institutions (the mean for multiple recipients is lower).

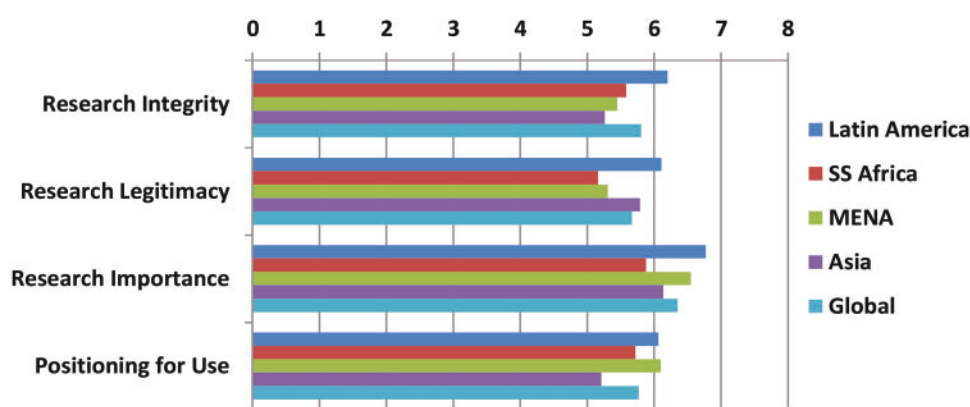


Figure 5. RQ+ quality dimensions by region of research focus.

Note: Total sample = 170. Within this: Latin America and the Caribbean = 54, Sub-Saharan Africa = 36, Middle East and North Africa = 11, Asia = 39, Global = 30.

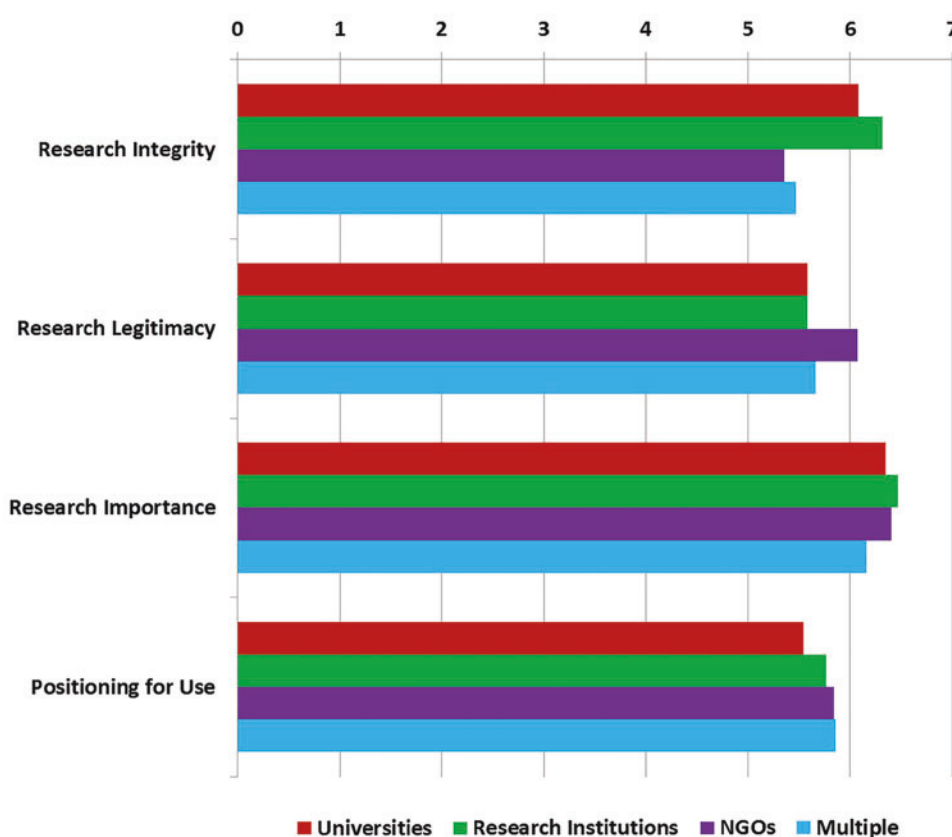


Figure 6. RQ+ quality dimensions by recipient institution type.

