Environmental Science and Policy 101 (2019) 192–203

Contents lists available at ScienceDirect

Environmental Science and Policy

journal homepage: www.elsevier.com/locate/envsci

Linking transdisciplinary research characteristics and quality to effectiveness: A comparative analysis of five research-for-development projects

Brian M. Belcher[a,b,*], Rachel Claus[a], Rachel Davel[a], Luisa F. Ramirez[a]

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

ARTICLE INFO

Keywords:
Transdisciplinary
Research quality
Engagement
Outcome evaluation
Research effectiveness

ABSTRACT

More and more effective research is needed to help address complex sustainability problems. Many research approaches have adopted more transdisciplinary characteristics as a way to improve effectiveness. However, empirical evidence of the extent to which and how transdisciplinary research design and implementation contribute to (more) effective scientific and social outcomes remains limited. This paper reports a comparative analysis of five research-for-development projects implemented in Peru and Indonesia to: characterize the extent to which projects employed transdisciplinary principles; assess the extent to which and how intended project outcomes were achieved; analyze the relationship between transdisciplinary research approaches and outcomes; and provide lessons from the experience of using a theory-based approach to evaluate a set of case studies. Our analysis demonstrates that the projects employing more transdisciplinary principles in their design and implementation make more diverse contributions and have a greater breadth of influence.

1. Introduction

The research presented in this paper responds to two interactive trends. First is the high and growing demand for more and more effective research to help address what Wiek et al. (2011b) describe as "sustainability problem constellations" (p.3). Despite mainstream political interest and widespread research on sustainability, humanity's progress "remains on largely unsustainable development trajectories" (Abson et al., 2017, p.30). There is increasing political pressure on research funding agencies to account for the scientific and social impacts of the research that they support, which in turn requires researchers to provide evidence of these impacts (Reale et al., 2017). This demand manifests in calls for improved research evaluation to demonstrate the value of research, and attract and be accountable for investments in research (Wilsdon et al., 2015). Research evaluation is also needed to support learning to improve research design and implementation (Boaz et al., 2009; Mitchell et al., 2015; Pohl, 2011).

The second related trend is a marked evolution in the way research is conducted to address pervasive challenges, with deliberate efforts to cross disciplinary and professional boundaries to engage stakeholders and intended research users as a way to deal with complexity and increase the relevance, uptake, use, and impacts of research. This evolution has been most prevalent in research fields dealing with complex social problems (e.g., public health, sustainability, and development) that are multi-faceted and therefore require integrated perspectives from which to develop solutions (Gehlert et al., 2014; Wiek et al., 2011a). Such approaches have been variously described as Mode 2 research (Funtowicz and Ravetz, 1993; Gibbons et al., 1994), transdisciplinary research (TDR) (Bergmann et al., 2012; Carew and Wickson, 2010; Jahn et al., 2012; Klein, 2006; Lang et al., 2012; Pohl and Hirsch Hadorn, 2007; Pohl et al., 2010; Walter et al., 2007; Wolf et al., 2013), and sustainability science (Brandt et al., 2013; Clark and Dickson, 2003; Heinrichs et al., 2016; Kates, 2017; Kates et al., 2001; Kauffman and Arico, 2014; Komiyama and Takeuchi, 2006; Roux et al., 2017), among others.

Despite the increase in theoretical discussion and actual application of TDR, there is still limited conceptual formulation of the kinds and scale of effects that can be realized, empirical evidence of the effectiveness of these approaches, and learning from experience to help improve the design and implementation of TDR (Jahn et al., 2012; Walter et al., 2007; Wiek et al., 2014). Some authors have discussed benefits of TDR approaches, including improved networks, capacity, and usable products (Flint et al., 2019; Ozanne et al., 2017; Walter et al., 2007; Wiek et al., 2014), though there have been few empirical studies to understand these effects (Wiek et al., 2014).

* Corresponding author at: Sustainability Research Effectiveness Program, College of Interdisciplinary Studies, Royal Roads University, Victoria V9B 5Y2, Canada.
E-mail address: brian.belcher@royalroads.ca (B.M. Belcher).

https://doi.org/10.1016/j.envsci.2019.08.013
Received 15 April 2019; Received in revised form 28 June 2019; Accepted 26 August 2019
Available online 10 September 2019
1462-9011/ © 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

Both trends have converged strongly in the international research-for-development field. Research funders (i.e., bilateral and multilateral development agencies and private foundations) demand more impact and more evidence of impact (Renkow and Byerlee, 2010; Savedoff et al., 2005; Stern et al., 2012) from the research organizations they support. Dedicated researchers, research managers, and research organizations have responded by adapting their approaches to be more engaged and deliberate to contribute to societal transformation (Court and Maxwell, 2005; Gaziulusoy et al., 2016).

For example, CGIAR[1], a large international agriculture and natural resources research consortium, has undertaken an organizational reform process which increased emphasis on social, economic, and environmental outcomes and impacts on top of the long-established commitment to produce high quality science. This shift was made explicit as a commitment to shared responsibility for impacts (ISPC, 2015), defined as reduced poverty, improved food and nutrition security for health, and improved natural resources and ecosystem services (CGIAR, 2016). One major aspect of the reform process was the creation of Consortium Research Programs (CRP) which intentionally facilitated broader and deeper partnerships, including with a range of policy and development actors (i.e., international conservation and development organizations, NGOs, policy-makers, etc.), as well as with other researchers and research organizations. CRPs were also required to develop theories of change (ToC) at the program-scale to model and plan the links from research through outcomes to intended impacts. This has resulted in a range of new research projects that incorporate many elements of TDR, with high variation in context, design, and implementation, creating an excellent opportunity to learn from individual TDR projects and conduct comparative analyses across projects.

This paper compares and contrasts five such research-for-development projects. The cases represent a range of research approaches, social and policy contexts, and outcomes. This analysis aims to learn whether and how TDR approaches in each case contributed to scientific and social outcomes, and draw more general lessons about TDR design, implementation, outcomes, and evaluation. The research question is: Does research that employs TDR principles contribute to (more) effective scientific and social outcomes?

The analysis takes advantage of a unique empirical dataset to assess both TDR design and implementation and scientific and social outcomes. The research builds on previous work to develop a framework to conceptualize and characterize TDR qualities (Belcher et al., 2016; Ramirez and Belcher, 2018) and develop and test a theory-based method to empirically assess the outcomes and impacts of TDR projects (Belcher et al., 2017; Halimanjaya et al., 2018). The framework (discussed in more detail below) adapts typical disciplinary research quality characteristics and adds a range of additional characteristics to reflect current theory about how TDR research should be done, with attention to problem orientation, stakeholder engagement, epistemological integration, and knowledge co-generation. This allows us to: characterize each project by the degree to which it conformed with principles of TDR design and implementation; assess the results of each project in terms of social and scientific outcomes; compare projects in terms of relationships between TDR design/implementation and outcomes; and provide lessons from the experience of using a theory-based evaluation approach.

We begin by introducing and defining key concepts and terms, and then summarize the research projects examined. We briefly present the methods used for each case study and the comparative analysis. Results are presented as: an overview of each project's design and implementation; analysis of the TDR characteristics of the five projects; and description and analysis of the social and scientific contributions made by the projects. The conclusions summarize the lessons learned from assessing, comparing, and contrasting the five cases in terms of how they were designed and implemented, what they achieved, and how they achieved change.

2. Key terms and concepts

The evaluation literature is rife with ambiguous and poorly-defined terms and concepts, which hinders communication, understanding, and progress (Belcher and Palenberg, 2018). In this paper, we take a systems perspective, assuming that any project operates within a complex system with other interacting actors and processes. We are interested in whether and how projects contribute to processes of change, but we have no expectation that any change will be fully attributable to one project. We conceptualize the change process within three spheres of declining control and influence (Earl et al., 2001). The project activities and outputs are (mostly) under the control of the project, in a sphere of control. Project outputs, which are the products, goods, and services of the research project, can influence other actors and their actions in a sphere of influence. In turn, those actors and actions can influence other actors and actions to contribute to positive social, economic, and/or environmental changes in a sphere of interest (Belcher et al., 2018). In this model, we define 'outcomes' as changes in knowledge, attitudes, skills, and/or relationships of actors in the system (individuals or organizations) that manifest as changes in behaviour as a result of the research process and its outputs. We define 'impacts' as changes in flow (e.g., higher annual income; increased water discharge from a watershed) or state (e.g., socio-economic status; water quality in a reservoir) resulting in whole or in part from a chain of events to which the research has contributed. These kinds of changes in poverty alleviation, food and nutrition security, and environmental conditions are the impacts that CGIAR research funders are seeking. These definitions of 'outcome' and 'impact' are more restricted and precise than are often used in evaluation, TDR literature, and everyday language. We use the term 'results' as a general term for any effect of an activity (i.e., output, outcome, or impact).

