RQ+
Research Quality Plus

Making a difference in the real world?
A meta-analysis of research for development

Robert McLean
Senior Program Specialist, Policy and Evaluation Division,
International Development Research Centre

Kunal Sen
Professor, Global Development Institute,
University of Manchester

IDRC | CRDI   Canada

Acknowledgements

The authors express sincere thanks to the following peers for their critical review, insights, and encouragement through the development of this paper. Any errors belong to the authors alone.

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.

In this paper, we provide a meta-analysis of the quality of research supported by Canada's International Development Research Centre. The meta-analysis examines a large and unique data-set that comprises 170 studies undertaken between 2010 and 2015. The research in the sample 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 applied – **Research Quality Plus, RQ+** – incorporates argumentation espoused in a growing international and interdisciplinary call for reform in the way research evaluation is conducted. As such, this paper presents a case study of doing research evaluation differently, and what the results can look like for research policy makers. The analysis suggests that contrary to conventional wisdom, there is no clear trade-off between the rigour 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 best positioned to innovate a solution. The results present novel evidence for consideration by those supporting, using, and doing research for development.

1 Introduction

Scientific research is an indispensable component of social progress. In the Global South, this holds just as true. High quality, use-oriented, and well communicated research can improve 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 United States, the Global Development Lab of USAID 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 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 2018).

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 in spite of 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. Two general postulates have dominated this sphere:

1) First, 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 amongst other researchers (Hicks et al. 2015).

2) Second, 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)

to demonstrate the social impact of research investments (Hicks et al. 2015; Wilsdon et al. 2015; Holmes 2015). 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 incorporated the assessment of research impact 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 a complex and iterative process of social interactions with research users, beneficiaries, and other intended and unintended stakeholders (D'Este 2018; Bowen & Graham 2015; Greenhalgh & Wieringa 2011; Nutley et al. 2007).

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

In this paper, 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 scientific research for the development priorities of Southern countries. The majority of the research is undertaken in Southern countries by Southern researchers, spans scientific disciplines from economics to neuroscience, and accepts multi- and trans-disciplinary approaches common in fields such as agriculture or climate change. (A detailed account of the historical experience of IDRC is available in Muirhead & Harpelle 2010.)

We conduct the meta-analysis using the Research Quality Plus (RQ+) approach. This approach was developed – by IDRC and its research community in collaboration with Zenda Ofir and Thomas Schwandt – in response to our own challenges in assessing the quality of the research we support in a way that reflects our organizational values. 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.

Roadmap

In the following section of this manuscript, we provide an overview of the RQ+ approach and the RQ+ assessment framework applied at IDRC – our dataset's underpinning evaluative framework and eligibility criteria for 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 paper, 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 across the research for development system.

2 The Research Quality Plus (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 multi-dimensional 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, and a reflection on the first implementation of the approach are presented in Ofir et al. 2016. Here we present a summary overview of the approach and the assessment framework in order to position our meta-analysis. To our knowledge, the RQ+ approach has been used primarily for the assessment of research for development. It is a publically accessible tool and we see no reason why it would not apply, given appropriate tailoring, outside of this context.

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, is solutions-oriented, and occurs within a diversity of contexts. The synergies, challenges, and tensions of producing socially relevant and scientifically meritorious research are well described and debated in the academic literature. Nevertheless, fewer practical contributions to how this research can be evaluated have been presented, and fewer still have been validated with systematic testing (D'Este et al. 2018; Greenhalgh et al. 2016; Bornmann 2013; Mendez 2012). Accordingly, developing 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 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

---

[1] See for example: Lebel & McLean 2018; McLean 2018; Ofir 2016; McLean & Feinstein 2016; Ofir et al. 2016; IDRC 2014; Singh et al. 2013; 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.

is no agreement on the quality dimensions that should be used to evaluate research. There are also large debates around the mechanisms used to evaluate research excellence (e.g., peer review and bibliometric analysis). This paper 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 dataset 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. The most impactful of these was arguably 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 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.

The RQ+ 'Non-negotiables'

The RQ+ approach embraces three central postulates for credible and comprehensive research quality evaluation. These are:

1) **Accept a multi-dimensional view of quality that is based on the values and objectives that drive a research agenda**. For IDRC scientific rigor is a non-negotiable. Given its interest in research for development, however, a complete picture of quality to IDRC 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 have 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, judgements 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).