Notes: 1) Total sample = 170. Within this: universities = 33, research institutions = 50, NGOs = 44, Multiple = 43. 2) 'NGOs' includes INGOs. 3) 'Multiple' includes any combination of two or more recipient types working together.

Conducting ANOVAs on the means of RQ+ dimensions by broad regions, we find that the null of no difference in means in RQ+ dimensions cannot be rejected, indicating that there is no statistically significant difference between the means of RQ+ dimensions by broad region (Table 4).

We then examine the correlations between the contextual factors and RQ+ quality dimensions to see if contextual factors within the research endeavor or in the external environment have any influence on research quality. We find strong correlation between

research capacity strengthening and research importance (a correlation coefficient of 0.40 and significant at 5% level) and between research capacity strengthening and research legitimacy (correlation coefficient of 0.34 and significant at the 5% level; Table 5). There is a negative correlation between risk in the research environment, on the one hand, and research integrity, research importance, and positioning for use, on the other hand. There is weaker correlation between other key influences and the main RQ+ dimensions.

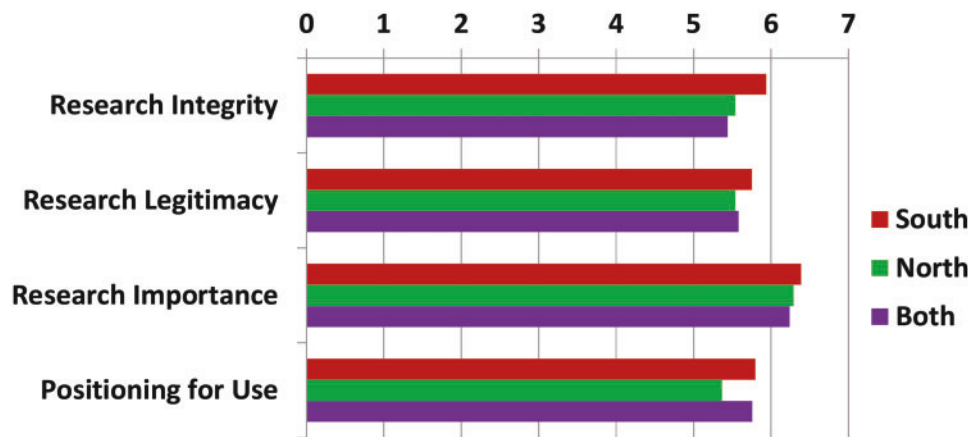


Figure 7. RQ+ quality dimensions by broad region of research.
 Note: Total sample = 170. Within this: North = 26, Both = 25, South = 119.

Table 2. Are the means of main research dimensions across regions the same?

Regional comparisons	Research integrity	Research legitimacy	Research importance	Positioning for use
Sub-Saharan Africa vs. Latin America	-1.71	-1.75	-3.22**	-1.05
Middle East and North Africa vs. Latin America	-1.34	-0.91	-0.50	0.08
Asia vs. Latin America	-2.67*	-0.63	-2.30	-2.63
Global vs. Latin America	0.10	-0.78	-1.28	0.63
Middle East and North Africa vs. sub-Saharan Africa	-0.22	0.18	1.47	0.72
Asia vs. sub-Saharan Africa	-0.85	1.19	0.86	-1.44
Global vs. sub-Saharan Africa	1.58	0.41	1.63	0.31
Asia vs. Middle East and North Africa	-0.35	0.55	-0.90	-1.69
Global vs. Middle East and North Africa	1.33	0.16	-0.33	-0.48
Global vs. Asia	2.41	-0.38	0.82	1.67
F-test on whether means by regions are the same	3.67**	0.86	2.93**	1.84

Notes: ** and * indicate whether *t*-statistic/*F*-statistic is significant at 5%, or 10% level of significance. In each cell, the means of RQ main dimensions by regions are compared, and *t*-statistics of pairwise comparisons of means are reported in each row, except last row, where *F*-statistic on whether means are different across region is reported. Positive values of *t*-statistics indicate that mean of first group compared is higher than the second group; negative values indicate the opposite. Tukey's method is used to calculate *t*-statistics.