3. Case studies

The CGIAR is a large global research network focused on agriculture and renewable natural resources management. There are 15 research centers and 12 cross-center CRPs, working with a total annual budget of USD850 million (CGIAR, 2017). The Forests, Trees and Agroforestry (FTA) CRP does research to enhance the role of forests, trees, and agroforestry in sustainable development, food security, and address climate change (FTA, 2019). FTA research aims to generate new knowledge and contribute to improved policy and practice for enhanced management and use of forests and landscapes and improved livelihoods.

The research projects included in this analysis were done under the auspices of FTA. They were done by CGIAR scientists in collaboration with a range of other research organizations, governments, and NGOs. None of the authors of this article were involved as researchers in any of the original FTA research projects. The completed research projects were selected for evaluation by FTA's Monitoring, Evaluation, Learning and Impact Assessment (MELIA) team for their learning potential. There were indications that each project had some success in policy engagement. The outcome evaluations (i.e., the case studies) were conducted to critically assess whether and how the research contributed to policy processes and/or other change processes. Each author was involved in at least one of the independent case study evaluations (design, data collection, analysis, and reporting). These outcome evaluations were done in a participatory mode with one or more of the authors, plus some other researchers, none of whom were part of the original research project teams.

A brief description of each project follows. Additional detail is provided in Table 1 and Appendix 1 (see Table S1). References to published on-line summaries of the case studies are provided in the first column of Table 1. Some case studies have been published (Brazil Nut Project (BNP), Sustainable Wetlands Adaptation and Mitigation

[1] CGIAR is the full name.

Table 1
Overview of project details for each case study.

Case Study	Research Topic	Scale	Budget	Duration	Does the project build on previous research and/or relationships? (Y/N)	Intended Audiences
BNP (Rockwell et al., 2015)	Impact of timber extraction on Brazil nut production (Peru)	• Subnational • National	500,000 – 1 million USD	3 years (2012-2015)	Y • Diagnostic study of timber extraction in Brazil nut concessions • Pre-existing relationships with policy-makers and NGO (boundary organization) key to support work with policy-makers	• National government (policy-makers) • Practitioners • Brazil nut concessionaires • Academia (field assistants, forestry faculty)
F&H (CIFOR, 2019a)	Management of fire and haze (Indonesia)	• National	$500,000 – $1 million USD	2 years (2014-2016)	Y • Lead researcher had pre-existing personal relationships with policy-makers • CIFOR has worked on fire and haze research for 18 years	• Policy-makers • Private sector • Community members
GCS-FTR (CIFOR, 2019b)	Forest tenure reform and implementation (Peru)	• Subnational • National • Multi-country	$1million – $2 million USD	4 years (2014-2018)	Y • Previous research on tenure reforms in ten countries in Africa, Asia, and Latin America help identify key issues and research gaps • Pre-existing relationships with policy-makers, regional government, NGOs, and academia	• National government (policy-makers) • Regional government • Practitioners • NGOs • Indigenous communities (local, regional, and national organizations) • International community (donors)
SUCCESS (ICRAF, 2019)	Management of agroforestry concessions (Peru)	• Subnational • National	< $250 000 USD	18 months (2016-2017)	Y • Diagnostic study of smallholder farmers in the Amazon • Pre-existing relationships with governments and partners	• National government (policy-makers) • Regional government • NGOs • Informal smallholder communities • Partner academic and development institutions
SWAMP (CIFOR, 2019c)	Sustainable wetlands management for climate adaptation and mitigation (Indonesia)	• Multi-country	> $2 million USD	6 years (2009-2015)	Y • Precursor study: TWINCAM	• National policy-makers • International policy-makers

Program (SWAMP)), and some are forthcoming (Fire and Haze Indonesia (F&H), Global Comparative Study on Forest Tenure Reform-Peru (GCS-FTR), Support to the Development of Agroforestry Concessions in Peru (SUCCESS)). The fact that all cases are from two countries, Peru and Indonesia, is coincidental. In the Peruvian case studies, there is some overlap among stakeholders from the national and regional governments, and one NGO was engaged as a partner in both the BNP and GCS-FTR cases. Otherwise, there is little overlap between cases in either country.

3.1. Brazil Nut Project (BNP)

The research aimed to provide a scientific basis for national policy on timber harvesting in Brazil nut concessions in Peru. The research was intended to support multi-use forest management to benefit smallholders' livelihoods and forest sustainability. The researchers worked with smallholder Brazil nut concessionaires and applied a quasi-experimental design to estimate an allowable timber harvest that would not affect Brazil nut production. The project liaised with government authorities and NGOs to support the integration of findings into policy. Independent outcome evaluations for this case study have been published (see Ramirez and Belcher, 2018, 2019).

3.2. Fire and Haze Indonesia (F&H)

The project aimed to investigate the political economy of peatland fire and haze in three provinces in Indonesia, to provide information, analysis, and tools to government, private sector, civil society organizations, and local community actors to identify and implement fire prevention strategies. There were three research streams: political assessment of fire; spatial mapping and modeling of fire events; and policy analysis. The project engaged with high-level government officials and directly participated in national and subnational policy processes.

3.3. Global Comparative Study on Forest Tenure Reform-Peru (GCS-FTR)

The research aimed to advance understanding of factors affecting the implementation of land tenure reform in Peru to support policy implementation. The project worked in two regions to convene multi-level government representatives, NGOs, community members, and indigenous federations in multi-stakeholder fora to support capacity-building, knowledge co-generation, and knowledge translation.

3.4. Support to the Development of Agroforestry Concessions in Peru (SUCCESS)

The research aimed to contribute to more coordinated, effective, and informed governance of agroforestry concessions (AFC) in Peru to support evidence-based and contextually appropriate implementation. The research approach integrated elements of technical (methods testing, mapping, PGIS, carbon emissions estimations, and policy analysis) and social research (anthropological assessment, interviews, surveys, workshops, and focus groups).

3.5. Sustainable Wetlands Adaptation and Mitigation Program (SWAMP)

SWAMP was a multi-national research program aimed to inform national and international climate policy and practice by developing tools and methods to quantify greenhouse gas (GHG) emissions, carbon stocks, and flux in tropical wetlands due to land use, land-use change, and forestry (LULUCF). SWAMP focused stakeholder engagement in Indonesia, but also engaged in international policy discourse. The technical research approach produced findings relevant for climate mitigation and adaptation strategies, which were presented through researcher engagement in relevant policy processes. This outcome evaluation has been published (Halimanjaya et al., 2018).

4. Methods

We used a comparative case study methodology. Case studies enable engagement with the complexity inherent in TDR and make it possible to retrieve reliable evidence from the intended users of the research (i.e., policy-makers and other stakeholders) about the degree to which the research project influenced their work or contributed to broader changes in the context (Boaz et al., 2009; Bornmann, 2013; ESRC, 2009).

4.1. Characterizing TDR design and implementation of each case

Belcher et al.'s (2016) TDR Quality Assessment Framework (QAF) provides criteria and an assessment rubric organized within the principles of relevance, credibility, legitimacy, and effectiveness. The criteria, identified from a systematic review of the literature on inter- and transdisciplinary research quality assessment, cover all aspects of TDR contextualization, problem definition, research design, collaboration, engagement, and communication. The QAF comprehensively reflects each of these elements of research design and implementation in one or more criteria. Therefore, it provides a checklist of TDR characteristics and a scale to assess the degree to which each characteristic is present. Each criterion has an explicit definition and rubric statement against which each project was scored (see Appendix 2: Table S2). The method specifies that assessment of each criterion should be done considering the project purpose. A score of 2 is assigned if a project fully satisfies the criterion, 1 if the criterion is partially satisfied, and 0 if the criterion is not at all satisfied. Projects were scored independently by the authors and average scores were recorded. The scores for individual criteria and as a set provide a TDR profile, indicating TDR characteristics that were strong, present but incomplete, or absent in each project.

4.2. Defining the theory of change for each case

A detailed project ToC serves as the analytical framework for the outcome evaluation for each project. A ToC models the change process, providing a description and explanation of how and why the project was expected to contribute to a process of change as a set of testable hypotheses. The ToC specifies the main actors involved in the change process and identifies their actions as a sequence of steps. Each research project had an implicit ToC but none had a fully articulated and documented ToC (GCS-FTR had developed a preliminary ToC at project inception). A sub-set of the authors worked with each of the research teams and partners after (or near) project completion in a two-day workshop to document and specify the project ToC. A draft version of the ToC was agreed during the workshop and subsequently refined in consultation with the research project team. Each ToC was documented as a flow diagram containing a set of activities, outputs, outcomes, and impacts (Fig. 1.) organized by impact pathways, and in narrative form. Impact pathways are defined by the primary actors or actions to be influenced. For example, a policy pathway represents the constellation of actors and actions that would, in theory, contribute to a change in policy.

