Research Evaluation, 28(2), 2019, 196–201
doi: 10.1093/reseval/rvy037
Advance Access Publication Date: 15 December 2018
Article

OXFORD

A response to Hansson and Polk (2018) "Assessing the impact of transdisciplinary research: The usefulness of relevance, credibility, and legitimacy for understanding the link between process and impact"

Brian M. Belcher[1,2,*], Luisa F. Ramirez[1], Rachel Davel[1] and Rachel Claus[1]

[1]Sustainability Research Effectiveness Program, College of Interdisciplinary Studies, Royal Roads University, Victoria V9B 5Y2, Canada and [2]Centre for International Forestry Research, P.O. Box 0113 BOCBD, Bogor 16000, Indonesia

*Corresponding author. Email: brian.belcher@royalroads.ca

Abstract

Hansson and Polk (2018, Research Evaluation, 27/2: 132–44) aim to assess the usefulness of the concepts of *relevance*, *credibility*, and *legitimacy* for understanding the link between process and impact in transdisciplinary (TD) research. However, the article misrepresents some of the ideas in the two main reference articles. It also uses definitions of the concepts it aims to test that are inconsistent with the definitions offered by the reference papers. The methods description is insufficient to know what data were collected or how they were analyzed. More importantly, the effort to understand relationships between process and impact in TD research needs more careful definitions of the concepts *outcome* and *impact* as well as more objective ways to assess outcomes and impacts. This letter discusses shortcomings in the article and makes suggestions to improve conceptual clarity and methods for empirically assessing TD research effectiveness.

Key words: transdisciplinary research; relevance; credibility; legitimacy; effectiveness; outcome; impact

1. Introduction

Hansson and Polk (2018) aim to assess the utility of three key concepts in transdisciplinary (TD) research. Specifically, they propose to explore the links between the quality of TD research processes, defined in terms of *relevance*, *credibility*, and *legitimacy* (hereafter RCL), and *impacts*. They refer to five TD projects as case studies, using interviews with TD project leaders and participants, and review project documents to assess participants' 'perceptions of the qualities of relevance, credibility, and legitimacy' (136) as well as 'how the perception of those qualities related to impact' (135–6). They conclude that framing the discussion using these three concepts provides useful insights, but find the concepts to be inadequate for evaluating TD research impact.

The intent of the article is important, addressing, as it does, key questions about whether and how TD research contributes to solving sustainability problems and therefore how to design and implement effective TD research. The RCL framework can and has been used in both an evaluative and prescriptive way (Tangney 2017) to assess and guide efforts to enhance the use of scientific knowledge and to bridge the science–policy gap (Sarkki et al. 2014, 2015; Dunn and Laing 2017). Notwithstanding challenges in defining and using the concepts (Heink et al. 2015), the RCL framework is potentially highly valuable as a heuristic tool to understand differing knowledge expectations and demands in linking knowledge to use (Tangney 2017), as well as in guiding the design and implementation of research that aims to be useful and used (Belcher et al. 2016).

Hansson and Polk's (2018) conclusions fundamentally question the potential of the RCL framework. However, there are a number of conceptual, definitional, and methodological problems in the article that undermine the discussion, analysis, and conclusions. This letter highlights the main shortcomings and makes suggestions to improve conceptual clarity and methods for empirically assessing TD research effectiveness.

2. Representation of relevance, credibility, and legitimacy concepts from the reference articles

Hansson and Polk (2018) draw heavily on articles by Cash et al. (2002) and Belcher et al. (2016), both of which define and use the RCL concepts explicitly [n.b. Cash et al. (2002) use the term *salience* instead of *relevance*, but the concepts are essentially the same]. Cash et al. (2002) conceptualize RCL as attributes of knowledge that are needed to be accepted and used by knowledge users, and therefore to be able to cross boundaries to link knowledge to action. Cash et al. (2002) discuss the fact that different actors perceive and value these attributes differently, and that there are thresholds, complementarities, and trade-offs among these attributes requiring 'boundary work' (1) to create and communicate knowledge that is simultaneously relevant, credible, and legitimate to multiple audiences. For the current discussion, a key point is that Cash et al.'s (2002) concepts of RCL refer to attributes of knowledge as perceived by knowledge users. Hansson and Polk (2018) do not refer to the RCL concepts in this way.

