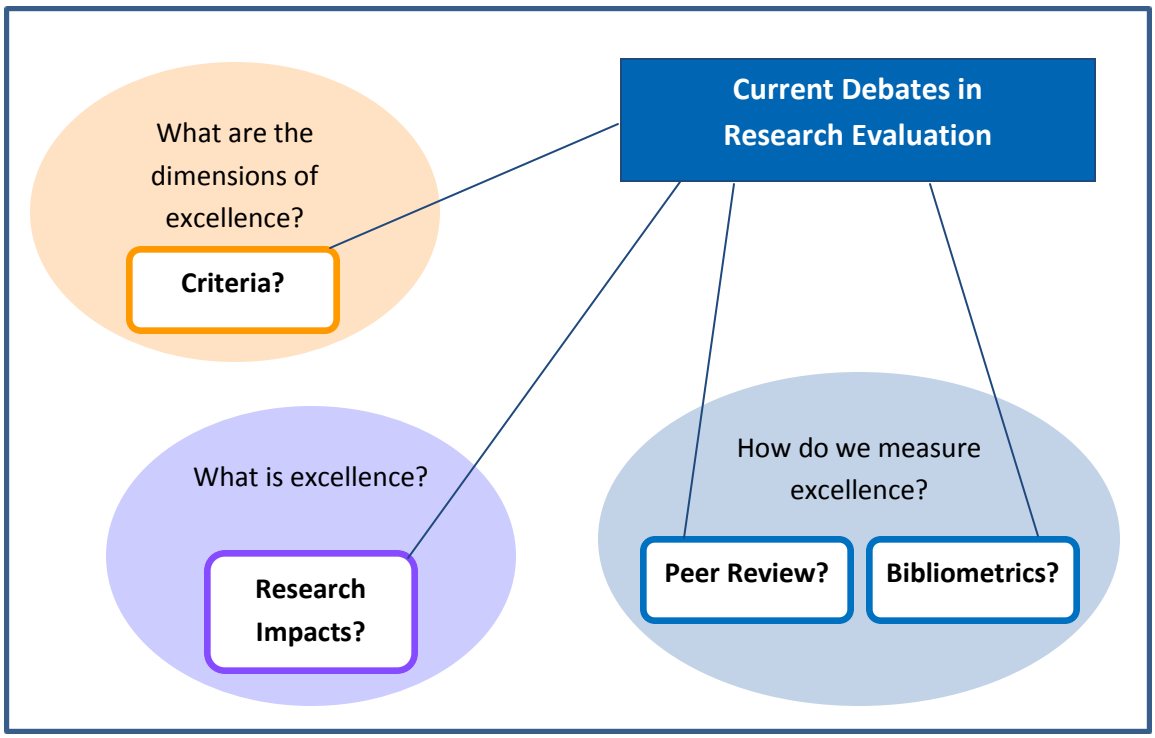
Evaluating Research Excellence[1]: Main Debates



Report written by Ethel Méndez

Introduction

The purpose of research for development goes beyond generating new knowledge to generating knowledge that can improve development outcomes. One could argue that excellence in research is desirable in any type of research; but, the stakes are higher when findings are meant to influence decisions that affect people's lives, the environment, governance, or other areas of development. Research findings gain credibility and are more likely to be used if they derive from excellent research. But, if excellence in research is important, how do we know the good from the bad? How do we evaluate excellence in research? Which criteria do we use?

The International Development Research Centre (IDRC), through its Evaluation Unit, is conducting a study on how to evaluate research excellence, particularly of applied interdisciplinary research for development. This brief draws on a literature review conducted as part of the study to present the main de-

[1] There is no consensus on the meaning of research excellence and if, and how, it differs from research quality. Some scholars think of research impacts as part of research quality (Yule, 2010; Boaz, 2003; OECD, 1997) while others note that quality and impact are two different elements that constitute research excellence (Grant, Brutscher, Kirk, Butler, & Wooding, 2010).

bates in research evaluation, namely questions around impact, peer review processes, metrics as indicators of excellence, and criteria that should be used in the evaluation of research excellence.