4.3. Data collection

The QAF and ToCs served as guides to identify what data were needed in each case to assess project design and implementation and test the ToC. Data were needed to document and evidence project activities, partnerships, outputs, and the range of hypothetical outcomes. We used mixed methods, including document review, key informant interviews, focus group discussions, and surveys to collect the data needed to understand project contexts, characterize project design and implementation, assess outcomes, and understand the processes and mechanisms that contributed to outcomes. The interview guides were designed to explore unexpected outcomes and alternative explanations to capture emergent phenomena characteristic of complex systems. Data were summarized and organized in an evidence table, which logs

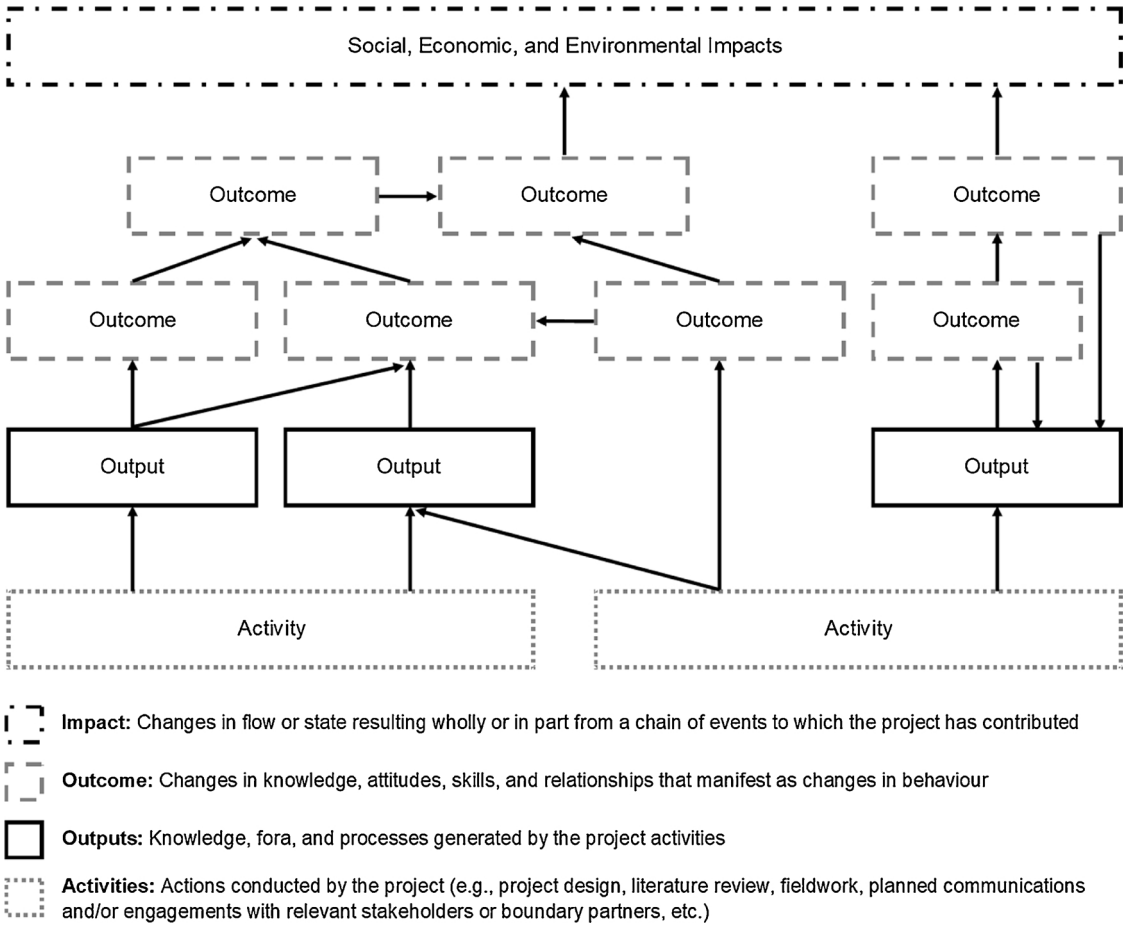

Fig. 1. Sample theory of change flow diagram demonstrating logical flow of activities, outputs, outcomes, and impacts.

each ToC result, evidence to confirm or disconfirm each result, and data sources for all evidence presented.

4.4. Assessing the scientific and social outcomes of each case

We applied the Outcome Evaluation method developed by Belcher et al. (2017) and Ramirez and Belcher (2018), building on theory-based evaluation approaches (Donovan and Hanney, 2011; Mayne et al., 2013; Mayne, 2012; Mayne and Stern, 2013; ODI, 2012; White and Phillips, 2012). The approach assesses whether and how a project contributed to the achievement of outcomes, using a project's ToC as the analytical framework and empirically testing it (Belcher et al., 2017; Ramirez and Belcher, 2018). Each step in the ToC was assessed to determine if it was realized and if there is evidence that the project contributed to its realization. We also explored complementary and alternative explanations for results in lieu of a true counterfactual (Mayne, 2012).

4.5. Comparative analysis

Each case is an example of a practical effort to apply (some) transdisciplinary principles in a research-for-development context, with a clear problem-orientation and engagement of actors beyond the project boundary. All cases were identified as successful projects in terms of making a contribution to changes in policy and/or practice. The individual case analyses identified the strengths and weaknesses of each project against current TDR standards and assessed the degree to which the research projects contributed to outcomes. The comparative analysis builds on the individual case studies, and their collective variability, to investigate the relationship between the design and implementation of TDR and outcomes within and across the five cases. This qualitative analysis explores relationships between the character and quality of TDR design and implementation and the kinds of outcome and impact contributions. We compared the ToCs and QAF scores, looking for common elements and idiosyncratic differences across cases. We also considered elements of design and implementation in conjunction with outcomes, looking for patterns across this set of cases.

5. Results and discussion

5.1. TDR design and implementation

A first notable observation is that each case study used multiple impact pathways. That is, each project aimed to contribute through knowledge creation, but also used other mechanisms, such as capacity-building, coalition-building, and agenda-setting to inform, support, and influence a range of actors in the system. Each case aimed to contribute to several high-level outcomes. Fig. 2 shows the ToC diagram for SWAMP. All five ToCs are available as supplementary material (see Appendix 3: Figures S3-S7). This multi-pathway model contrasts the typical linear model of science in which knowledge is created, disseminated, taken up, and used (Pielke, 2007). Rather, these cases are consistent with a stakeholder model of science in which users have some role in knowledge production, and where considerations of how science is used in decision-making are important (Pielke, 2007; Sarewitz, 1996; Stokes, 1997).

5.2. QAF characterization

Fig. 3 illustrates the scores by QAF principle and criterion (see Appendix 4: Tables S8-S12 for the evaluators' scores and justifications). Each case included many TDR characteristics in different combinations.