Table 3. Are the means of main research dimensions across recipient institutions the same?

Institutional comparisons	Research integrity	Research legitimacy	Research importance	Positioning for use
Research institution vs. university	0.62	-0.80	0.41	0.68
NGO vs. university	-1.92	0.90	0.15	0.84
Multiple vs. university	-1.61	0.14	-0.62	0.90
NGO vs. research institution	-2.80**	1.60	-0.27	0.18
Multiple vs. research institution	-2.46*	0.94	-1.13	0.24
Multiple vs. NGO	0.33	-0.80	-0.82	0.06
F-test on whether means by recipient institutions are the same	3.57**	0.88	0.45	0.32

Notes: ** indicates whether *t*-statistic/*F*-statistic is significant at 5%, level of significance. In each cell, the means of RQ main dimensions by recipient institution are compared, and *t*-statistics of pairwise comparisons of means are reported in each row, except last row, where *F*-statistic on whether means are different across recipient institution is reported. Positive values of *t*-statistics indicate that mean of first group compared is higher than the second group; negative values indicate the opposite. Tukey's method is used to calculate *t*-statistics.

Between RQ+ main dimensions, we find strong associations between these measures, with correlation coefficients in the range of 0.4–0.7, and statistically significant. This suggests that projects that score highly in one main dimension also score highly in other dimensions (Table 5).

With respect to the correlation between contextual factors and RQ+ subdimension measures (Table 6), we find limited evidence of strong associations, with the exception of a strong correlation between research capacity strengthening and originality (correlation coefficient of 0.45 and statistically significant).

Table 4. Are the means of main research dimensions across broad regions the same?

Broad regional comparisons	Research integrity	Research legitimacy	Research importance	Positioning for use
North vs. South	-1.61	-1.02	-0.88	-0.57
Both vs. South	-0.27	-0.65	-0.86	0.01
Both vs. North	1.04	0.25	0.00	0.44
F-statistic on whether means by broad regions are the same	1.30	0.10	0.28	0.10

Notes: In each cell, the means of RQ main dimensions by broad regions are compared, and *t*-statistics of pairwise comparisons of means are reported in each row, except last row, where *F*-statistic on whether means are different across broad region is reported. Positive values of *t*-statistics indicate that mean of first group compared is higher than the second group; negative values indicate the opposite. Tukey's method is used to calculate *t*-statistics. Where *N* = 170 and is composed of: South = 119, North = 26, both = 25.

Table 5. Correlations between key influences and RQ+ main dimensions

	Mat	Cap	RiskD	RiskR	RiskP	Resint	Resleg	Resimp	Posuse
RQ+ contextual factors									
Mat	1.00								
Cap	0.03	1.00							
RiskD	-0.08	-0.04	1.00						
RiskR	-0.05	-0.20*	0.52*	1.00					
RiskP	0.10	-0.06	0.18*	0.35*	1.00				
RQ+ dimensions									
Resint	0.02	0.25*	-0.14	-0.25*	0.01	1.00			
Resleg	-0.09	0.34*	-0.05	-0.05	0.03	0.43*	1.00		
Resimp	0.15	0.40*	-0.14	-0.20*	0.17*	0.59*	0.69*	1.00	
Posuse	0.12	0.27*	-0.04	-0.29*	-0.03	0.50*	0.48*	0.63*	1.00

Notes: Correlation coefficients in cells. Mat = maturity of research field; Cap = research capacity strengthening; RiskD = risk in the data environment; RiskR = risk in the research environment; RiskP = risk in the political environment; Resint = research integrity; Resleg = research legitimacy; Resimp = research importance; Posuse = positioning for use. * indicates significance at 5% level or less.