Impact

The debate around research impacts can be summarized into three questions: What is impact? How do we measure impact? Should impact be a dimension of research excellence evaluation?

A) What is impact?

There is no universal definition for research impacts. Sandra Nutley et al. make a distinction between the conceptual use of research which "brings about changes in levels of understanding, knowledge and attitude" and the instrumental use of research which "results in changes in practice and policy making" (2003, p. 11). Based on those distinctions they identify different forms of research impact: changes in access to research; changes in the extent to which research is considered, referred to, or read; citation in documents; changes in knowledge and understanding; changes in attitudes and beliefs; and changes in behaviour (Ibid, 2003).

The London School of Economics (LSE), through their Impact of Social Sciences project, defines research impact as "an occasion of influence and hence it is not the same thing as a change in outputs or activities as a result of that influence, still less a change in social outcomes" (LSE Public Policy Group, 2011, p. 21). They categorize research impacts into academic impacts, which are instances when research influences actors in academia or universities and external impacts, or instances when research influences other actors outside of academia (LSE Public Policy Group, 2011).

B) How do we measure impact?

The lack of consensus on the definitions of research impact has resulted in different ways of measuring it. For example, using LSE's definition, the internal or academic impacts can be measured by citations in other academic work while the external or non-academic impacts can be measured by references in the "trade press or in government documents or by coverage in the mass media" (LSE Public Policy Group, 2011, p. 5). Other approaches to measuring impact include case studies, non-bibliometric indicators, and self-evaluation (Grant et al., 2010; Tatavarti, Sridevi, & Kothari, 2010; Aizenman et al., 2011).

However, even if impact is well-defined, it takes an uncertain amount of time for research to have influence. This makes it difficult to evaluate impacts in the short-term (Hammerseley, 2008). Similarly, since research often does not directly influence policy, tracing impacts and attributing causation to research is also difficult (Weiss, 1980). Another issue in evaluating research impacts is the subjectivity of reviewers who may hold conflicting views based on the values and epistemological paradigms that underpin their thinking (Yates, 2005).

C) Should impact be considered in research excellence evaluation?

Beyond the issues of definition and measurement, the notion that all types of research should be subject to evaluation of impacts, as is suggested by the Research Excellence Framework (REF) in England, has sparked debate. Tying funding to research evaluation may generate incentives for researchers to focus on topics where they can have faster results, hence compromising academic freedom and the production of research that can have long-term effects (Yates, 2005; Hammersley, 2008).

Peer Review

A British study conducted in 2002 showed that most researchers choose a peer review mechanism for research evaluation even after determining that it is the system with the "least good features and most bad features" in comparison to other options (Wooding & Grant, 2003, p. 25). Despite experimentation to improve the peer review process [i.e.: use of technology and variations in models like market based alternatives, author pays[2], post-publication, open review, and cascade review[3] (Ware, 2011; Rowland, 2002)], it continues to be mired by criticism of its effectiveness and efficiency. The latter has not dislodged it from its predominance in evaluating research.

A) Effectiveness

Criticism of peer review's effectiveness often centers on subjectivity (Yates, 2005), where the results of the evaluation depend on the epistemological and value-based opinions of the reviewers. The subjectivity of judgement unfolds into different types of problems, such as accusations of game-playing, conservative judgements, and change-adverse behaviors in certain disciplines (O'Gorman, 2008). Researchers in new disciplines or who conduct revolutionary research within existing disciplines often find it challenging to break through existing paradigms and to find peer reviewers who have the expertise to judge their work. Peer review has also been criticized for generating unhealthy competition among researchers that can lead to negative reviews or to purposefully delaying comments to have other authors publish first (Rowland, 2002; Roebber & Schultz, 2011; Petit-Zeman, 2003). The effectiveness of peer review in identifying plagiarism or fraud, in looking at policy relevance, or in guaranteeing methodological standards has also been questioned.