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. It should hence 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 analysed 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: i) research quality dimensions and sub-dimensions, ii) contextual factors, and iii) evaluative rubrics. These components are presented in turn hereafter.

i) Research quality dimensions and sub-dimensions

Ofir et al. (2016) describe one 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'.

These values were technically categorized as research quality dimensions and sub-dimensions. The four principal quality dimensions in RQ+ as applied in this exercise were: a) research integrity, b) research legitimacy, c) research importance, and d) 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 selected research outputs. In their scoring, 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.

Research legitimacy considered the extent to which research results were produced by a process that was procedurally fair and accounted for and included the concerns, insights, values, and perspectives of relevant stakeholders. 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 sub-dimensions: i) addressing negative consequences – that is, the potentially negative consequences and outcomes for populations, ii) gender responsiveness – that is, how reactive to gender issues is the project, iii) inclusiveness – that is, whether the project is inclusive of vulnerable populations, and iv) 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 sub-dimensions: i) originality of the research, and ii) the relevance of the research.

Positioning for use considered the extent to which the research process was managed, and research products/outputs prepared in a way that enhanced the probability of use, influence, and impact. The incorporation of this dimension in the RQ+ framework was guided by the understanding that the uptake of research is an inherently political process. Preparing for it therefore required attention to user contexts, accessibility of products, and 'fit for purpose' engagement and dissemination strategies. It also required 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. Strategies to integrate potential users into the research process itself, wherever feasible and desirable, were also considered.

Figure 1 presents a visual representation of the multi-dimensional nature of research quality expressed in the RQ+ approach (it includes the dimension of Research Integrity, and all sub-dimensions).

Figure 1. Research Quality as Multi-Dimensional

ii)   Contextual factors

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

The first is **the maturity of the research field**, which is the extent to which theoretical and conceptual frameworks, from which well-defined hypotheses have been developed and subjected to testing, are established. It also considers whether a substantial body of conceptual and empirical research in the research field exists.

The second factor is **research capacity strengthening**, which is the extent to which the research endeavor or project focuses on strengthening research capacities. This includes the provision of financial and technical support to enhance capacities for identifying and analyzing development challenges, and for conceiving, conducting, managing, and communicating 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. "Supportive" can, for example, refer to institutional priorities, incentives, and infrastructure.

The fourth factor is **risk in the political environment**, which is the extent to which external risks could arise as a result of political and governance challenges, and affect the conduct of the research or its positioning for use. These risks 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. This factor also considers whether the research environment is data rich or data poor.

Figure 2. The Dynamism of Research Quality

iii) Evaluative rubrics

The final component of RQ+, **the evaluative rubrics**, set judgement criteria for reviewers, clarifying how performance should be measured for each dimension and sub-dimension 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 a systematic approach to evaluation judgement, which allowed for the meta-evaluation that follows in this manuscript.

In terms of research quality dimensions and sub-dimensions, the rubrics used graduated levels of achievement. Each sub-dimension for Research Legitimacy, Research Importance and Positioning for Use, and the principal dimension of Research Integrity was scored from 1 to 8. Scores of 1 or 2 indicate unacceptable levels of achievement. Scores of 3 or 4 are less than acceptable, while scores of 5 or 6 are acceptable to good. Scores of 7 or 8 signify very good achievement. A range of 1 to 8 allowed enough variation in the levels of achievement of the project in different sub-dimensions/dimensions. Once scores were allocated for the sub-dimensions 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 employed a 3-point rubric. In terms of maturity of the research field, projects in established fields scored 1 point, those in emerging fields scored 2, and those in new fields scored 3. Projects where research capacity strengthening is of low focus scored 1 point. Projects scored 2 points where research capacity strengthening is of medium focus, and 3 where research capacity strengthening is of high focus. Projects judged as low risk in their research, political or data environments scored 1 point. Projects judged as medium risk scored 2 points, while those judged as high risk scored 3. Below, in Figure 3, we provide an example for how the RQ+ framework, focusing on the RQ+ sub-dimension of engagement with local knowledge, is applied in practice. Figure 4 provides a picture of all three parts of the RQ+ framework as a whole.

Figure 3. Example of the Evaluative Rubric for Engagement with Local Knowledge

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 4. The Components of the IDRC-tailored RQ+ Assessment Framework

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.

LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4
N/A 1 2	3 4	5 6	7 8

Research quality rubrics and qualitative measures
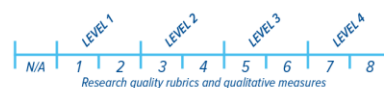
3 Methods

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

Meta-Analysis & 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 dataset. This increases the point precision and generalizability of results (Gurevitch et al. 2018; Liu 2015).

In 2015, seven external evaluations of IDRC supported research, which had embedded the RQ+ approach (as well as assessments of program strategies and outcomes) were completed. The RQ+ data from these seven evaluations comprise the metadata we analyze and present in this research paper. These studies were selected based on their common methodological approach to assessing research quality (RQ+), the commonality of the data, and eligibility for 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 interviewed the project staff responsible for administering the projects, researchers involved in the project and, to the extent possible, key research users (such as policy-makers 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. We believes this empirical element adds strength to the validity and reliability of the meta-analysis results. The data collection for the RQ+ assessment was integrated into the broader program evaluation process which helped to mitigate the time and resources required.

The aggregate metadata includes 170 components from 130 discretely funded research projects funded by IDRC between 2010 and 2015. This research happened around the world, with the majority in Africa, Asia, the Caribbean, Latin America, and the Middle East. Table 1 below presents some further descriptive characteristics of the metadata.

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

Table 1. Descriptive Characteristics of the Metadata

Aggregate Sample Size	170 (170 unique components of 130 discretely funded research projects; all projects received funding from IDRC.)
Broad Disciplines/ Areas of Research	Climate change Water Ecohealth (Environment and health intersections) Governance Security Justice Urbanization and cities Inclusive growth Enterprise development Employment opportunities Non-communicable diseases Tobacco policy and taxation Healthy diets and lifestyles Health equity Health systems governance and financing Information and networks for communications Open data for development
General Characteristics of IDRC-supported Research for Development	Use-oriented toward the development priorities of greatest local need Multi, inter, trans – disciplinary Mixed-method Aims to address complexity Aims to respect and integrate local voice, knowledge, and ways of knowing Aims to be responsive to vulnerabilities
Regional Focus of Research	Global Sub-Saharan Africa Middle East and North Africa Caribbean and Latin America Asia
Institutional Location of Research	Academic institutions (e.g., university) Research institutions (e.g., independent think tank) Government agencies Non-governmental organizations (both international and local)
Research Project Financial Size	Range from 50,000 CAD to 5,000,000 CAD

Statistical Analysis

We first analysed the data using summary statistics – mean, standard deviation, minimum and maximum values of each RQ+ dimension/ sub-dimension score for the 170 components.[3] We next conducted one way Analysis of Variance (ANOVA) tests for different categorisations 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 categorisations.[4] We conducted omnibus F tests, where the null hypothesis of no difference between the means of the population sub-samples was tested across each of the data categorisations in the sample. If the null hypothesis is rejected, then we can infer that at least one of the population sub-sample means is different from the other. However, the F test cannot tell us which mean is different from the others. To determine which means are different, we used a multi-comparison method, the Tukey t-test. This allows us to test which mean of a specific RQ+ dimensions for a particular population sub-sample is different from the means of the same RQ+ dimension for the other population sub-samples. The test compares the difference between 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/ sub-dimension scores to assess the relationship within and between contextual factors and research quality. We used non-parametric Spearman correlations due to the ordinal nature of the data. The level of significance was set at 5 per cent. The analysis was undertaken using STATA version 14.0.

---

[3] 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 sub-dimension that were 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.

[4] We preferred ANOVA over multivariate regression methods (such as Ordinary Least Squares) in our analysis of the data as 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).

4 Results

We begin with an examination of the contextual factors measured across 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 2). For the other contextual factors, most projects were in established or emerging fields, or low to medium risk.

Table 2. Results of RQ+ Analysis for the Entire Sample

	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
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

Turning to the RQ+ quality 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 as 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. Engagement with local knowledge exhibits the highest level of achievement, with a mean of 6.29. Within the Research Importance dimension, relevance has a significantly higher score (6.71) than the originality sub-dimension (5.98). Within the Positioning for Use dimension, there is little difference in the level of achievements between the two sub-dimensions – knowledge accessibility and sharing, and timeliness and accessibility scored 5.94 and 5.65, respectively.

When we disaggregate the RQ+ quality dimensions by regions, we find that the highest levels of achievement are in Latin America, while 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).

