Environmental Science and Policy 103 (2020) 32–40

Contents lists available at ScienceDirect

Environmental Science and Policy

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

Combining the roles of evaluator and facilitator: Assessing societal impacts of transdisciplinary research while building capacities to improve its quality

L. Verwoerd*, P. Klaassen, S.C. van Veen, R. De Wildt-Liesveld, B.J. Regeer

Athena Institute, VU University Amsterdam, Faculty of Science, De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Transdisciplinary research
Evaluation
Impact assessment
Capacity building
Researcher roles
Research quality

ABSTRACT

Participation of relevant stakeholders, knowledge integration, responsive and emergent design and effective boundary management are four key features of transdisciplinary research (TDR). These features pose significant challenges to both undertaking TDR and evaluating its societal impact. We argue that TDR's context specificity and complexity warrant an evaluation approach that supports the coordinating team in developing these key features. In light of this, this article aims to reconcile two distinct foci of TDR evaluation, namely supporting transdisciplinary capacity building and impact evaluation. We share the results from a combined approach in which the authors acted both as facilitators and evaluators of a TDR project, to conduct an embedded, formative evaluation. Our findings show that the approach allowed for better access to the participants and sensitivity to their perspectives on impact, and for enhanced understanding of complex internal and external project dynamics and how these shaped the project. This resulted in a meaningful assessment of TDR's societal impacts and enabled attributing these to specific process elements. Moreover, the approach supported the coordinating TDR team's capacities for developing key TDR features. Four TDR capacities were identified: *building TDR ownership, openness and transparency for integrating divergent TDR needs, purposeful responsiveness to emergent TDR needs* and *navigating institutional realities and TDR ambitions.* The approach presented may serve as stepping stone for the TDR community to further the conversation on (the impact of) inclusive, reflexive and responsive research.

1. Introduction

Accompanying today's many complex and critical societal issues – such as climate change, depletion of natural resources, global food security – is a rise in demand for transdisciplinary research (TDR). In TDR, academics and societal actors collaborate to integrate knowledge and develop socially robust answers to real-world issues (Pohl, 2011). It is assumed that TDR is better equipped to contribute to solving the complex problems facing society than mono- or interdisciplinary research (Hansson and Polk, 2018). The growing practice of TDR makes it more pressing to look beyond its intentions, and to ask how its impacts can be shown and its quality assessed, and ultimately tied to these impacts.

Four key TDR features include participation of relevant stakeholders, knowledge integration, responsive and emergent designs that allow for the research to develop as insights unfold, and managing boundary dynamics (Hoffmann et al., 2017; Regeer and Bunders, 2009; Scholz and Steiner, 2015). These features, however, also make it particularly challenging to evaluate TDR (Walter et al., 2007). As TDR, by definition, crosses disciplinary boundaries and aims to achieve impacts beyond the scientific realm, conventional academic quality or impact criteria are inadequate (Belcher et al., 2016). Even if standardized criteria were readily available, they would have to be sufficiently flexible to accommodate every project's contextual specificity (Carew and Wickson, 2010). Furthermore, rather than a linear process of formulating a problem, research, and achieving impact, TDR is more like iterative and experimental interactions between actors from different domains (Regeer et al., 2009). TDR's many contingent internal and external project dynamics are hard to accommodate in evaluations that build upon predefined quality criteria (Hansson and Polk, 2018).

The literature has proliferated with analytical and methodological frameworks to measure TDR impacts (Walter et al., 2007) or to assess the quality of TDR conduct (Wickson et al., 2006). Others have focused on developing principles and criteria for TDR processes, emphasizing on-going reflection and learning by researchers and practitioners to ensure quality (Lang et al., 2012). Here, another purpose of evaluation comes to the fore: next to assessing impact and quality, scholars have argued that evaluation may also support researchers and practitioners in dealing with TDR complexities and contingencies, and support learning-by-doing (Bergmann et al., 2005; Zscheischler et al., 2018).

* Corresponding author.
E-mail address: l.verwoerd@vu.nl (L. Verwoerd).

https://doi.org/10.1016/j.envsci.2019.10.011
Received 15 April 2019; Received in revised form 14 October 2019; Accepted 16 October 2019
Available online 01 November 2019

This article aims to reconcile these two distinct foci – transdisciplinary capacity building and impact evaluation – of TDR evaluation. We argue that TDR's complex nature indeed warrants an evaluation approach that is simultaneously supportive of this complexity and hypothesize that combining these evaluation foci meets this purpose. While others have previously reported on similar endeavors (e.g. Gaziulusoy et al., 2016; Roux et al., 2010; Zscheischler et al., 2018), we explicitly examine how simultaneously supporting and assessing TDR might be complementary and allow for more meaningful impact evaluation and enhanced TDR quality.