5. Discussion

This study provided a meta-evaluation of the quality of research supported by Canada's IDRC. The analysis was based on a large and unique data set that comprises 170 independent, third-party expert reviews of research projects supported over the period 2010–2015, spanning scientific disciplines and regions of the globe. In the previous section we provided our analysis technique and results. Based on these results, we draw the following inferences about research for development:

1. Our results show that scientifically excellent research is useful research. Conventional wisdom suggests a trade-off between the rigor and the utility of research. In other words, the policy-making can move too quickly to wait for the best designed and executed scientific studies. In our analysis, a strong positive correlation between research integrity and positioning for use suggests the opposite. We suggest that this provides evidence for great attention to scientific integrity for those investing in research to achieve development outcomes.
2. We find that in research for development, risk and opportunity are diversified. The incidence of internal and external environmental contextual factors is mixed across regions and disciplines, and there is little evidence of correlation between these factors. Traditional assumptions about the generalized risk of undertaking

research in the South are disputed with these data. Instead, the environment is similar to the science and research environment of the global North, where risk and opportunity are considered on a case-by-case basis. We suggest that this implies idiosyncratic funding program design and funding decisions, attention to contextual detail in monitoring and evaluation of research projects, and the avoidance of sweeping risk assessment claims regarding research for development led in the South.

3. At the same time, we find that research context indicates some broad trends in terms of correlation with research quality. In other words, knowing more about the environment in which research takes place can help one to understand the quality it achieves. For instance, risk in the research environment is overall negatively associated with research quality, and so too is risk in the data environment. Whereas, risk stemming from an immature field and/or capacity strengthening is in fact generally positively correlated to quality, and quite strongly in the case of capacity-strengthening efforts. Political environments have little correlation to quality, except when it comes to the importance of research where positive (though weak) association with quality is evident. We suggest this furthers the case for thoughtful review of research environments, to fully understand quality determinants and draw reasonable conclusions on the quality of any research process.
4. Our results indicate that capacity-strengthening efforts are positively correlated with the quality of research projects, including with scientific integrity. This contradicts a potent assumption—that research requiring attention to training and support to skills development will also be poor-quality research. We suggest an implication being that research which requires or includes a focus on capacity strengthening should not be avoided due to a desire for excellence in traditional views of scientific rigor.
5. We find several compelling correlation coefficients relate to capacity strengthening and research originality (a subdimension of research importance). Max Planck famously noted that, 'A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.' Our finding seems to provide support to the hypothesis that innovative, original research is undertaken by those who are new to a field.⁷ A strong positive correlation between the effort spent on capacity building and originality of research supports this. Further, we find that research capacity-strengthening effort is positively correlated with the scientific merit of a project. But, our analysis demonstrates a particularity about Planck's assertion he may have overlooked. The only factor more strongly correlated to originality of science than the fact it

Table 6. Correlation matrix between key influences and RQ+ subdimensions

	Mat	Cap	RiskD	RiskR	RiskP	ResInt	Addneg	Genres	Inc	Lockn	Orig	Rel	Know	Timel
RQ+ contextual factors														
Mat	1.00													
Cap	0.08	1.00												
RiskD	-0.04	0.05	1.00											
RiskR	-0.05	-0.20*	0.52*	1.00										
RiskP	0.10	-0.06	0.19*	0.35*	1.00									
RQ+ subdimensions														
Resint	0.02	0.25*	-0.14	-0.25*	0.01	1.00								
Addneg	0.05	0.36*	-0.11	-0.13	0.07	0.39*	1.00							
Genres	-0.14	0.03	-0.06	-0.01	0.12	0.22*	0.41*	1.00						
Incl	-0.21*	0.10	-0.10	-0.03	0.11	0.36*	0.44*	0.71*	1.00					
Lockn	0.01	0.28*	-0.19*	-0.27*	-0.07	0.51*	0.42*	0.39*	0.57*	1.00				
Orig	0.18	0.45*	-0.13	-0.16*	0.13	0.56*	0.45*	0.31*	0.36*	0.54*	1.00			
Rel	0.08	0.25*	-0.12	-0.20*	0.18*	0.48*	0.55*	0.40*	0.39*	0.47*	0.60*	1.00		
Know	0.02	0.22*	-0.01	-0.21*	0.08	0.36*	0.35*	0.22*	0.32*	0.38*	0.40*	0.53*	1.00	
Timel	0.21	0.21	-0.13	-0.29	-0.08	0.46	0.43	0.21*	0.32*	0.51*	0.52*	0.59*	0.67*	1.00