Figure 5. RQ+ Quality Dimensions by Region of Research Focus
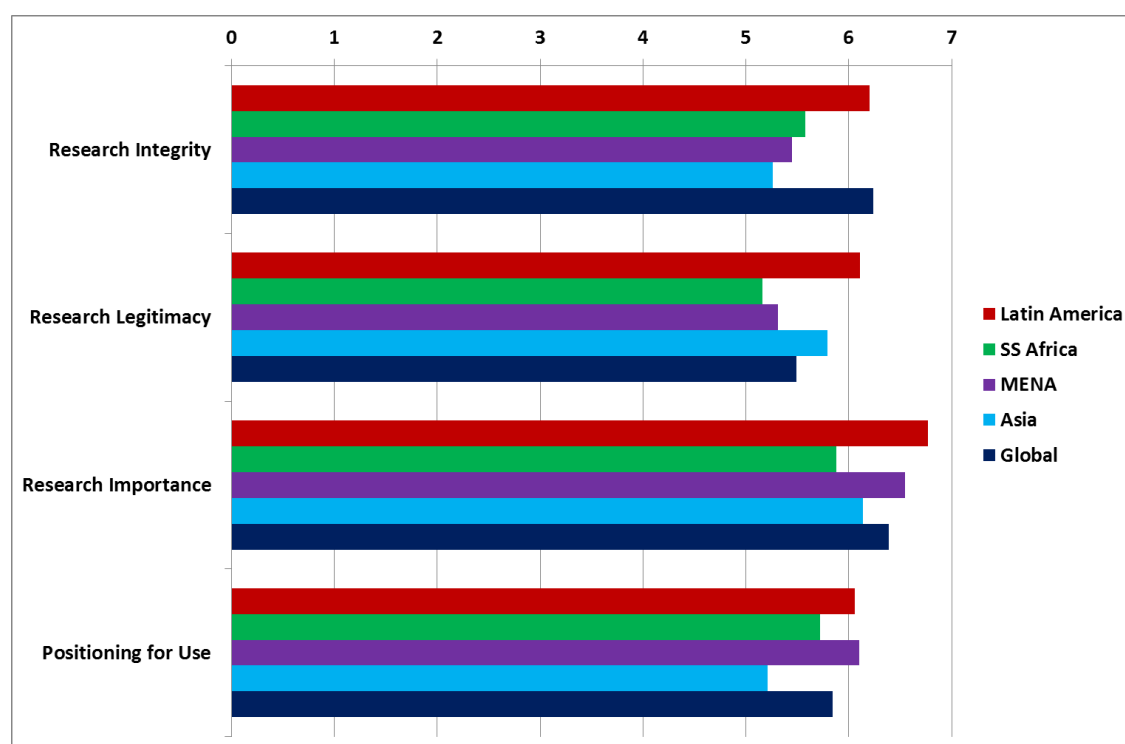
Notes: Total sample = 170. Within this: LAC = 54, SSA = 36, MENA = 11, Asia = 39, Global = 30.

Disaggregating the RQ+ quality dimensions by recipient institution type, we find the average score for Research Integrity is highest for research institutions. For Research Legitimacy, it is highest for NGOs/INGOs. For Research Importance, research institutions achieve the highest score. For the Positioning for Use dimension, the combination of multiple types of organizations working together scores highest (Figure 6).