Belcher et al. (2016) use the RCL concepts, as well as a fourth concept of *effectiveness* (to be discussed below), as organizing principles to frame sets of criteria for the assessment of research quality in a TD context. These criteria were derived from a systematic review of literature discussing definitions and measures of interdisciplinary and TD research quality. Belcher et al. (2016) follow Cash et al. (2002) in conceptualizing RCL as attributes of knowledge, and draw upon theoretical and empirical discussions in the literature to propose criteria for TD research that would be expected theoretically to produce knowledge that is perceived to be relevant, credible, and legitimate by intended audiences. Those evaluation criteria, which embody the richness of the RCL concepts, are mostly ignored by Hansson and Polk (2018), with the exception of a brief reference to two legitimacy criteria. A key element of the TD Research Quality Assessment Framework (QAF) developed by Belcher et al. (2016) is that it is to be used to assess research according to the purpose of the research being assessed, in a *sui generis* way. In practice, that means anticipating (in an *ex ante* assessment) or assessing the needs of intended or actual users to evaluate the research design and implementation according to whether it will or has produced outputs that are considered to be relevant, credible, and legitimate by those users. The same tool can also be used in a prescriptive way to guide research design and implementation.

Hansson and Polk (2018) misrepresent important elements of both Cash et al.'s (2002) and Belcher et al.'s (2016) articles. Notably, neither of these articles propose to use RCL as criteria for 'evaluating the contributions of a TD approach to societal impact' (Hansson and Polk 2018: 133) as implied by the authors. Likewise, Hansson and Polk's (2018) assertion that the reference papers 'seem to assume that stakeholders can implement usable results once they understand their relevance, and find them credible and legitimate' (133) is incorrect. The point is that stakeholders will be more likely to find research relevant when it asks and answers the right questions, which can be achieved through a comprehensive understanding of the problem context and engagement of stakeholders, as outlined in Belcher et al. (2016).

Considering that the aim of Hansson and Polk's (2018) article is to assess the usefulness of the RCL concepts, it would have been helpful to start with clear definitions of those concepts. In fact, the authors do not provide explicit definitions of the terms as they are using them, and also do not provide full definitions of the terms provided by Cash et al. (2002) or Belcher et al. (2016). Rather, they refer to elements of definitions from the reference papers, often with comments about how they themselves are interpreting and using those definitions [for ease of reference, we provide both sets of original definitions as well as the partial definitions provided by Hansson and Polk (2018) in Table 1]. This leaves the definitions ambiguous, and also misrepresents elements of the others' concepts.

For example, the authors note that Cash et al.'s (2002) definition of salience (relevance) '"deals with the needs of decision-makers"' (as cited in Hansson and Polk 2018: 136), and then they list the kinds of decision-makers involved in their own case study projects. However, their list only includes formal decision-makers (various civil servants, in this case); the discussion does not give due attention to the distinguishing idea in Cash et al. (2002) and Belcher et al. (2016) that knowledge needs to be considered relevant, credible, and legitimate by any stakeholder and not only formal decision-makers. In addition, Hansson and Polk (2018) state that '[*c*]*redibility* is usually defined in scientific standards' (italics original, 136), citing both Cash et al. (2002) and Belcher et al. (2016), implicitly suggesting that they hold exclusively to a traditional scientific definition of credibility. The authors then explain that their own interpretation includes non-scientific input and expertise, as if this is a novel element. This ignores the fact that Belcher et al. (2016) substantially expand the definition of credibility beyond typical disciplinary standards to include TD data, analysis, and methodological and theoretical integration and interpretation. For example, the QAF reflects this in two of the credibility criteria, notably Broad Preparation and Research Approach Fits Purpose (see Belcher et al. 2016: 9, 10, respectively). Moreover, Section 4.3.2 in Belcher et al. (2016) deals explicitly with 'New meanings for familiar terms' (13), which explains and justifies the extension of traditional definitions to encompass social and environmental contextual nuance for TD research. However, Hansson and Polk (2018) make no reference to these key sections in their review. The authors later return to this discussion, stating '[o]ur study further questions scientific credibility as a sufficient trait of TD research processes' (141), which seems to imply that the reference articles propose that credibility is a sufficient trait, which they definitely do not. Furthermore, Hansson and Polk (2018) give little attention to the concept of legitimacy, with only a brief reference to the idea from Cash et al. (2002) that it refers to '"fair and unbiased" processes and results' (Hansson and Polk 2018: 136). They then state '[i]n our TD projects, legitimacy is seen through the extent to which different actors, both research and practice-based, take responsibility for ensuring or creating such "fair and unbiased" processes and impacts' (136). This does not fully reflect the idea that these concepts are intended to apply to users' perceptions, valuation of knowledge, and the way it has been generated (see Table 1). Again, this suggests that the concepts they are testing are inconsistent with the concepts proposed by the reference articles.