Notes: Mat = maturity of research field; Cap = research capacity strengthening; RiskD = risk in the data environment; RiskR = risk in the research environment; RiskP = risk in the political environment; Resint = research integrity; Addneg = addressing negative consequences; Genres = gender-responsiveness; Inc = inclusiveness; Lockn = engagement with local knowledge; Orig = originality; Rel = Relevance; Know = knowledge accessibility and sharing; Timel = timeliness and actionability. * indicates level of significance at 5% level or less.

is being undertaken by new researchers is the degree to which the research is incorporating local knowledge (a subdimension of research legitimacy). In other words, those most closely linked to a problem appear best placed to innovate a solution to it.⁸

6. We find that Southern research demonstrates high quality, in all RQ+ dimensions. In fact, Southern research demonstrates superior research quality to Northern research⁹ and to partnered North–South research. This is not to say that the research happening in the South is categorically better than the North. It is important to recall the data set examined in this study comprised research projects with objectives to improve social outcomes in the global South. As such, this analysis demonstrates the validity of Southern-led research for development. When a problem is local, locals appear best placed to address it. Further to this, we suggest that North–South research partnerships may hold great value for interdisciplinary expansion, internationalization of science, and shared problem-solving. However, we should not assume that Northern partners are improving the capacity of Southern ones or improving the quality of the science undertaken. Rather, North–South partnerships should be predicated on mutually strategic benefits.

5.1 Limitations

A comprehensive discussion of limitations of the RQ+ approach to evaluating research, and the limitations of the seven RQ+ external evaluations undertaken in 2015 that have been aggregated for this meta-analysis, is provided in Ofir et al. (2016).

Here we note limitations of our meta-analysis.

1. First, we note that a bias was existent, and entirely intentional, in the construction of the dimensions and contextual factors examined. We have measured and thus highlighted elements that are particularly important to IDRC. We forgo the analysis of other dimensions of quality in doing so. For an example, we stake no claim about researcher or research project ‘productivity’, which is a common measure of research project success and

is widely defined as the number of research outputs per unit. We measure what has mattered primarily to IDRC.

2. We hold concern that the comprehensive nature of the RQ+ approach has yielded meta-analysis that is, on the one hand, unique and path-breaking, but, on the other hand, setting a high bar. We admit concern that the examination of these comprehensive sets of variables may lead to the development of another set of challenges for researchers and research organizations wishing to assess the quality of their work.
3. We suggest the metadata examined could be diversified and the learning potential amplified by the inclusion of projects supported by alternative funders. For the reason identified in the first limitation, or for others that are yet to be uncovered, there may be implicit bias in the data that we cannot identify without source comparison. To mitigate this limitation, we openly call on other funders to replicate or reproduce the study approach.
4. Finally, we note the limitations of this meta-study emerging from our strictly quantitative approach. In future iterations, the synthesis of qualitative data will lend significant value to unpacking the meaning behind study results identified in a quantitative approach. Quantitative meta-analysis has helped us to identify relationships between variables; qualitative synthesis may help us to understand *how* and *why* these relationships hold. In future applications of RQ+ at IDRC, and synthesis of findings, we will aim to undertake quantitative and qualitative synthesis. There is much to learn by doing so.

Notes

1. See, for example, Lebel and McLean (2018), McLean (2018), Ofir (2016), Ofir et al. (2016), Singh et al. (2013), and Mendez (2012).
2. We recommend these studies for readers seeking to more fully deconstruct the underpinnings of the RQ+ approach. For the purposes of presenting our analysis of RQ+ metadata, we do not unpack the literature and empirical review they provide in this manuscript.