Figure 6. RQ+ Quality Dimensions by Recipient Institution Type
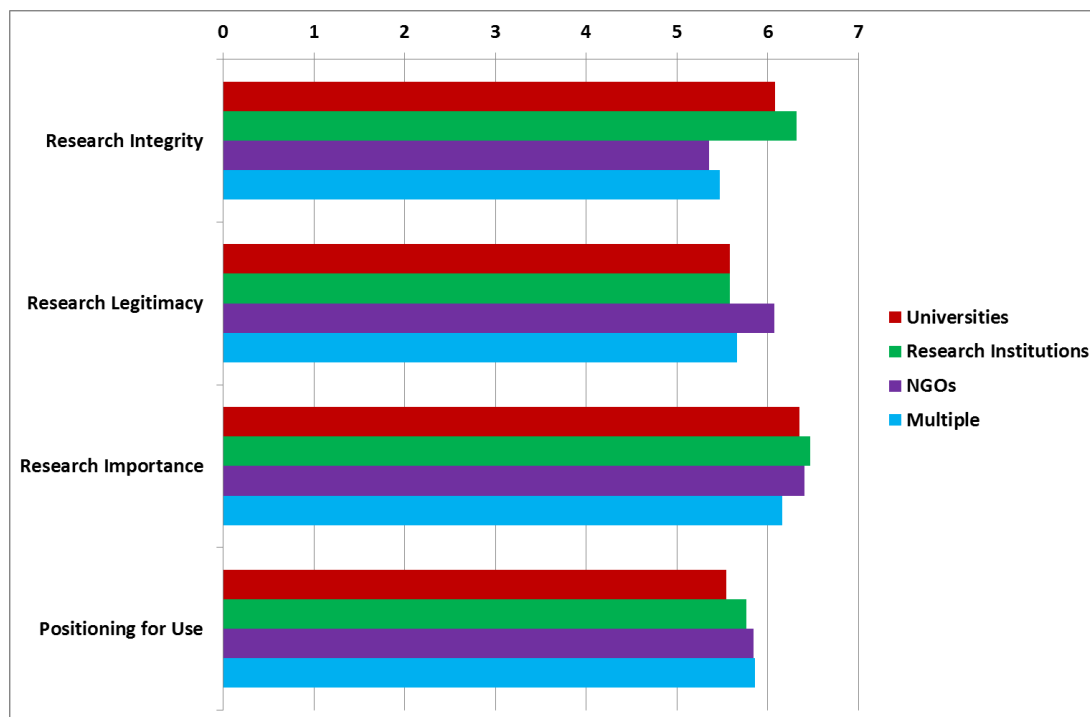
Notes: 1) Total sample = 170. Within this: Universities = 33, Research institutions = 50, NGOs = 44, Multiple = 43. 2) 'NGOs' includes INGOs. 3) 'Multiple' includes any combination of 2 or more recipient types working together.

Categorising the grants by region where the researchers are located (South, North, or Both), we find that Southern-based projects obtain the highest scores in all RQ+ main dimensions (Figure 7).

Figure 7. RQ+ Quality Dimensions by Broad Region of Research
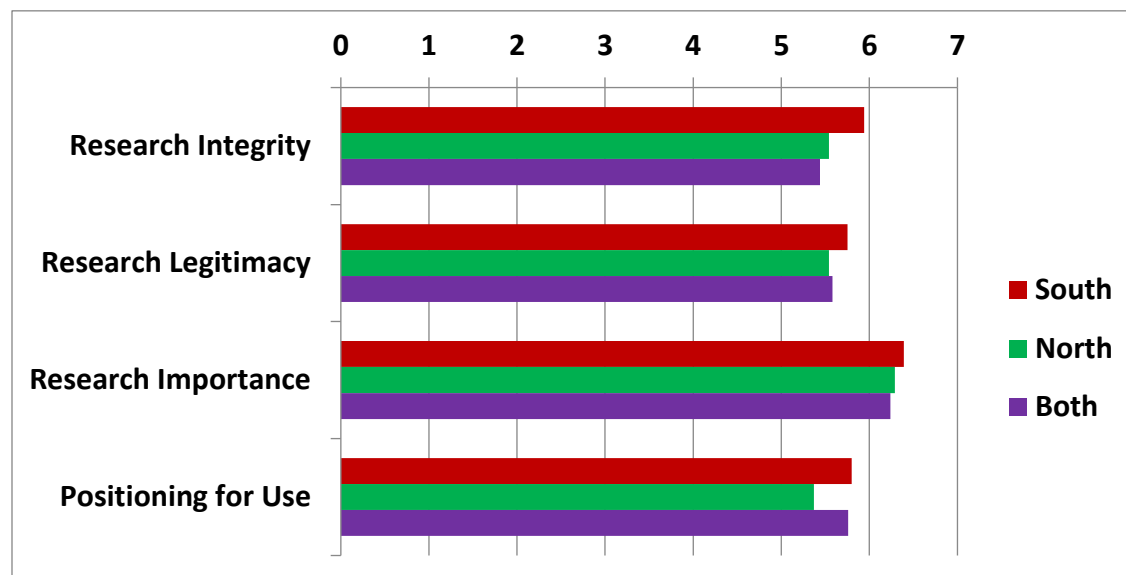
Notes: Total sample = 170. Within this: North = 26, Both = 25, South = 119.