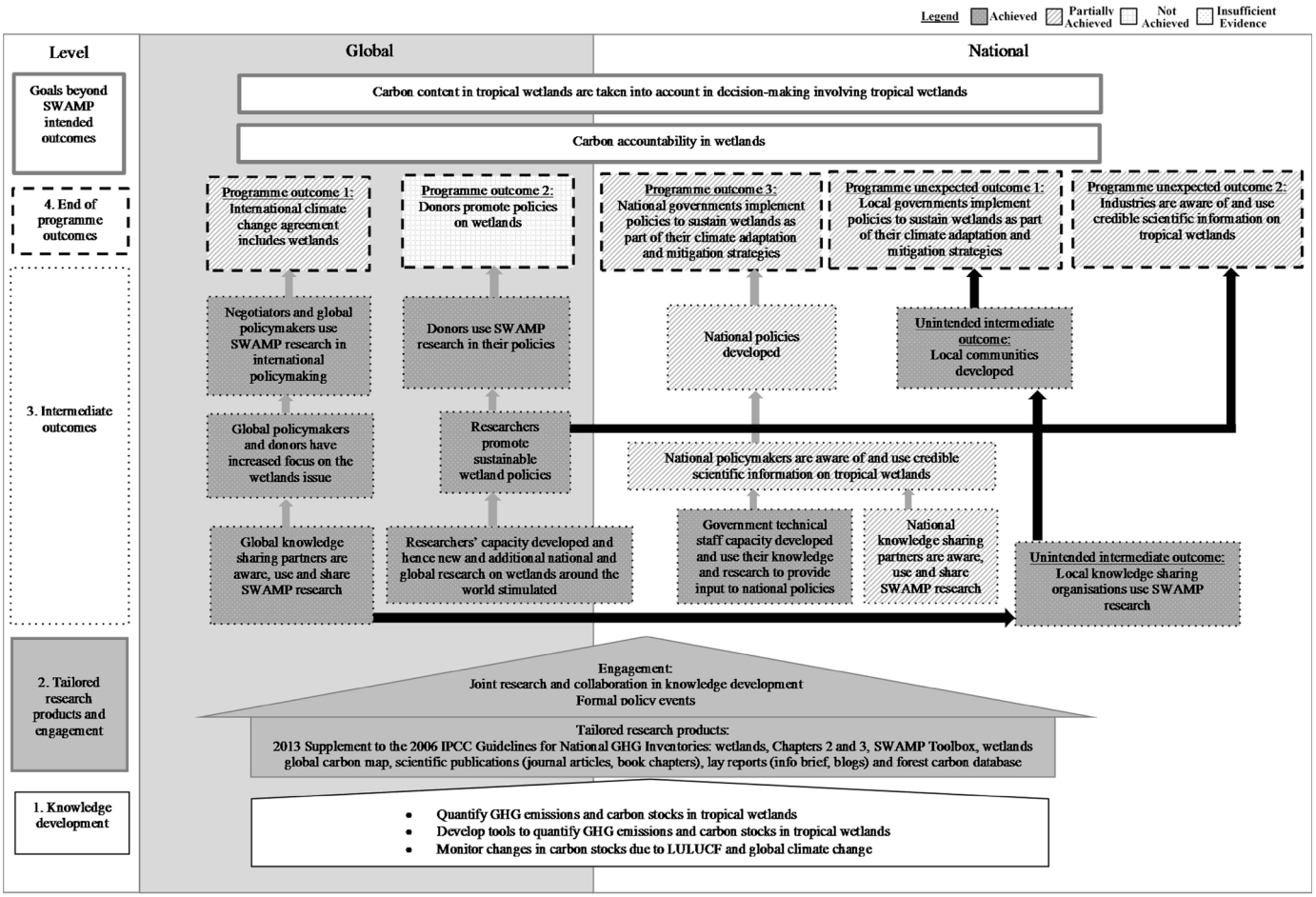

Fig. 2. SWAMP theory of change, with outcomes coded to reflect the extent of outcome achievement (adapted from Halimanjaya et al., 2018).

The following discussion highlights some examples and summarizes and explains key findings by principle.

5.2.1. Relevance

Fig. 3a illustrates that all projects demonstrated understanding of the context and were designed to operate and contribute in that context. This reflects that the researchers involved in each project had previous experience, knowledge, and relationships. It also reflects the deliberate intent to do research that would be used in a research-for-development context. All projects scored lower on the criterion 'Effective communication'. This is a challenging criterion to satisfy as it is currently defined because it considers communication during all research stages with all relevant actors; omission of one necessary actor group at any stage results in a score of 1. Likewise, all projects scored low on 'Explicit theory of change'. GCS-FTR had done work to develop and document its ToC at inception, but emphasis was at the global project-scale and specificity was needed at the national-scale (which was the focus of the evaluation). Still, the lead researchers found the ToC conceptualization useful to guide the project and support adaptive management. All other projects had implicit ToCs, but the ToC workshops revealed differences in understanding among project teams and partners when made explicit.

SUCCESS scored high on many 'relevance' criteria. The research team had previous research experience in the study regions, and leveraged pre-existing relationships with regional government actors. One team member had high social capital in one region's study communities. The technical analysis of AFC policy was relevant to the change the project aimed to influence, and government informants perceived the research findings to be highly relevant to their work.

BNP responded well to a policy window for management guideline revisions. However, while the ToC anticipated the research would benefit Brazil nut concessionaires, our interviews with concessionaires and other informants indicated that they did not feel that the research was relevant to them. They suggested that the research was too technical and did not ask what they thought were the right questions. This criticism is not fair or correct if the project is evaluated as a biophysical research project meant to inform forest policy; however, if it had been developed deliberately as a TDR project, these intended users would have had the opportunity to help define the research question and the project might have taken a different course.

5.2.2. Credibility

There was greater variability across projects under the principle of 'credibility' (Fig. 3b). The highest scoring criteria across projects were 'Adequate competencies', 'Clear problem definition', and 'Objectives stated and met'. This is partly owing to the fact that the projects were led by researchers working for well-resourced international research organizations whose funders require explicit documentation of problem definitions and research objectives.

BNP and SWAMP had the greatest focus on biophysical research and employed approaches that were most strongly rooted in disciplinary scientific traditions. As such, they scored well on criteria that relate closely to disciplinary scientific criteria. The technical research question and the multi-country scale of SWAMP required well-developed replicable methods with clear documentation. The methods were transferable and scalable by design. SWAMP was well-resourced in funding and scientific capacity to meet project objectives and adapt as needed during the course of the project. However, SWAMP scored

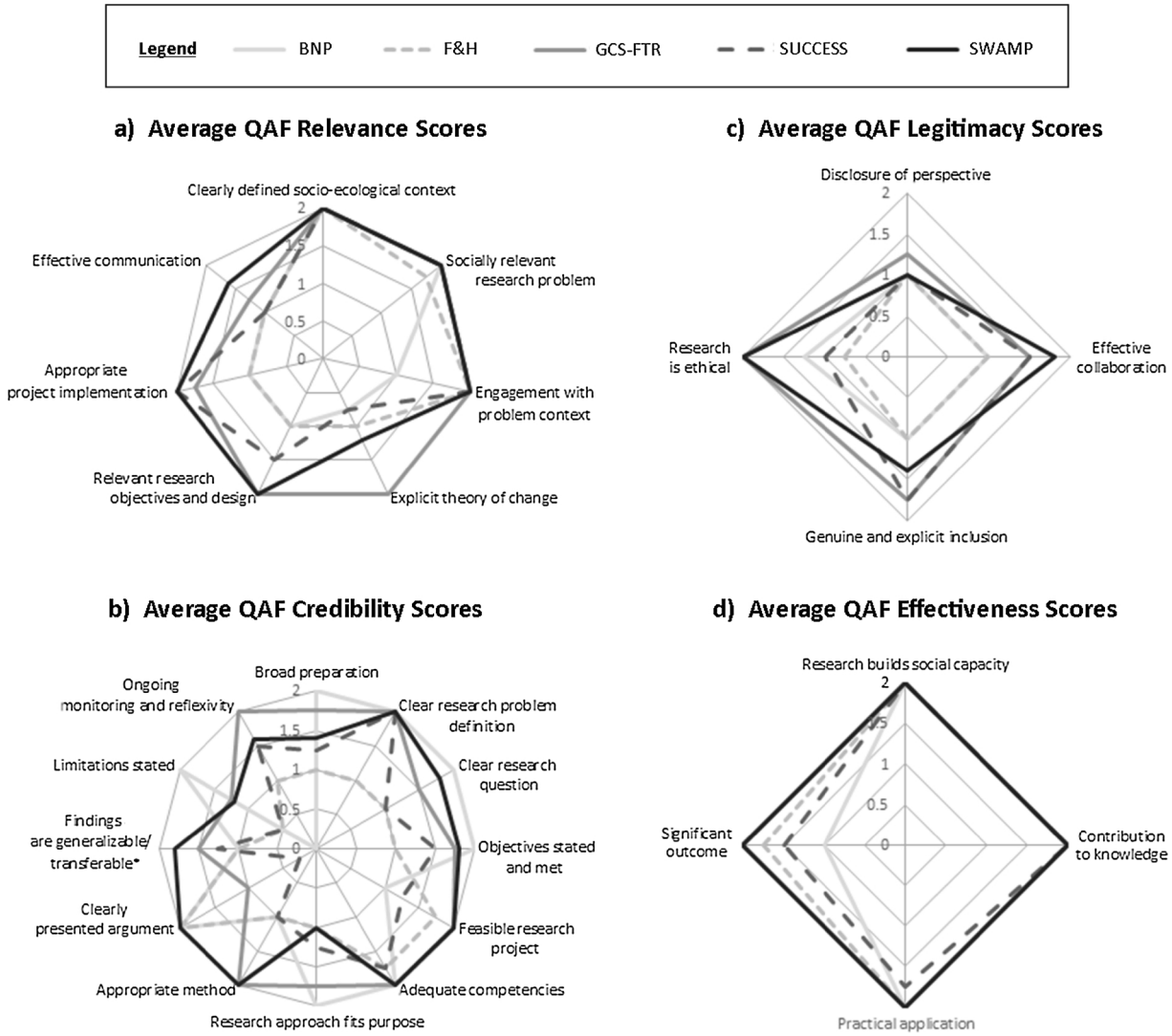

Fig. 3. QAF scoring diagrams of (a) relevance, (b) credibility, (c) legitimacy, and (d) effectiveness.

lower on some criteria indicative of TDR process characteristics. As SWAMP was primarily driven by a scientific rationale, less attention was given to consultation and engagement for 'Broad preparation', and the criterion 'Research approach fits purpose' was not fully satisfied because the project claimed to be interdisciplinary but did not explicitly explain methodological and/or epistemological integration.