3. We note that the decision on what evidence was required to reach a judgment on any particular dimension was left to the expert opinion of each external review team. IDRC provided teams with a package of research outputs and a list of relevant stakeholders for each independent project in the sample. How these data were interrogated and weighed was independently decided upon by the reviewers to ensure neutrality. Reviewers were allowed and encouraged to move beyond the initial resources provided by IDRC.
4. The score for each main dimension for each of the 170 components was obtained by taking the simple average of the individual scores for each subdimension that was part of the main dimension. For example, to obtain the score for positioning for use, the average of the scores for knowledge accessibility and sharing and timeliness and actionability was obtained. If there were no scores for any of the subdimensions, that particular score was not computed for the corresponding dimension. That is to say there was no downward bias on aggregate scores from a null or zero score being assigned before aggregation.
5. We preferred ANOVA over multivariate regression methods (such as ordinary least squares) in our analysis of the data because the former approach makes less stringent assumptions on the structure of the data (e.g. ANOVA does not assume that the explanatory variables are not collinear).
6. Note that there were fewer observations available for research legitimacy than for the other dimensions. This is primarily due to the fact that reviewers did not score the subdimension 'Addressing Negative Consequences', as this subdimension was deemed 'not applicable' or 'unable to assess' in the projects that were being reviewed. As noted in text earlier, in our meta-analysis, this does not lead to downward bias aggregate scores for any dimension.
7. It should be noted that even those who are not new to a field may also undergo capacity building, though this is less likely.
8. This does not mean that all local knowledge is necessarily wholly generated within a particular national or subnational context: the role of external experts is often crucial in enhancing the knowledge base of local researchers.
9. Here, by Northern research, we mean research projects that are led by Northern-based researchers but which may also have Southern researchers in the team. And vice versa for the Southern-based data. We did not assess the citizenship of all researchers in our sample, or any other indicator of origin such as place of birth. The data are based on the location of the grant recipient, and where grant monies were managed from and expended.

Acknowledgements

We are grateful to two anonymous reviewers of this manuscript and its lead editor Dr Diana Hicks for thoughtful and important critique. We acknowledge the colleagues and peers who contributed through the development of a working paper version of this manuscript (these reviewers listed in: McLean and Sen 2018). We acknowledge multiple editors at *Nature* for their feedback during a seminar presented on the RQ+ approach. Finally, we acknowledge our colleagues at the International Development Research Centre and its research and evaluation community who contributed to the many phases of this project.