Referring to Cash et al. (2002) in a discussion on the different perceptions of RCL, Hansson and Polk (2018) state: '[h]owever, their

Table 1. Definitions of concepts by Cash et al. (2002), Belcher et al. (2016), and Hansson and Polk (2018)

Concept	Cash et al. (2002)	Belcher et al. (2016)	Hansson and Polk (2018)
Relevance (salience)	'Salience refers to the relevance of information for an actor's decision choices, or for the choices that affect a given stakeholder' (4)	'The importance, significance, and usefulness of the research problem, objectives, processes, and findings to the problem context' (9)	'In our analysis, the quality of *relevance* was therefore assessed via different degrees of sensitivity [*sic*] the project context' (italics original, 136)
Credibility	'Credibility refers to whether an actor perceives information as meeting standards of scientific plausibility and technical adequacy. Sources of knowledge must be deemed trustworthy and/or believable, along with the facts, theories, and causal explanations invoked by these sources' (4)	'The research findings are robust and the sources of knowledge are dependable. This includes clear demonstration of the adequacy of the data and the methods used to procure the data including clearly presented and logical interpretation of findings' (9)	'*Credibility* is usually defined in scientific standards (Cash et al. 2003: 8086; Belcher et al. 2016: 8). However, in our TD processes, credibility also includes the importance of how science judges, evaluates, and integrates non-scientific input and expertise into the project process, as well as how practitioners evaluate the appropriateness of different types of scientific expertise for the project goals' (italics original, 136)
Legitimacy	'Legitimacy refers to whether an actor perceives the process in a system as unbiased and meeting standards of political and procedural fairness. Legitimacy involves the belief that S&T systems are "fair" and consider appropriate values, interests, concerns, and specific circumstances from multiple perspectives. Audiences judge legitimacy based on who participated and who did not, the processes for making those choices, and how information is produced, vetted, and disseminated. When connecting knowledge to action, choices are made about which problems and potential solutions will be considered, and which ones will not' (5)	'The research process is perceived as fair and ethical. This encompasses the ethical and fair representation of all involved and the appropriate and genuine inclusion and consideration of diverse participants, values, interests, and perspectives' (10)	'In our TD projects, legitimacy is seen through the extent to which different actors, both research and practice-based, take responsibility for ensuring or creating such "fair and unbiased" processes and impacts' (136)

reasoning implies somewhat homogenous groups where relevance, credibility, and legitimacy are experienced according to similar criteria' (138), suggesting that there are more individualized differences. This is a straw man argument—Cash et al. (2002) fully appreciate and discuss that different stakeholders will evaluate and use knowledge according to their own unique context and perceptions, and Belcher et al. (2016) deliberately apply ideas of appropriateness and adequacy in their framework to encompass the necessary flexibility and specificity for application in diverse contexts.

On the topic of stakeholder involvement and engagement, a defining aspect of TD research, Hansson and Polk (2018) write:

> Belcher et al. also define relevance, credibility, and legitimacy in terms of the content of the research, not in terms of the involvement of stakeholders, except under legitimacy (referring to effective collaboration, genuine, and explicit inclusion) [...] This suggests that researchers take more responsibility for the relevance and credibility of the project, and/or that the involvement of practitioners is taken care of under legitimacy. (142)

This misunderstands and misrepresents Belcher et al. (2016), who strongly emphasize that the engagement of stakeholders is important at all stages in TD research, with specific criteria that address the incorporation of stakeholder ideas and values as well as methodological pluralism and integration reflected in each of the RCL principles. Moreover, Section 4.1.2 further clarifies their position on stakeholder involvement throughout the research process:

> Additional and modified criteria are needed to address the integration of epistemologies and methodologies and the development of novel methods through collaboration, the broad preparation and competencies required to carry out the research, and the need for reflection and adaptation when operating in complex systems. Having researchers actively engaged in the problem context and including extra-scientific actors as part of the research process helps to achieve relevance and legitimacy of the research; it also adds complexity and heightened requirements of transparency, reflection, and reflexivity to ensure objective, credible research is carried out. (8)

It seems that Hansson and Polk (2018) have misconstrued several aspects of the RCL concepts as they were used in the reference articles, which has led to misinformed conclusions about the adequacy of RCL for TD research evaluation.