GCS-FTR displayed good use of 'Ongoing monitoring and reflexivity', which the researchers used to adapt the project as new opportunities emerged, while other projects exhibited less consistent scores under this criterion.

It was surprising that not all projects included a 'Clearly stated research question'. In some cases, the research question was implicit and could be gleaned from objectives. Discussions with project teams revealed that this is partly an artifact of donor funding proposal formats, which in a development context may not require an explicit research question. We also noted low scores for 'Limitations stated'. Though donors may not require these elements in the projects they fund, we recommend their inclusion for transparency.

5.2.3. Legitimacy

The scores on the four criteria under 'legitimacy' also showed high variation (Fig. 3c). GCS-FTR and SUCCESS scored well on 'Research is ethical', 'Genuine and explicit inclusion', and 'Effective collaboration'. This reflects that both projects invested heavily in stakeholder engagement and collaborated extensively with partner NGOs, indigenous organizations, research institutions, and/or international organizations to support their work. In GCS-FTR, the lead researchers utilized soft skills and deliberately created opportunities to include women and indigenous voices in research processes and outputs. The project facilitated knowledge co-production and validation of the research findings through its collaborative approach. Intra-team collaboration was also important for both cases, as roles, partnerships, and decision-making structures were mutually agreed upon and made explicit.

SWAMP was less engaged than these other two projects, but it was appropriate to the project purpose and therefore still scored relatively high. BNP and F&H scored lower on these criteria. All projects scored low on 'Disclosure of perspective', suggesting a need for researchers to be more aware or more forthcoming in declaring their interests and perspectives. This may be another artifact of donor proposal requirements or may reflect a difference between the research-for-development context and academia, where there is increasing emphasis on the importance of reflection on positionality, potential bias, and bias implications in modern methodology courses.

F&H was an outlier that demonstrates the importance of having both effective internal and external collaboration for a project. External collaboration with national-level policy-makers was strong because of one researcher's pre-existing personal and professional relationships. This provided the project access to key decision-makers and processes

Table 2
Types of research contributions made by each case study.

		BNP	F&H	GCS-FTR	SUCCESS	SWAMP
Knowledge contributions	Develop, test and improve theory and methods			✓	✓	✓
	Provide theoretical and/or empirical analysis and possible solutions	✓	✓	✓	✓	✓
	Provide evidence-based recommendations and guidance for improved policy and practice	✓	✓	✓	✓	✓
Social process contributions	Build social and/or scientific capacity	✓		✓	✓	✓
	Encourage and influence public discourse		✓	✓	✓	✓
	Facilitate negotiated solutions		✓	✓	✓	✓
	Network development		✓	✓	✓	✓
	Influence research agendas			✓	✓	✓
	Influence policy and practice	✓	✓	✓	✓	✓

based primarily on the reputation and trust of the researcher, and, by extension, their research organization. However, internal collaboration and communications were weaker in comparison; the three research streams functioned as separate sub-projects, and the opportunity to integrate the streams and leverage access to policy processes for the full range of research outputs was missed.

5.2.4. Effectiveness

The scores on the four 'effectiveness' criteria are displayed in Fig. 3d. These QAF criteria were originally designed to be applied based on review of project proposal and final reports. In this study, the outcome evaluations provide much richer detail on contributions to knowledge, building of social capital, practical application, and significant outcomes in each case, so we turn to that now (see Appendix 5: Tables S13-S17 for summaries of the outcome assessments). All projects made tangible contributions to change processes with sufficient evidence to make a reliable assessment.

Most outcomes achieved by BNP related to changed knowledge, awareness, and understanding, affecting stakeholder attitudes toward science in forest policy development. BNP supported capacity-building, though it was not central to the project. BNP results were considered and referenced in the national forest policy, though ultimately the timber harvest recommendations were not adopted.

Most F&H outcomes were achieved through changed understanding, awareness, and attitudes toward fire prevention. The project supported a shift in discourse and policy from a predominant emphasis on fire suppression toward fire prevention, stimulating new action. The project supported capacity-building, though it was not central to the project. However, most of the outcomes in the policy realm resulted from just one component of the project. This is indicative that there was scope for more effective collaboration among the project team (including more explicit interdisciplinary integration) to augment outcome achievements. Many interview informants expressed interest in but were not aware of findings from other components of the project.

Outcomes achieved by GCS-FTR related to changed understanding and awareness; access, uptake, and sharing of knowledge and tools; enhanced capacities; and strengthened relationships and trust. It is too early to assess outcomes relating to support, advocacy (attitudes), and coordination (changed relationships) for tenure reform implementation.

SUCCESS achieved outcomes relating to changed understanding, recognition, and support for new methods trialed by the project and application of AFCs for conservation and livelihood objectives. It is too early to assess outcomes relating to changed policy.

Outcomes achieved by SWAMP related to changed awareness and recognition of the research findings, in addition to changed capacities. Changed knowledge in terms of quantifying below-ground carbon and GHG emissions from wetlands highlighted the role of wetlands in mitigating climate change and subsequent omission of wetlands in international policy. This knowledge, shared through conventional academic means, but also, importantly, through direct engagement in national and international policy processes by project scientists, supported uptake and use of the research. This is indicative of an effective contribution to problem-framing and solution development as a part of the policy process (Hanberger, 2001).

5.3. Scientific and social contributions of projects

As noted above, the ToCs for each project are complex, with multiple interacting impact pathways. The ToC diagrams are colour-coded to indicate which outcomes were achieved, partially achieved, or not achieved (Fig. 2; see Appendix 6: Figures S18-S22), and which outcomes had insufficient evidence to judge. Table 2 indicates the kinds of contributions each project made.

All projects made both knowledge and social process contributions. Projects that employed more TDR characteristics achieved more diverse contributions. Contrasting BNP and GCS-FTR demonstrates this finding clearly. BNP, which would be characterized as the least transdisciplinary, made less diverse research contributions overall and markedly less social process contributions. GCS-FTR, which would be characterized as the most transdisciplinary, made many kinds of research contributions and had a greater range of knowledge and social process contributions.

This may be explained by the range of mechanisms applied in transdisciplinarity research. Table 3 indicates the kinds of mechanisms leveraged by the fives projects to achieve outcomes.

BNP, F&H, SUCCESS, and SWAMP technical scientific outputs likely could have been produced without a TDR approach. However, they went beyond purely scientific approaches in order to improve their contributions to policy and practice outcomes. Notably, SWAMP researchers engaged directly in Indonesian and international policy processes, suggesting that employing some TDR elements, even in a disciplinary research project, can help translate and mobilize knowledge. BNP took advantage of a policy window, developed a cost-effective method, and produced data and analysis that filled a knowledge gap and could be used directly in policy development. The team also invested time and effort in reaching out to stakeholders, especially building relationships with key actors in the policy process, which helped ensure that research results were utilized. F&H had a mixed record. The project engaged with a high-profile political issue. One researcher in particular had a high level of interaction and influence on national policy processes on the basis of their and their organization's reputation. Moreover, the project invested in processes that helped build a coalition of interest, but was less effective at leveraging that potential to make use of the full range of research results.

SUCCESS and GCS-FTR were more comprehensive in applying TDR principles. GCS-FTR went furthest in its engagement of stakeholders to guide research direction and facilitate knowledge co-generation. Genuine engagement led to high accountability and legitimacy of the research process. The outcome evaluation found that the research team was open to feedback and input from participants (mainly from NGOs, academia, and government), which increased research legitimacy. Use of participatory tools and the meaningful engagement of participants from native communities and NGOs contributed to capacity-strengthening and coalition-building around tenure reform. SUCCESS also employed a high level of engagement to build capacity and coalitions of

Table 3
Impact pathways and mechanisms leveraged in each case study.

	International policy	National and sub-national policy	NGO policy and/or practice	Private sector policy and/or practice	Individual resource manager practice	Academic
BNP		▲●✦	■		■	▲☽➤✚
F&H		▲■●✦◈	■	◈■	✚■	
GCS-FTR	❖➤✚■✦	▲☽❖◈✚■●	▲❖✚■		❖✚	▲☽➤✚
SUCCESS		▲◈✚■✦	☽◈■		✚	
SWAMP	▲☽◈■●✦	▲☽✚■●✦	✚■	■		▲☽➤✚✦

Legend

- ▲ scientific knowledge increased/knowledge gap filled
- ☽ methods developed and/or refined
- ❖ knowledge co-produced
- ➤ research agenda influenced
- ◈ alignment of research with parallel issues/initiatives
- ✚ capacity of actors in system improved
- ■ coalitions created or strengthened
- ● policy window opportunity realized
- ✦ reputation leveraged and/or enhanced

interest. The project also successfully aligned itself with another high-profile policy process by demonstrating how AFCs could help realize national commitments to climate change objectives.