References

- Bill and Melinda Gates Foundation (BMGF) (2017) Grand Challenges Website. <<http://gcgh.grandchallenges.org/>>
- Bornmann, L. (2013) 'What is Societal Impact of Research and How Can it be Assessed? A Literature Survey', *Journal of the American Society for Information Science and Technology*, 64/2: 217–33.
- Bowen, S., and 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'', *International Journal of Health Policy and Management*, 4/8: 545–7.
- Carden, F. (2009), *Knowledge to Policy: Making the Most of Development Research*, Los Angeles: Sage Publications and IDRC.
- Cash, D. et al. (2003) 'Salience, Credibility, Legitimacy and Boundaries: Linking Research, Assessment and Decision Making (November 2002)', KSG Working Papers Series RWP02-046. <<https://ssrn.com/abstract=372280>>
- Centre for Research on Evaluation, Science and Technology (CREST) (2014) Science Granting Councils in Sub-Saharan Africa. Commissioned Report. Stellenbosch: Centre for Research on Evaluation, Science and Technology.
- D'Este, P. et al. (2018) 'How Do Researchers Generate Scientific and Societal Impacts? Toward an Analytical and Operational Framework', *Science and Public Policy*, 45/6: 752–63.
- DFID (2014) What is the evidence on the impact of research on international development?, <<https://assets.publishing.service.gov.uk/media/57a089aced915d622c000343/impact-of-research-on-international-development.pdf>>
- DFID (2016) *DFID Research Review*. <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/564075/Research-review4.pdf>
- Greenhalgh, T., and Wieringa, S. (2011) 'Is It Time to Drop the 'Knowledge Translation' Metaphor? A Critical Literature Review', *Journal of the Royal Society of Medicine*, 104: 501–9.
- Greenhalgh, T. et al. (2016) 'Research Impact: A Narrative Review', *BMC Medicine*, 14: 78. <<http://doi.org/10.1186/s12916-016-0620-8>>
- Gurevitch, J. et al. (2018) 'Meta-Analysis and the Science of Research Synthesis', *Nature*, 555: 175–82.
- Hicks, D. et al. (2015) 'The Leiden Manifesto for Research Metrics', *Nature*, 520: 429–31.
- Holmes, B. (2016) 'The Rise of the Impact Agenda', Keynote Speech Delivered at the Fuse International Conference on Knowledge Exchange, April 27th 2016. Newcastle. <<https://drive.google.com/file/d/1rjfOo7hF9or1Rdd0TI0lp4-98ZWNuTKP/view>>
- IDRC Digital Library. (2017) External Program Reviews 2015. <www.idl-bnc.idrc.ca/dspace/simple-search?query=external+reviews+2015>
- International Development Research Centre (2017) *Towards Research Excellence for Development: The Research Quality Plus Assessment Instrument*. Ottawa, Canada. <https://www.idrc.ca/sites/default/files/sp/Documents%20EN/idrc_rq_assessment_instrument_september_2017.pdf>
- Lebel, J., and McLean, R. (2018) 'A Better Measure of Research from the Global South', *Nature*, 559/7712: 23–6.
- Liu, J. (2015) 'Statistical Power in Meta-Analysis', Doctoral dissertation. <<http://scholarcommons.sc.edu/etd/3221>>
- McLean, R. (2018) *Credibility and Research Quality: Time for a Paradigm Shift?* London, UK: On Think Tanks Annual Review. <<https://onthinktanks.org/publications/on-think-tanks-2017-annual-report-credibility/>>
- McLean, R. K. D., and Feinstein, O. N. (2016) Research Quality plus (RQ+)—Reflections on Application at IDRC', 21st International Conference on Science and Technology Indicators. Valencia, Spain. September 14-16, 2016. <<http://ocs.editorial.upv.es/index.php/STI2016/STI2016/paper/viewFile/4543/2327>>
- McLean, R.K.D., and Sen, K. (2018) *Making a Difference in the Real World? A Meta-analysis of Research for Development*. Ottawa, Canada: IDRC. <<http://hdl.handle.net/10625/56969>>
- Mendez, E. (2012) *What's in Good?* Ottawa, Canada: IDRC. <<https://www.idrc.ca/sites/default/files/sp/Documents%20EN/Lit-review-Final-English.pdf>>

- Muirhead, B., and Harpelle, R. N. (2010) *IDRC: 40 Years of Ideas, Innovation, and Impact*, Wilfrid Laurier University Press. <<https://www.wlupress.wlu.ca/Books/I/IDRC>>
- Nutley, S., Walter, I., and Davies, H. T. O. 2007. *Using Evidence: How Research Can Inform Public Services*. Bristol, UK: The Policy Press.
- Ofir, Z. (2016) The Research Quality plus Assessment Framework. On Think Tanks—MEL for Think Tanks Blog Series. <<https://onthinktanks.org/articles/the-research-quality-plus-rq-assessment-framework/>>
- Ofir, Z. et al. (2016) *Research Quality Plus (RQ+)—A Holistic Approach to Evaluating Research*. Ottawa, Canada: IDRC. <www.idrc.ca/sites/default/files/sp/Documents%20EN/Research-Quality-Plus-A-Holistic-Approach-to-Evaluating-Research.pdf>
- Science Granting Councils Initiative (SGCI) (2016) Science Granting Councils Initiative Website. <<http://sgciafrica.org>>accessed 10 Dec 2016.
- Singh, S. et al. (2013) *Excellence in the Context of Use-Inspired Research: Perspectives of the Global South*. Ottawa, Canada: IDRC. <<https://www.idrc.ca/sites/default/files/sp/Documents%20EN/Perspectives-of-the-global-south-Full-paper.pdf>>
- Stern, N. (2016) Building on Success and Learning from Experience: An Independent Review of the Research Excellence Framework. <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/541338/ind-16-9-ref-stern-review.pdf>
- USAID (2017) Website of the Global Development Lab of USAID. <<https://www.usaid.gov/GlobalDevLab>>
- Wilsdon, J. et al. (2015) *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. Bristol: Higher Education Funding Council of England.