3. Methods

The description of the methods used by Hansson and Polk (2018) lacks detail and precision, so it is difficult to know what data were

collected. The first paragraph of the methods section refers to interviews and focus group discussions (the actual number of each is unclear) with 'project leaders, project participants as well as the coordinators of the Center partners' (135). No specific information is provided on what questions were asked, but we are informed that the interviews focused on: (1) quality of the project process in terms of the extent of TD executed; (2) project participants' assessment of outcomes; (3) information about internal and external factors 'that support or hinder the successful enactment of the projects' (135); and (4) success factors for TD process and the production of usable results.

Likewise, the methods description does not explain how the data were analyzed. The authors do refer to the use of the 'action-value attribution framework, developed by Hellström (2015)' (Hansson and Polk 2018: 135), but provide little detail on how the framework was used. Regarding the analysis of interviews, the last paragraph of the methods description states: '[i]n the interviews we looked both for how the participants value the internal process and how the projects were received by external actors, both their own organizations and others. We also looked for how the coordinators in the partner organizations perceive relevance, credibility, and legitimacy in relation to their institutional and political context' (136). For the current discussion, it is important to note that Hansson and Polk (2018) emphasize the perceptions of participants on internal project processes and 'how the projects were received by external actors' (136). This attention to perceptions of the project (instead of perceptions of the RCL of the project outputs) is expanded with a focus on the 'internal-external dynamic between projects and project context' (135) and a distinction between what happens internally (within a project) and what happens externally, 'primarily in the organizations of the stakeholders, but also in society at large' (135). The focus on internal project dynamics shifts attention from perceptions of RCL by stakeholders to a discussion of perceptions about collaborative practice within projects by project participants.

The methods description is also insufficient to know how *impact* was assessed. It says only '[t]o explore the process-impact link, we analyzed the material in terms of how the enactment of the TD process contributed to and shaped the perception of relevance, credibility, and legitimacy of the results, as well as how the perception of those qualities related to impact' (Hansson and Polk 2018: 135–6). Hence, the claim that '[o]ur study shows that there are no clear mechanisms that link participatory features to impact' (141) leaves the reader unclear as to how these conclusions were reached. The authors' misinterpretation and misrepresentation of RCL definitions, in combination with unclear methods description, add to the conceptual inconsistencies across the results and discussion.

4. Conceptual inconsistencies

As noted above, the concepts being tested were not clearly defined by Hansson and Polk (2018), and the definitions-in-use shift throughout the article. In the methods section, it is stated that '[r]elevance and legitimacy, for example, cannot be achieved without credibility within and between groups' (136). This use of credibility does not refer to users' perceptions of the quality of knowledge produced by a project; rather, it seems to refer to relationships between participants (individuals and groups) in a research process. This shift in perspective continues in Section 4.3, where a distinction is made between 'internal relevance' (138) and 'external relevance' (139). For instance, the authors discuss how the partner organizations define relevance in terms of capacity building within their own organizations, silo breakdown, and relation-building beyond the contribution of results to solve problems. Another example can be found in the authors' finding that 'the most important component of credibility in these projects was mutual respect and exchange between participants' (139) and an identified need for 'unaligned space' (139) for collaborators to work outside of normal organizational bounds. These are valuable insights for the conduct of TD research, but they relate more to whether project participants and participating organizations appreciated and approved of the collaborative process itself. Hansson and Polk (2018) do not address outputs or analyze how knowledge users evaluated the relevance or credibility of outputs produced by the case study projects.