With respect to the kind and depth of engagement, projects mostly aimed to inform (e.g., BNP) and consult (e.g., SWAMP, F&H, SUCCESS) with their target audiences, with some engagement throughout the project. GCS-FTR actively sought and encouraged a high level of participation through partnerships. Table 3 illustrates that GCS-FTR utilized the greatest breadth of mechanisms, and therefore the greatest breadth of influence, which correlates to its breadth and depth of engagement.

It appears that when a project employs more transdisciplinary elements, it has more mechanisms at its disposal to make a broader range of knowledge and social process contributions. With more diverse contributions and mechanisms to leverage, a project has the potential to have greater influence across more impact pathways. Other projects looking to expand their research contributions, influence, and overall effectiveness could benefit from the qualities enabled by TDR in the design and implementation of their research.

6. Limitations

The research projects were not randomly selected, so the analysis of overall effectiveness cannot necessarily be extrapolated to the larger universe of FTA or CGIAR research projects. While the cases represent a wide range of topics and research approaches, they were mostly focused in two countries, Peru and Indonesia; different geographic, social, and political systems will require different approaches. Lessons about outcome achievements are context-specific, but lessons regarding research design and implementation can be expected to have general application as long as researchers consider the geographic context in which a project is situated (i.e., account for aspects of relevance).

The ToCs were developed retrospectively, so it was not possible to accurately document original intentions, expectations, and assumptions. The ToC used to analyze each project was, at least in part, a reflection of what had actually happened. While this would be considered a weakness in a summative evaluation of the original project design, it is appropriate to evaluate a TDR project, which is expected to adapt to new knowledge and emergent conditions, by its most up-to-date ToC.

Higher-level outcomes inevitably take time to be realized and in some cases evaluation data were collected near the project completion, with some outputs still to be produced (e.g., GCS-FTR, SUCCESS). In that sense, the outcome evaluations are a snapshot of a continual process.

The QAF was originally designed to be used for evaluating project documents (project proposals or final project reports and/or scientific publications). Therefore, some criteria do not make full use of the rich data available from interviews and wider document review. We modified some criteria (see footnote 1 in Appendix 2). QAF scores should not be interpreted as measures of excellence, but only as indications of the presence and strength of TDR characteristics. This test of the QAF will help to revise and improve it for broader application, as a checklist and guide for project planning and/or monitoring and as a framework for proposal (*ex ante*) or project (*ex post*) assessment.

7. Conclusion

One leverage point in the sustainability research agenda is "rethinking how knowledge is created and used in pursuit of sustainability" (Abson et al., 2017, p.30). We have described, compared, and contrasted five international research-for-development projects that each employed elements of transdisciplinarity. The particular combination of TDR elements was characterized by assessing the degree to which each project conformed to a set of TDR criteria, as organized in the QAF. The kinds of contributions intended by each project and hypotheses about the processes and mechanisms by which those contributions were to be realized were documented as detailed project ToCs. The ToCs were tested empirically to determine whether or not intended outputs and outcomes were realized and assess whether and how each project contributed. We wanted to know whether research that employs TDR principles contribute to (more) effective scientific and social outcomes. Based on this set of cases, the answer is yes. Our results demonstrate that projects incorporating TDR elements can achieve both scientific and social outcomes by expanding their breadth of influence. This is likely because each project was able to take advantage of a range of mechanisms to increase effectiveness. This evidence contrasts with the linear model of research uptake and use.

None of the projects set out deliberately to do TDR. Rather, each had a deliberate focus on moving beyond knowledge production to influence policy and practice. To do that, they successfully employed a variety of strategies that crossed disciplinary bounds and engaged a range of partners and stakeholders at different levels.

All projects started from a problem focus (as opposed to an abstract theoretical focus). That is, the research problem and research objectives were designed to help solve a social or environmental problem. In most cases, the research teams were already well-integrated in the problem context whether it was at a local scale (e.g., BNP, SUCCESS) or an international policy scale (e.g., SWAMP), and they invested in ensuring the relevance of their work.

The participation of various system actors contributed to the projects' relevance and strongly contributed to the uptake and use of the research. This is a fundamental premise of TDR (Boaz et al., 2009;

Burkhardt-Holm and Zehnder, 2018; Gaziulusoy et al., 2016; Jahn et al., 2012; Lang et al., 2012). Each project used participation and stakeholder engagement to some degree. Projects that invested most in developing and facilitating participation (e.g., GCS-FTR, SUCCESS) were the most successful in generating social learning and coalition-building. Projects that employed the most traditional scientific models (e.g., BNP, SWAMP) still invested in outreach and engagement, and this contributed to the realization of outcomes.

In several of the projects, researchers had opportunities to get involved in and influence policy processes, partly because of their reputations and/or the reputations of their organizations. That is, they gained standing and were able to contribute knowledge and ideas that were not necessarily derived directly from the research project.

All projects had multiple impact pathways, with deliberate attention to multi-scale government policy processes, but also contributed to policy and practice in private and NGO sectors as well as scientific progress. Encouraging and facilitating participation and engagement opened opportunities to influence change in a variety of ways, through capacity-building, coalition-building, and agenda-setting, which all helped achieve core project objectives and supported complementary outcomes. We feel that this is an important observation. It strongly contrasts with the classic linear model of science, reinforces the need for TDR approaches, and calls into question traditional disciplinary-oriented research evaluation approaches.

There was an indication of trade-offs between some desired elements of TDR. This may happen simply owing to practical considerations. For example, the high investment in public outreach in F&H may have been done at the cost of reduced collaboration and integration within the project. Any project will have decisions to make about how to allocate resources. There are also inherent contradictions between elements; for instance, producing timely research findings that align with decision-making cycles may shorten time for data collection and analysis, affecting aspects of research design and credibility (see Cash et al., 2002; Sarkki et al., 2014). The current analysis cannot delve into trade-offs, but we recommend that researchers give due consideration to the outcomes they hope to contribute to (and not only the outputs they aim to produce) and allocate resources accordingly.

This study also served as a test of methods used to evaluate change-oriented research projects in complex systems. The ToC workshops were universally appreciated by the participating research teams. The process of articulating and documenting the ToCs helped develop common understanding, even though it was done retrospectively. ToC also facilitates critical thinking, integration, collective visioning, transparency, and accountability (Belcher et al., 2019). FTA has found the approach useful enough to make it policy that future projects will develop ToCs as part of their design. The ToCs served as useful analytical frameworks to identify intended outcomes and empirically test whether they were achieved. The tool is sufficiently flexible to be applied at the programme-level, and has been trialed by the authors in a diverse range of programming including non-profit, higher education, and research.

The QAF was helpful to identify and assess the TDR characteristics of each project. This exercise demonstrated that the QAF can be applied to research projects that fall anywhere along the disciplinary-transdisciplinary spectrum. As a multi-functional tool, the QAF can be used for project planning, adaptive management, and evaluation. The QAF was originally intended to be applied to evaluate project proposals and/or reports of completed projects. The much richer data set created by the outcome evaluations revealed shortcomings in some concepts, definitions, and scoring. This will be addressed in forthcoming work.

Our analysis demonstrates that projects can make more contributions to scientific and social outcomes – thereby expanding their breadth of influence – by employing transdisciplinary principles. In these cases, the social processes employed helped translate and broker knowledge outputs, and made substantial additional contributions through capacity-building as well as initiating and supporting discourse and relationship-building. Research intending to influence policy and practice change should consider integrating and reflecting on TDR characteristics more intentionally throughout the research process. The QAF and ToC tools can help facilitate this process. Further testing and comparison of TDR evaluation frameworks (e.g., Blackstock et al., 2007; Cook et al., 2018; Phillips et al., 2018; Ozanne et al., 2017; Polk, 2014; Talwar et al., 2011; Wiek et al., 2014; etc.) needs to be done in transdisciplinary contexts to learn more about how research contributes to change processes and support greater convergence in the TDR field.

Declaration of Competing Interest

None.

Acknowledgements

The Sustainability Research Effectiveness Program is supported by the Canada Research Chairs Program, Canada (950-229114) and the Canadian Social Sciences and Humanities Research Council, Canada (435-2019-0096). The case study evaluations were partially supported by FTA (all), the UK Department for International Development KNOWFOR Project (BNP, F&H, SWAMP), European Commission (FCS-FTR), IFAD (GCS-FTR), and GEF (GCS-FTR). Aidy Halimanjaya and Daniel Suryadarma contributed to the SWAMP outcome evaluation, and Bethany Davies provided input during ToC workshops. Stephanie Jones assisted with QAF scoring and manuscript editing. We are grateful to all participants in the case studies, including CIFOR and ICRAF researchers and the many interview respondents.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.envsci.2019.08.013.