[2] Although this arrangement would place some researchers, including many from the global South, at a disadvantage.

[3] In cascade review the rejected author is asked whether "they wish to have their paper resubmitted with its previous reviewer reports attached, thereby reducing the amount of reviewing required by the next journal." (Ware, 2011, p.35)

B) Efficiency

Peer review has been heavily criticized for being resource-intensive. The long and convoluted review process has discouraged reviewers from accepting more papers. Some also complain about the seemingly arbitrary process through which papers are assigned, which they believe places a greater burden on some reviewers (O'Gorman, 2008). Researchers submitting papers also find it disheartening to have to wait months before getting reviews on their papers.

Peer review is also questioned for being expensive (Boaz & Ashby, 2003). For example, the cost of the peer review based Research Assessment Exercise (RAE) in England was one of the main criticisms that prompted its replacement with the REF (Bridges, 2009).[4]

Metrics

Metrics have gained a degree of popularity in research evaluation because they are perceived to be cheaper, less burdensome, and more transparent and objective than peer review processes (Wooding & Grant, 2003). However, even supporters agree that metrics are far from perfect indicators of research excellence (Andras, 2011).

The debate around the use of metrics in research evaluation centers on their validity as measures of quality. Metrics such as citation counts, impact factors, and publication rates rely on the assumption that published research is research of good quality because it has gone through a careful quality review process. Metrics rely on the subjectivity of a peer review process and thus replicate the biases for which peer review is criticized (Boaz & Ashby, 2003). Other factors, such as the reputation of the author or the persistence of a researcher to get published, can play a role in what is published and what is not (Ware, 2011). Therefore, metrics are questioned because they are not indicators of intrinsic quality of research but proxies that can be altered by other factors.

Using metrics in research evaluation is also contested because not all disciplines, geographical regions, and languages are covered by metrics in the same way, which results in bias against certain groups (Coryn, 2006). This bias is evident in the research for development field, where researchers in the global South may not be as concerned about getting published in a top ranked journal as they may be about influencing development outcomes (Tijssen *et al.*,2006). Their research may not be published and the metrics around it may create the impression that it is not high quality research. This becomes more troublesome when funding is tied to quality. Researchers then face the question of whether to conduct research that matters in their context or conduct research that is likely to be published (Tijssen *et al.*, 2006).

[4] The REF was initially proposed as a metrics-based system which would be less costly than the RAE.

Despite these criticisms, metrics hold some promise as complementary elements in research evaluation (OECD, 1997; David, 2008; Tatavarti, Sridevi, & Kothari, 2010).

Common Criteria

Virtually every research methods textbook or handbook includes one or various sets of quality dimensions. Research funders draw from these sources and also craft their own criteria to apply in evaluating research excellence. The literature reviewed for this strategic evaluation cites at least 30 sets of criteria which vary in length, detail, and approach. Looking across the 30 sets revealed recurring conceptual elements and specific criteria used in evaluating research excellence. The conceptual elements - purposivity, relevance, originality, scientific merit, ethics, and impact – and the more specific criteria that unfold from them are summarized in the annex.

Conclusion

These debates highlight two larger issues in research evaluation. First, there is no agreement on what is meant by excellence in research. Disciplines have discussed this at length in the past, with questions around methods and discipline-specific criteria and even criteria according to the type of research that is conducted. The debate about research impact is a more recent addition to the wider attempt to define excellence. Second, the debates on peer review and metrics emphasize another broader question in research evaluation: how do we measure research excellence? Defining research excellence is necessary before measurements are defined. However, in something as broad and diverse as research it is worth asking whether a definition and a type of measurement are even possible or worth pursuing.

This brief outlines the main debates discussed in "What's in Good?," a review of the literature on research excellence. The annex to the brief presents conceptual elements and criteria drawn from 30 frameworks which are examined in more detail in the literature review. The full reference list can also be found in the literature review.