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 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 3). However, when we conduct a pair wise comparison of means, we find that the t-ratio on difference in means for Asia as compared to Latin America is significant in the case of Research Integrity (with the mean for Asia being lower than the mean for Latin America). We also find that the t-ratio on difference in means for Sub-Saharan Africa as compared to Latin America is significant for Research Importance (again, with the mean for Sub-Saharan Africa being 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 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 4). The only t-ratios for difference in means by RQ+ dimension which are significant are for NGOs vs Research Institutions (the mean for NGOs is lower) and for Multiple Recipients vs Research Institutions (the mean for Multiple Recipients is lower).

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 5).

Table 3. 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 1, 5 or 10 per cent level of significance. In each cell, the means of RQ+ main dimensions by regions (as reported in Tables 2-6) are compared, and t-statistics of pair-wise comparisons of means are reported in each row, except the last row, where F-statistic on whether means by regions are different 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 4. 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: ***,** and * indicate whether t-statistic / F-statistic is significant at 1, 5 or 10 per cent level of significance. In each cell, the means of RQ+ main dimensions by recipient institution (as reported in Tables 7-10) are compared. T-statistics of pair-wise comparisons of means are reported in each row, except the 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.

Table 5. 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: ***,** and * indicate whether t-statistic / F-statistic is significant at 1, 5 or 10 per cent level of significance. In each cell, the means of RQ+ main dimensions by broad regions (as reported in Tables 16-18) are compared. T-statistics of pair-wise comparisons of means are reported in each row, except the 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 comprised of: South = 119, North =26, Both = 25.

We then examine the correlations between 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 per cent level) and between research capacity strengthening and Research Legitimacy (correlation coefficient of 0.34, and significant at the 5 per cent level) (Table 5). There is a negative correlation between risk in the research environment on one hand, and Research Integrity, Research Importance, and Positioning for Use on the other. There is a weaker correlation between other contextual factors and the main RQ+ dimensions.

With respect to the correlation between contextual factors and RQ+ sub-dimension measures (Table 7), 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).

Between RQ+ main dimensions, we find strong associations, with statistically significant correlation coefficients in the range of 0.4-0.7. This suggests that projects that score highly in one main dimension also score highly in other dimensions (Table 6).

Table 6. Correlations between Contextual Factors and RQ+ Main Dimensions

	Mat	Cap	RiskD	RiskR	RiskP	Resint	Resleg	Resimp	Posuse
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 per cent or less.

Table 7. Correlation Matrix between Contextual Factors and RQ+ Sub Dimensions

	Mat	Cap	RiskD	RiskR	RiskP	ResInt	Addneg	Genres	Inc	Lockn	Orig	Rel	Know	Timel
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+ Sub Dimensions														
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.

5 Discussion

This study provided a meta-evaluation of the quality of research supported by Canada's International Development Research Centre. The analysis was based on a large and unique dataset that comprises 170 independent expert reviews of research projects supported between 2010 and 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.

Scientifically excellent research is useful research. Conventional wisdom suggests a trade-off between the rigour and the utility of research. In other words, policy-making must often 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. This finding provides evidence for attention to scientific integrity for those investing in research in order to achieve development outcomes.

In research for development, risk and opportunity are diversified. The incidence of internal and external environmental factors is mixed across regions and disciplines, and there is little evidence of correlation between these factors. This data undermines traditional assumptions about the generalized risk of undertaking research in the South. Instead, the environment of Global South 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.

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 helps in understanding its quality.** 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 positively correlated with quality; and quite strongly in the case of capacity strengthening efforts. Political environments have little correlation with quality, except in the case of 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 in order to fully understand quality determinants and draw reasonable conclusions on the quality of any research process.

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 hence suggest that research that requires or includes a focus on capacity strengthening should not be avoided based on a desire for excellence in traditional views of scientific rigour.

We find several compelling correlation coefficients relate to research originality (a sub-dimension 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."*

We confirm the hypothesis that innovative, original research is undertaken by those who are new to a field. 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 that he may have overlooked. **A factor more strongly correlated with originality of science than the fact that it is being undertaken by new researchers, is the degree to which local knowledge is incorporated** (a sub-dimension of Research Legitimacy). In other words, those most closely linked to a problem appear to be best positioned to innovate its solution.