References

Abson, D.J., Fischer, J., Leventon, J., Newig, J., Schomerus, T., Vilsmaier, U., von Wehrden, H., Abernethy, P., Ives, C.D., Jager, N.W., Lang, D.J., 2017. Leverage points for sustainability transformation. Ambio 46 (1), 30–39. https://doi.org/10.1007/s13280-016-0800-y.

Belcher, B., Claus, R., Davel, R., Jones, S., Ramirez, L., 2019. Theory of Change: A Practical Tool for Planning and Evaluating Change-oriented Research. Retrieved from. https://researcheffectiveness.ca/wp-content/uploads/sites/7/2019/08/Theory-of-Change-Toolkit.pdf.

Belcher, B.M., Rasmussen, K.E., Kemshaw, M.R., Zornes, D.A., 2016. Defining and assessing research quality in a transdisciplinary context. Res. Eval. 25 (1), 1–17. https://doi.org/10.1093/reseval/rvv025.

Belcher, B., Palenberg, M., 2018. Outcomes and impacts of development interventions: toward conceptual clarity. Am. J. Eval. 39 (4), 478–495. https://doi.org/10.1177/1098214018765698.

Belcher, B., Suryadarma, D., Halimanjaya, A., 2017. Evaluating policy-relevant research: lessons from a series of theory-based outcomes assessments. Palgrave Commun. 3, 1–16. https://doi.org/10.1057/palcomms.2017.17.

Bergmann, M., Jahn, T., Knobloch, T., Krohn, W., Pohl, C., Schramm, E., 2012. Methods for Transdisciplinary Research: A Primer for Practice. Campus, Frankfurt.

Blackstock, K.L., Kelly, G.J., Horsey, B.L., 2007. Developing and applying a framework to evaluate participatory research for sustainability. Ecol. Econ. 60, 726–742.

Boaz, A., Fitzpatrick, S., Shaw, B., 2009. Assessing the impact of research on policy: a literature review. Sci. Public Policy 36 (4), 255–270. https://doi.org/10.3152/030234209X436545.

Bornmann, L., 2013. What is societal impact of research and how can it be assessed? A literature survey. J. Am. Soc. Inf. Sci. 64 (2), 217–233. https://doi.org/10.1002/asi.22803.

Brandt, P., Ernst, A., Gralla, F., Luederitz, C., Lang, D.J., Newig, J., Reinert, F., Abson, D.J., von Wehrden, H., 2013. A review of transdisciplinary research in sustainability science. Ecol. Econ. 92, 1–15. https://doi.org/10.1016/j.ecolecon.2013.04.008.

Burkhardt-Holm, P., Zehnder, A.J.B., 2018. Fischnetz: assessing outcomes and impacts of a project at the interface of science and public policy. Environ. Sci. Policy 82, 52–59. https://doi.org/10.1016/j.envsci.2018.01.010.

Carew, A.L., Wickson, F., 2010. The TD Wheel: a heuristic to shape, support and evaluate transdisciplinary research. Futures 42 (10), 1146–1155. https://doi.org/10.1016/j.futures.2010.04.025.

Cash, D., Clark, W., Alcock, F., Eckley, N., Jäger, J., 2002. Salience , credibility,