Further inconsistent usage of terms is found in statements such as: '[i]n one project the high status connected to a local technical university gave credibility to the practitioners, and indirectly to the project and uptake of the project results as well (WISE)' (140), and 'a high degree of legitimacy between a limited number of individuals' (140) among long-standing collaborators. As used in the article, these terms seem to refer to attributes of project participants and socially hierarchical relationships among them as perceived by other project participants; they do not refer to attributes of knowledge as perceived by knowledge users, as advocated by Cash et al. (2002) and Belcher et al. (2016).

Likewise, Hansson and Polk (2018) discuss the importance of joint leadership, ownership, mutual commitment, trust, and adequate time to work as elements of what they refer to as internal legitimacy. This is a narrow meaning of legitimacy, which again strays from the definitions and intent in the reference articles. However, the observations about collaboration and what collaborators value in a collaboration are important. TD research, which is inherently collaborative and requires a broader range of collaborative partnerships than disciplinary research, will need to learn and improve in terms of teamwork. Much can be learned from the literature on organizational behavior, effective teamwork, collaboration (Katzenbach and Smith 1993a, 1993b; Beyerlein et al. 2003; Schein 2004; Bozeman et al. 2016), and the science of team science (Bennett, Gadlin, and Levine-Finley 2010). Belcher et al. (2016) do address some (but not all) of these ideas in their QAF, but with a focus on what is needed to produce outputs that will be perceived as relevant, credible, and legitimate by intended audiences and not on characterizing the RCL of a particular collaboration. The focus by Hansson and Polk (2018) on the characterization by collaborators of other collaborators and of the collaborative process itself as relevant, credible, and legitimate means that their assessment of the usefulness of the concepts largely misses the mark.

5. Assessing outcomes and impacts

The stated aim of the article is to assess the usefulness of RCL for understanding the link between TD research processes and impact. This is indeed a valuable aim, but as noted, the methods description does not provide a clear explanation of how this was done, and the results and subsequent discussion do not seem to realize the aim. To properly assess the link, we would need: (1) a clear characterization of the research process using the RCL framework to assess the outputs of the research process, and (2) a separate and independent assessment of impact.

As discussed above, Hansson and Polk's (2018) testing of the RCL framework is compromised by the ambiguous and inconsistent use of the RCL concepts. The discussion of outcomes and impacts is

undermined by the fact that only research project leaders, participants, and coordinators were interviewed; there do not appear to be data with which to assess external users' perceptions, acceptance, or use of project outputs, let alone any more objective assessment of outcomes and impact. This may be the reason that the brief discussion on external knowledge users' perceptions tends to conflate the three concepts, stating that they are 'tightly coupled' (Hansson and Polk 2018: 140), and '[i]t can be argued that the technical character of the research through its high credibility provided additional external legitimacy to the project' (140). These statements fail to apply the original definitions (Table 1), and therefore miss the opportunity to accurately analyze the usefulness of the RCL concepts as conceptualized by Cash et al. (2002) and Belcher et al. (2016). Other statements about external users' perceptions and appreciation of research outputs are similarly weak. For example, Hansson and Polk (2018) surmise that 'the fact that these projects were responding to preexisting and ongoing processes, goals, and strategies, where mind shifts had already been made, increased their external relevance' (140), which is a tautological argument that projects that addressed topical issues were considered relevant.

The article uses the following terms and definitions for various kinds of project results: outputs are 'usable products' (Hansson and Polk 2018: 134); outcomes are 'enhanced capacity and network effects' (134); and impacts are 'transformational change, such as structural changes' (134). These vague definitions limit the ability to identify outcomes and impacts, and therefore hinder the assessment of relationships between research processes and results.[1]

Such an assessment is further confounded with the puzzling statement that 'TD approaches are built on the assumption that the intermediate or direct effects of participatory research contribute indirectly to transformational societal change, since the latter is difficult to measure, as it is often significantly delayed as well as hard to attribute to specific research' (134). It is not clear why measurement challenges should influence theory about how TD research contributes to change processes. In any case, without a clear definition of outcome or impact, and without any measure, indicator, or even any external opinion about the results of the case study projects, how did the authors or the interview respondents assess *impact*? This remains unanswered.

Although Hansson and Polk (2018) do not discuss it in any depth, Belcher et al. (2016) add a fourth principle, *effectiveness*, along with related assessment criteria. Effectiveness is defined as '[t]he research generates knowledge and stimulates actions that address the problem and contribute to solutions and innovations' (11). Belcher et al. (2016) suggest that potential effectiveness can be assessed *ex ante* at the research proposal stage, and actual effectiveness can be assessed *ex post*. To assess the relationship between TD research design and implementation as independent variables, as well as outcomes and impact as dependent variables, we need reliable empirical methods to assess project results *ex post*.