Please contact Colleen Duggan (cduggan@idrc.ca) or Amy Etherington (aetherington@idrc.ca) to learn more about IDRC's Strategic Evaluation on Research Excellence.

References

Aizenman, J., Edison, H., Leony, L., & Sun, Y. (2011). *Evaluating the Quality of IMF Research: A Citation Study.* Washington : Independent Evaluation Office - International Monetary Fund.

Boaz, A., & Ashby, D. (2003). *Fit for purpose? Assessing research quality for evidence based policy and practice.* Retrieved 2011, from ESRC UK Centre fro Evidence Based Policy and Practice: Working Paper 11: http://www.kcl.ac.uk/content/1/c6/03/46/04/wp11.pdf

Bridges, D. (2009). Research quality assessment in education: impossible science, possible art? *British Educational Research Journal* , 497-517.

Coryn, C. L. (2006). The Use and Abuse of Citations as Indicators of Research Quality. *Journal of MultiDisciplinary Evaluation* , 115-121.

David, M. E. (2008, 7, 1,). Research Quality Assessment and the Metrication of the Social Sciences. *European Political Science*, 52-63, Palgrave Macmillan Ltd.

Grant, J., Brutscher, P.-C., Kirk, S. E., Butler, L., & Wooding, S. (2010). *Capturing Research Impacts: A review of international practice.* Cambridge, UK: Rand Europe.

Hammersley, M. (2008). Troubling criteria: A critical commentary on Furlong and Oancea's framework for assessing educational research. *British Educational Research Journal* , 747-762.

Higher Education Funding Council of England . (n.d.). *Higher Education Funding Council of England* . Retrieved January 25, 2012, from Research Excellence Framework : http://www.hefce.ac.uk/research/ref/

LSE Public Policy Group. (2011). *Maximizing the impacts of your research: A Handbook for Social Scientists.* London: LSE Public Policy Group.

Nutley, S., Percy-Smith, J., & Solesbury, W. (2003). *Models of research impact: a cross-sector review of literature and practice.* London: Learning and Skills Research Centre.

O'Gorman, L. (2008, January). *The (frustrating) state of peer review.* Retrieved July 2011, from International Association for Patter Recognition Newsletter: http://iapr.org/docs/newsletter-2008-01.pdf

Organisation for Economic Co-operation and Development. (1997). *The evaluation of scientific research: Selected experiences.* Paris: OECD.

Petit-Zeman, S. (2003, January 16). Trial by peers comes up short. *The Guardian*, http://www.guardian.co.uk/science/2003/jan/16/science.research.

Roebber, P. J., & Schultz, D. M. (2011, April 12). *Peer Review, Program Officers and Science Funding.* Retrieved July 25, 2011, from PLoS One : http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3075261/

Rowland, F. (2002, October). *The Peer Review Process.* Retrieved July 2011, from Joint Information Systems Committee : http://www.jisc.ac.uk/uploaded_documents/rowland.pdf