legitimacy and boundaries: linking research assessment and decision making. KSG Working Papers Series RWP02-046. SSRNhttps://doi.org/10.2139/ssrn.372280.
CGIAR, 2016. CGIAR Strategy and Results Framework 2016-2030: Overview. Retrieved from https://cgspace.cgiar.org/handle/10947/3865.
CGIAR, 2017. Funding and Finance Highlights From 2017. Retrieved from https://www.cgiar.org/funding-and-finance-highlights-from-2017/.
CIFOR, 2019a. Fire and Haze: Background. Retrieved from https://www.cifor.org/fire-and-haze/research/background/.
CIFOR, 2019b. Global Comparative Study on Forest Tenure Reform: Research Background. Retrieved from https://www.cifor.org/gcs-tenure/research/research-background/.
CIFOR, 2019c. SWAMP: Sustainable Wetlands Adaptation and Mitigation Program. Retrieved from https://www.cifor.org/swamp/.
Clark, W.C., Dickson, N.M., 2003. Sustainability science: the emerging research program. Proc. Natl. Acad. Sci. U. S. A. 100 (14), 8086–8091. https://doi.org/10.1073/pnas.1231333100.
Cook, T., Boote, J., Buckley, N., Vougioukalou, S., Wright, M., 2018. Accessing participatory research impact and legacy: developing the evidence base for participatory approaches in health research. Educ. Action Res. 24 (4), 473–488. https://doi.org/10.1080/09650792.2017.1326964.
Court, J., Maxwell, S., 2005. Policy entrepreneurship for poverty reduction: bridging research and policy in international development. J. Int. Dev. 17 (6), 713–725. https://doi.org/10.1002/jid.1234.
Donovan, C., Hanney, S., 2011. The "Payback Framework" explained. Res. Eval. 20 (3), 181–183. https://doi.org/10.3152/095820211X13118583635756.
Earl, S., Carden, F., Smutylo, T., 2001. Outcome Mapping: Building Learning and Reflection into Development Programs. International Development Research Centre, Ottawa.
ESRC, 2009. Measuring the Impact of ESRC Funding. Retrieved from https://esrc.ukri.org/files/research/research-and-impact-evaluation/measuring-the-impact-of-esrc-funding/.
Flint, C.G., Dean, K.T., Yang, B., Jackson-Smith, D., 2019. Socio-scientific research and practice: evaluating outcomes from a transdisciplinary urban water systems project. Socio-Ecol. Pract. Res. 1, 55–66.
FTA, 2019. Forest Trees and Agroforestry. Retrieved from http://foreststreesagroforestry.org/.
Funtowicz, S.O., Ravetz, J.R., 1993. Science for the post-normal age. Futures 25 (7), 739–755. https://doi.org/10.1016/0016-3287(93)90022-L.
Gaziulusoy, A.I., Ryan, C., McGrail, S., Chandler, P., Twomey, P., 2016. Identifying and addressing challenges faced by transdisciplinary research teams in climate change research. J. Clean. Prod. 123, 55–64. https://doi.org/10.1016/j.jclepro.2015.08.049.
Gehlert, S., Hall, K., Vogel, A., Hohl, S., Hartman, S., Nebeling, L., Redline, S., Schmitz, K., Patterson, R., Thompson, B., 2014. Advancing transdisciplinary research: the transdisciplinary research on energetics and Cancer initiative. J. Transl. Med. Epidemiol. 2 (2), 1032–1047. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25356437.
Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., Trow, M., 1994. The New Production of Knowledge, SAGE. Retrieved from https://ia601600.us.archive.org/21/items/mode1_2/mode1_2.pdf.
Halimanjaya, A., Belcher, B., Suryadarma, D., 2018. Getting forest science to policy discourse : a theory-based outcome assessment of a global research programme. Int. For. Rev. 20 (4), 469–487.
Hanberger, A., 2001. What is the policy problem? Methodological challenges in policy evaluation. Evaluation 7 (1), 45–62. https://doi.org/10.1177/13563890122209513.
Heinrichs, H., Martens, P., Michelsen, G., Wiek, A. (Eds.), 2016. Sustainability Science: An Introduction. Springer, Dordrecht.
ICRAF, 2019. Support to the Development of Agroforestry Concessions in Peru. Retrieved from http://www.worldagroforestry.org/project/support-development-agroforestry-concessions-peru.
ISPC, 2015. Strategic Study of Good Practice in AR4D Partnership. Rome, Italy. Retrieved from https://ispc.cgiar.org/sites/default/files/ISPC_StrategicStudy_Partnerships.pdf.
Jahn, T., Bergmann, M., Keil, F., 2012. Transdisciplinarity: between mainstreaming and marginalization. Ecol. Econ. 79, 1–10. https://doi.org/10.1016/j.ecolecon.2012.04.017.
Kates, R.W., 2017. Sustainability science. In: Richardson, D., Castree, N., Goodchild, M.F., Kobayashi, A., Liu, W., Marston, R.A. (Eds.), The International Encyclopedia of Geography: People, the Earth, Environment and Technology. Wiley-Blackwell, pp. 1–4.
Kates, R., Clark, W.C., Correll, R., Hall, M., Jaeger, C., Lowe, I., McCarthy, J.J., Schellnhuber, J.S., Bolin, B., Dickson, N.M., Faucheux, S., Gallopin, G.C., Grübler, A., Huntley, B., Jäger, J., Jodha, N.S., Kasperson, R.E., Mobogunje, A., Matson, P., Mooney, H., Moore, B., O'Riordan, T., Svedin, U., 2001. Sustainability science. Science 641–642. https://doi.org/10.1126/science.1059386.
Kauffman, J., Arico, S., 2014. New directions in sustainability science: promoting integration and cooperation. Sustain. Sci. 9 (4), 413–418. https://doi.org/10.1007/s11625-014-0259-3.
Klein, J.T., 2006. Afterword: the emergent literature on interdisciplinary and transdisciplinary research evaluation. Res. Eval. 15 (1), 75–80. https://doi.org/10.3152/147154406781776011.
Komiyama, H., Takeuchi, K., 2006. Sustainability science: building a new discipline. Sustain. Sci. 1, 1–6. https://doi.org/10.1007/s11625-006-0007-4.
Lang, D.J., Wiek, A., Bergmann, M., Stauffacher, M., Martens, P., Moll, P., Swilling, M., Thomas, C.J., 2012. Transdisciplinary research in sustainability science: practice, principles, and challenges. Sustain. Sci. 7 (SUPPL.1), 25–43. https://doi.org/10.1007/s11625-011-0149-x.
Mayne, J., 2012. Contribution analysis: Coming of age? Evaluation 18 (3), 270–280. https://doi.org/10.1177/1356389012451663.
Mayne, J., Stern, E., 2013. Impact Evaluation of Natural Resource Management Research Programs: a Broader View. ACIAR Impact Assessment Series Report No.84. . Retrieved from https://pdfs.semanticscholar.org/3830/68b2f4f81014327ea49d7d4db0540e2fef35.pdf.
Mayne, J., Stern, E., Douthwaite, B., 2013. AAS Practice Brief: Evaluating Natural Resource Management Programs. CGIAR Research Program on Aquatic Agricultural Systems, AAS-2013-23, Penang, Malaysia. Retrieved from http://www.worldfishcenter.org/resource_centre/AAS-2013-23.pdf.
Mitchell, C., Cordell, D., Fam, D., 2015. Beginning at the end: the outcome spaces framework to guide purposive transdisciplinary research. Futures 65, 86–96. https://doi.org/10.1016/j.futures.2014.10.007.
ODI, Overseas Development Institute, 2012. RAPID Outcome Assessment. Retrieved from https://www.odi.org/publications/6800-rapid-outcome-assessment.
Ozanne, J.L., Davis, B., Murray, J.B., Grier, S., Benmecheddal, A., Downey, H., Ekpo, A.E., Garnier, M., Hietanen, J., Le Gall-Ely, M., Seregina, A., Thomas, K.D., Veer, E., 2017. Assessing the societal impact of research: the relational engagement approach. J. Public Policy Mark. 36 (1), 1–14.
Phillips, P., Moutinho, L., Godhino, P., 2018. Developing and testing a method to measure academic societal impact. Higher Educ. Quart. 72, 121–140.
Pielke, R.A., 2007. The Honest Broker. Cambridge University Press, Cambridge.
Pohl, C., 2011. What is progress in transdisciplinary research? Futures 43 (6), 618–626. https://doi.org/10.1016/j.futures.2011.03.001.
Pohl, C., Hirsch Hadorn, G., 2007. Principles for Designing Transdisciplinary Research. Oekom, Munich.
Pohl, C., Rist, S., Zimmermann, A., Fry, P., Gurung, G.S., Schneider, F., Speranza, C.I., Kiteme, B., Boillat, S., Hirsch Hadorn, E.S.G., Wiesmann, U., 2010. Researchers' roles in knowledge co-production: experience from sustainability research in Kenya, Switzerland, Bolivia and Nepal. Sci. Public Policy 37 (4), 267–281. https://doi.org/10.3152/030234210X496628.
Polk, M., 2014. Achieving the promise of transdisciplinarity: a critical exploration of the relationship between transdisciplinary research and societal problem solving. Sustain. Sci. 9, 439–451.
Ramirez, L.F., Belcher, B.M., 2018. Crossing the science-policy interface: lessons from a research project on Brazil nut management in Peru. For. Policy Econ. https://doi.org/10.1016/J.FORPOL.2018.07.018.
Ramirez, L.F., Belcher, B.M., 2019. Stakeholder perceptions of scientific knowledge in policy processes: a Peruvian case-study of forestry policy development. Sci. Public Policy. https://doi.org/10.1093/scipol/scz003.
Reale, E., Avramov, D., Canhial, K., Donovan, C., Flecha, R., Holm, P., Larkin, C., Lepori, B., Mosoni-Fried, J., Primeri, E.O.E., Puigvert, L., Scharnhorst, A., Schubert, A., Soler, M., Soòs, S., Sordé, T., Travis, C., van Horik, R., 2017. A review of literature on evaluating the scientific, social and political impact of social sciences and humanities research. Res. Eval. 27 (4), 298–308. https://doi.org/10.1093/reseval/rvx025.
Renkow, M., Byerlee, D., 2010. The impacts of CGIAR research: a review of recent evidence. Food Policy 35, 391–402. https://doi.org/10.1016/j.foodpol.2010.04.006.
Rockwell, C.A., Guariguata, M.R., Menton, M., Arroyo Quispe, E., Quaedvlieg, J., Warren-Thomas, E., Fernandez Silva, H., Rojas, E.E.J., Arrunátegui, J.A.H.K., Vega, L.A.M., Vera, O.R., Hancco, R.Q., Tito, J.F.V., Panduro, B.T.V., Salas, J.J.Y., 2015. Nut production in *Bertholletia excelsa* across a logged forest mosaic: implications for multiple forest use. PLoS One 10 (8), e0135464. https://doi.org/10.1371/journal.pone.0135464.
Roux, D.J., Nel, J.L., Cundill, G., O'Farrell, P., Fabricius, C., 2017. Transdisciplinary research for systemic change: who to learn with, what to learn about and how to learn. Sustain. Sci. 12 (5), 711–726. https://doi.org/10.1007/s11625-017-0446-0.
Sarewitz, D.R., 1996. Frontiers of Illusion: Science, Technology, and the Politics of Progress. Temple University Press, Philadelphia, PA.
Sarkki, S., Niemelä, J., Tinch, R., van den Hove, S., Watt, A., Young, J., 2014. Balancing credibility, relevance and legitimacy: a critical assessment of trade-offs in science-policy interfaces. Sci. Public Policy 41 (2), 194–206. https://doi.org/10.1093/scipol/sct046.
Savedoff, W.D., Levine, R., Birdsall, N., 2005. When will we ever learn? Recommendations to improve social development through enhanced impact evaluation. Center Glob. Dev. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.483.4704&rep=rep1&type=pdf.
Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., Befani, B., 2012. Broadening the Range of Designs and Methods for Impact Evaluations (Working Paper No.38). Retrieved from https://www.oecd.org/derec/50399683.pdf.
Stokes, D.E., 1997. Pasteur's Quadrant: Basic Science and Technological Innovation. The Brookings Institution, Washington, D.C.
Talwar, S., Wiek, A., Robinson, J., 2011. User engagement in sustainability research. Sci. Public Policy 38, 379–390.
Walter, A.I., Helgenberger, S., Wiek, A., Scholz, R.W., 2007. Measuring societal effects of transdisciplinary research projects: design and application of an evaluation method. Eval. Program Plann. 30 (4), 325–338. https://doi.org/10.1016/j.evalprogplan.2007.08.002.
White, H., Phillips, D., 2012. Addressing Attribution of Cause and Effect in Small N Impact Evaluations: Towards an Integrated Framework. Better Evaluation (Working Paper No.15). Retrieved from https://betterevaluation.org/en/resources/overview/attribution_cause-effect_impact_evaluations.
Wiek, A., Talwar, S., O'Shea, M., Robinson, J., 2014. Toward a methodological scheme for

capturing societal effects of participatory sustainability research. Res. Eval. 117–132. https://doi.org/10.1093/reseval/rvt031.

Wiek, A., Withycombe, L., Redman, C.L., 2011a. Key competencies in sustainability: a reference framework for academic program development. Sustain. Sci. 6 (2), 203–218. https://doi.org/10.1007/s11625-011-0132-6.

Wiek, A., Withycombe, L., Redman, C., Mills, S.B., 2011b. Moving forward on competence in sustainability research and problem solving. Environment 53 (2), 3–12. https://doi.org/10.1080/00139157.2011.554496.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S.R., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., Johnson, B., 2015. The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. https://doi.org/10.13140/RG.2.1.4929.1363.

Wolf, B., Lindenthal, T., Szerencsits, M., Holbrook, J.B., Heß, J., 2013. Evaluating research beyond scientific impact: how to include criteria for productive interactions and impact on practice and society. Gaia 22 (2), 104–114.