There are some promising approaches for TD research project outcome assessment such as Outcome Mapping (Earl, Carden, and Smutylo 2001), Contribution Analysis (Mayne 2008, 2012), Collaborative Outcomes Reporting Technique (Dart and Roberts 2014), and the RAPID Outcome Assessment method (ODI 2012). More recently, Belcher, Suryadarma, and Halimanjaya (2017) developed and tested a theory-based research evaluation approach that could be applied in case studies such as those discussed by Hansson and Polk (2018). Belcher, Suryadarma, and Halimanjaya (2017) collaborated with researchers (and in some cases, with stakeholders) to articulate a theory of change (ToC) for completed research projects. A ToC is a set of testable hypotheses about relationships between an intervention and its intended results. The ToC explains how, why, and in what contexts an intervention, such as a research project, is expected to lead to particular outcomes (defined as changes in knowledge, attitudes, skills, and relationships, manifested as changes in behavior that result in whole or in part from the research and its outputs) and impacts (defined as changes in flow or a change in state, resulting in whole or in part from a chain of events to which the research has contributed). Using a ToC, it is possible to identify indicators and define measures of outcomes and impacts at different levels to empirically test whether those outcomes have been achieved and to assess how and why the project has (or has not) contributed to a change process.

6. Conclusion

While we greatly appreciate the aim of the article, we feel that it falls short in several ways to advance understanding of either the concepts of relevance, credibility, or legitimacy in a TD research context, or to assess the relationship between those concepts and TD project outcomes and impacts. Hence, the conclusion that 'our study shows that while the relationship between the traits of relevance, credibility, and legitimacy on knowledge systems and their societal impact is theoretically convincing [. . .] it has limited applicability in TD research' (Hansson and Polk 2018: 141) is not justified. The finding that '[w]hile designing the analysis around these terms gave many interesting reflections, the overall benefit was in understanding how poorly the accepted definitions of such traits mirrored or could be applied to a TD context' (141) reflects on how the concepts were defined and used inconsistently by the authors and not as they were defined or proposed to be used in the reference articles. This weakness is reflected in the concluding statement that 'internal and external dynamics affect the diversity of meanings of relevance, credibility, and legitimacy that are experienced and played out in a specific political context' (142–3); the idea that different stakeholders will evaluate knowledge differently depending on their own particular perspectives is fundamental to both of the reference papers. However, Cash et al. (2002) and Belcher et al. (2016) both take the position that it is useful to analyze how stakeholders evaluate knowledge using fixed concepts of RCL. This approach is more reliable than allowing the meaning of the concepts themselves to shift with different users as Hansson and Polk (2018) have done. Moreover, both reference papers offer suggestions to help make knowledge relevant, credible, and legitimate for stakeholders. Belcher et al. (2016) go further, adding the principle of effectiveness and several related criteria for assessing TD research effectiveness.

For Hansson and Polk (2018) to effectively test the concepts proposed by Cash et al. (2002) and Belcher et al. (2016), the article would need to:

1. provide clear definitions of the RCL concepts (ideally using the definitions from the reference papers) and apply those definitions consistently in the data collection and analysis;
2. accurately represent the key ideas that were developed and presented in the reference papers;
3. provide a clear and adequate description of the methods for data collection and analysis used in the case study research to allow readers to understand fully what was done and how (and allow for replicability);

4. reliably assess the degree to which the case study projects' produced outputs were perceived as relevant, credible, and legitimate by key audiences/stakeholders; and
5. reliably assess the degree to which the case study projects contributed to intended outcomes and impacts.

In the absence of a rigorous test of the RCL concepts, we would suggest not to abandon the RCL framework, but instead to continue work to refine, test, and improve the framework to help realize its potential for evaluating and guiding TD research and other research that aims to create knowledge that is useful and used.

Note

1. Belcher and Palenberg (2018) have reviewed the range of definitions of the concepts *outcome* and *impact* as used in international development and also within a research-for-development context. They recognize the sequential and iterative nature of research contributions to change processes, and make recommendations for good definitional practice, including to clearly distinguish each kind of result.