Southern research demonstrates quality, in all RQ+ dimensions. In fact, Southern research demonstrates superior research quality to Northern research and to partnered North-South research. This is not to say that research conducted in the South is categorically superior to research conducted in the North. The dataset examined in this study was comprised of research projects with the objective of improving social outcomes in the Global South. As such, this analysis reinforces the validity of Southern-led research *for* development. When a problem is local, locals appear well placed to address it. Further to this, we suggest that South-North research partnerships may hold great value for interdisciplinary expansion, internationalisation 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 other mutually strategic benefits.

References

Bill and Melinda Gates Foundation. (2017). Global Grand Challenges. Retrieved April 27, 2018, from 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., & 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), 545–547.

Carden, F. (2009). *Knowledge to Policy: Making the Most of Development Research*. Los Angeles: SAGE Publications Ltd and IDRC.

Centre for Research on Evaluation (CREST). (2014). *Science Granting Councils in Sub-Saharan Africa*. Stellenbosch, South Africa.

D'Este, P., Ramos-Vielba, I., Woolley, R., & Amara, N. (2018). How Do Researchers Generate Scientific and Societal Impacts? Toward an Analytical and Operational Framework. *Science and Public Policy*, 1–12.

DFID. (2016). *DFID Research Review*. London, UK. Retrieved April 27, 2018 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/564075/Research-review4.pdf

DFID. (2014). *Rates of Return to Research: A Rapid Review*. London, UK.

Greenhalgh, T., Raftery, J., Hanney, S., & Glover, M. (2016). Research impact: a narrative review. *BMC Medicine*, *14*(1), 78.

Greenhalgh, T., & Wieringa, S. (2011). Is it time to drop the "knowledge translation" metaphor? A critical literature review. *Journal of the Royal Society of Medicine*, *104*(12), 501–509.

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, *555*, 175.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429.

Holmes, B. (2016). The Rise of the Impact Agenda. In *Fuse International Conference on Knowledge Exchange*. Newcastle, U.K.

IDRC. (2016). External Program Reviews. Retrieved April 27, 2018, from https://www.idrc.ca/en/article/external-program-reviews-2015

IDRC. (2017). *Towards Research Excellence for Development: The Research Quality Plus Assessment Instrument*. Ottawa, Canada. Retrieved April 27, 2018 from https://www.idrc.ca/sites/default/files/sp/Documents EN/idrc_rq_assessment_instrument_september_2017.pdf

Lebel, J., & McLean, R. (in press). Research Quality Evaluation – A Southern Experience. *Nature*.

Liu, J. (2015). *Statistical Power in Meta-Analysis*. University of South Carolina.

McLean, R. (2018). Credibility and research quality- time for a paradigm shift? *On Think Tanks 2017 Annual Review: Credibility*.

McLean, R.K.D., & Feinstein, O.N. (2016). Research Quality Plus (RQ+) – Reflections on Application at IDRC. In *21st International Conference on Science and Technology Indicators*. Valencia, Spain.

Méndez, E. (2012). *What's in Good?* Ottawa, Canada.

Muirhead, B., & Harpelle, R. N. (2010). *IDRC: 40 Years of Ideas, Innovation, and Impact*. Wilfrid Laurier University Press.

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

Ofir, Z. (2016, September 30). The Research Quality Plus (RQ+) Assessment Framework. *On Think Tanks - MEL for Thunk Tanks Blog Series*. Retrieved April 27, 2018 from https://onthinktanks.org/articles/the-research-quality-plus-rq-assessment-framework/

Ofir, Z., Schwandt, T., Duggan, C., & McLean, R. (2016). *Research Quality Plus (RQ+) – A Holistic Approach to Evaluating Research*. Ottawa, Canada. Retrieved April 27, 2018 from www.idrc.ca/sites/default/files/sp/Documents EN/Research-Quality-Plus-A-Holistic-Approach-to-Evaluating-Research.pdf

SGCI. (2018). Science Granting Councils Initiative. Retrieved December 10, 2018, from http://sgciafrica.org

Singh, S., Dubey, P., Rastogi, A., & Vail, D. (2013). *Excellence in the Context of Use-Inspired Research: Perspectives of the Global South*. Ottawa, Canada. Retrieved April 27, 2018 from https://www.idrc.ca/sites/default/files/sp/Documents EN/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*. London, U.K. Retrieved April 27, 2018 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/541338/ind-16-9-ref-stern-review.pdf

USAID. (2017). U.S. Global Development Lab. Retrieved April 27, 2018, from https://www.usaid.gov/GlobalDevLab

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., … Johnson, B. (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management.