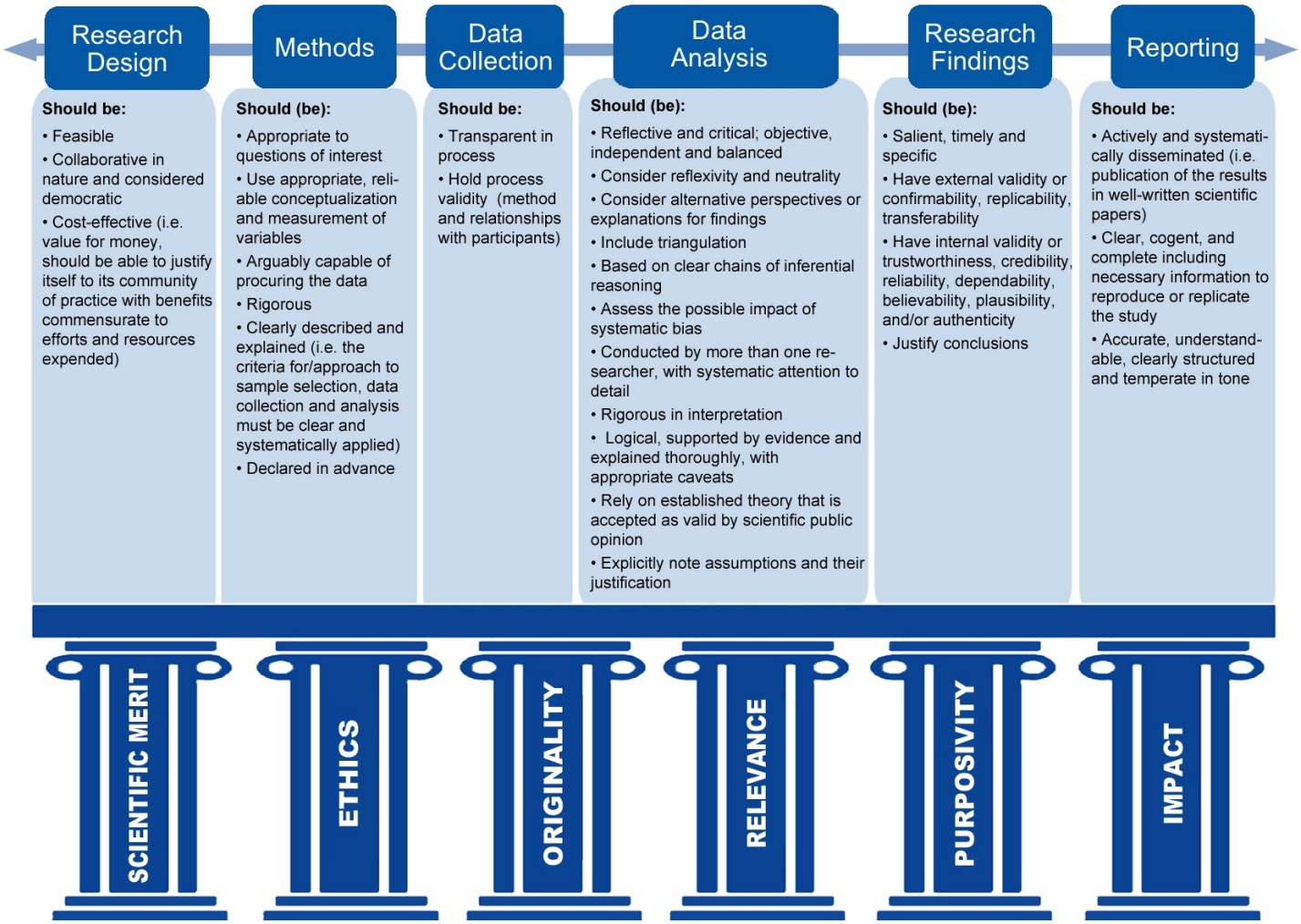
Tatavarti, R., Sridevi, N., & Kothari, D. (2010). Assessing the quality of university research – the RT factor. *General Science*, 1015-1019.

Tijssen, R. J., Mouton, J., van Leeuwen, T. N., & Boshoff, N. (2006, December). How relevant are local scholarly journals in global science? A case study of South Africa. *Research Evaluation, Volume 15* (number 3), 163-174.

Ware, M. (2011). Peer review: recent experience and future direction. *New Review of Information Networking*, 23-53.

Weiss, C. (1980). Knowledge creep and Decision Accretion. *Science Communication*. Volume 1, no.3. (381-404)

Wooding, S., & Grant, J. (2003). *Assessing Research: The Researcher's View.* Retrieved from RA Review: http://www.ra-review.ac.uk/reports/assess/AssessResearchReport.pdf

Yates, L. (2005). Is Impact a measure of Quality? Some Reflections on the Research Quality and Impact Assessment Agendas. *European Educational Research Journal, Volume 4* (Number 4), 391-403.

Annex 1: Common Conceptual Elements and Criteria of Research Excellence in the Research Process