Acknowledgements

The Sustainability Research Effectiveness Program is supported by funding from the Canada Research Chairs program and the Canadian Social Sciences and Humanities Research Council (SSHRC).

References

Belcher, B., and Palenberg, M. (2018) 'Outcomes and Impacts of Development Interventions: Toward Conceptual Clarity', *American Journal of Evaluation*, 39/4: 478–95.

Belcher, B. et al. (2016) 'Defining and Assessing Research Quality in a Transdisciplinary Context', *Research Evaluation*, 25/1: 1–17.

Belcher, B., Suryadarma, D., and Halimanjaya, A. (2017) 'Evaluating Policy-Relevant Research: Lessons from a Series of Theory-Based Outcomes Assessments', *Palgrave Communications*, 3: 1–16.

Bennett, L. M., Gadlin, H., and Levine-Finley, S. (2010) *Collaboration and Team Science: A Field Guide*. Bethesda, MD: National Institutes of Health.

Beyerlein, M. M. et al. (2003) *Beyond Teams*. San Francisco, CA: Jossey-Bass/Pfeiffer.

Bozeman, B. et al. (2016) 'Research Collaboration Experiences, Good and Bad: Dispatches from the Front Lines', *Science and Public Policy*, 43/2: 226–44.

Cash, D. et al. (2002) 'Salience, Credibility, Legitimacy and Boundaries: Linking Research, Assessment and Decision Making', KSG Working Papers Series RWP02-046, November 2002. SSRN. <https://papers.ssrn.com/sol3/papers.cfm? abstract_id=372280>.

Dart, J., and Roberts, M. (2014) *Collaborative Outcomes Reporting* <http://betterevaluation.org/sites/default/files/COR.pdf>.

Dunn, G., and Laing, M. (2017) 'Policy-Makers Perspectives on Credibility, Relevance and Legitimacy (CRELE)', *Environmental Science and Policy*, 76: 146–52.

Earl, S., Carden, F., and Smutylo, T. (2001) *Outcome Mapping: Building Learning and Reflection into Development Programs*. Ottawa: International Development Research Centre.

Hansson, S., and Polk, M. (2018) 'Assessing the Impact of Transdisciplinary Research: The Usefulness of Relevance, Credibility, and Legitimacy for Understanding the Link between Process and Impact', *Research Evaluation*, 27/2: 132–44.

Heink, U. et al. (2015) 'Conceptualizing Credibility, Relevance and Legitimacy for Evaluating the Effectiveness of Science–Policy Interfaces: Challenges and Opportunities', *Science and Public Policy*, 42/5: 676–89.

Katzenbach, J. R., and Smith, D. K. (1993a) *The Wisdom of Teams: Creating the High-Performance Organization*. Boston: Harvard Business School Press.

Katzenbach, J. R., and Smith, D. K. (1993b) 'The Discipline of Teams', *Harvard Business Review*, 71/2: 111–20.

Mayne, J. (2008) 'Contribution Analysis: An Approach to Exploring Cause and Effect', in *Institutional Learning and Change Brief*. Rome: Institutional Learning and Change (ILAC) Initiative. <http://dmeforpeace.org/sites/default/files/Mayne_Contribution Analysis.pdf>

Mayne, J. (2012) 'Contribution Analysis: Coming of Age?' *Evaluation*, 18/3: 270–80.

ODI. (2012) *RAPID Outcome Assessment Toolkits*. London: Overseas Development Institute. <https://www.odi.org/publications/6800-rapid-outcomeassessment>.

Sarkki, S. et al. (2014) 'Balancing Credibility, Relevance and Legitimacy: A Critical Assessment of Trade-Offs in Science-Policy Interfaces', *Science and Public Policy*, 41/2: 194–206.

Sarkki, S. et al. (2015) 'Adding 'Iterativity' to the Credibility, Relevance, Legitimacy: A Novel Scheme to Highlight Dynamic Aspects of Science-Policy Interfaces', *Environmental Science and Policy*, 54: 505–12.

Schein, E. H. (2004) *Organizational Culture and Leadership*. San Francisco, CA: John Wiley & Sons.

Tangney, P. (2017) 'What Use Is CRELE? A Response to Dunn and Liang', *Environmental Science and Policy*, 77: 147–50.