To make this case, the next section discusses challenges of TDR evaluation in greater depth. Section 3 presents our case description: the evaluation of a TDR project on Dutch nature policy during which we combined the roles of facilitators and evaluators. Section 4 shows that our approach simultaneously supporting transdisciplinary learning by the coordinating TDR team regarding the operationalization of four key TDR features, thereby improving the quality of the TDR process, and for meaningful assessment of the project's societal impacts and attributing these to specific process features. Section 5 critically reflects on these findings and the approach.

2. Evaluation of TDR projects

As TDR becomes increasingly common it becomes necessary to demonstrate its societal effects and account for the resources invested. Various attempts have been made to empirically capture TDR impacts and TDR has been linked to more usable research products, denser stakeholder networks, enhanced decision-making capacities and policy change (Walter et al., 2007). Wiek et al. (2014) differentiate between TDR *outputs* (usable products) and *outcomes* (network effects and enhanced capacities), both considered intermediate effects which, indirectly and in complex interplay, contribute to societal *impacts* (structural changes and action). The latter tends to occur with significant delay and is found harder to attribute to the specific TDR project (Hansson and Polk, 2018).

To explain – and ultimately advance – these various effects, scholars have been seeking for ways to measure the quality of TDR processes. For instance, Belcher et al. (2016) suggest the perceived *credibility, legitimacy* and *relevance* of TDR research by stakeholders as determinant for impact. However, the complex (political) contexts in which TDR projects take place may make for highly diverse stakeholder views on a project's credibility, legitimacy and relevance, complicating the use of these concepts in guiding the TDR process (Hansson and Polk, 2018). Others have focused on differentiating between types of 'productive interactions' between researchers and stakeholders as the key to quality and societal impact (de Jong et al., 2016), for which Wiek et al. (2014) distinguish between the *nature* (number, type and sequence of interactions) and the *quality* (representation of perspectives, addressing conflict) of participatory processes.

Evaluators, however, are often faced with a lack of high-quality data to sufficiently test these conceptualizations, due to low participation rates and time-lag, which affects stakeholders' memories (Wiek et al., 2014). To fully grasp the complexity that surrounds a project and make informed judgements on the value of its effects and the quality of its process, it is vital to have access to practitioners' perspectives. One reason why TDR evaluation has yet to fully address these challenges is that evaluations run the risk of becoming *decoupled* from the project in question. The evaluator tasked to assess societal impact does so at a relative distance. From this position it is almost impossible to comprehend how a project developed in response to internal and external dynamics, and what, through this lens, constitutes impact, or what might be the appropriate criteria to assess the quality of the project's process (Regeer et al., 2009). To address this, we suggest that an embedded approach geared towards monitoring is better suited for assessing impacts and attributing these to the TDR process.

This is consistent with the observation that the complex character of TDR requires an evaluation approach that is supportive of this complexity (Carew and Wickson, 2010; Klaassen et al., 2019). This is pertinent, because it is unlikely that the features the literature suggests as key to successful TDR are all in place when a project commences. Think, for instance, of stakeholders' commitment to collaborate, openness to other worldviews and capacities for bridging epistemic cultures. A novice TDR team can hardly be expected to meet such conditions immediately, or even to know how to develop these from the outset. Because of each project's unique nature this even holds for experienced teams, as each project differs regarding, for example, the relevant stakeholders, their interests and cultures. There is no one-size-fits-all recipe – or 'blueprint' – for inducing effective collaboration in relation to diverse contextual factors (Bracken et al., 2014:5).

In light of this, various scholars have argued that researchers who 'do' TDR require additional capacities (Pohl et al., 2010), such as critical awareness of stakeholders' diverse assumptions, values and worldviews and how these shape participatory research processes (Popa and Guillermin, 2017). Such capacities may be promoted through participatory action–research approaches (Gaziulusoy et al., 2016; Roux et al., 2010), in which insights into the experiences with the ongoing TDR project function as direct feedback mechanism to improve its quality. Pohl et al. (2010) identify the role of facilitators: those researchers who are tasked with promoting joint reflection and transdisciplinary learning. We would argue that these facilitators may also be the most appropriate to evaluate the TDR project. Indeed, a facilitator who helps a TDR team build the required capacities to develop key TDR features may also have the best access to researchers' and practitioners' perspectives on its impacts. Being a facilitator allows for a 'view from the trenches' – a comprehensive understanding of the project's intricacies and how these relate to its outcomes, which we argue is necessary for meaningful impact evaluation.

3. Methodology

3.1. Case study

We present a case study of the first period of the Natuurpact research program: a large-scale, long-term TDR program in the Netherlands. Conflicting agricultural interest and nature conservation goals have resulted in Dutch nature policy becoming increasingly polarized. Uncoordinated attempts on the part of national and provincial governments to address this polarization were unsuccessful and generated a level of conflict between these governmental bodies. As a step forwards, national and provincial governments, and a number of societal organizations, signed the Natuurpact agreement (2013). In this agreement they finalized the decentralization of nature policy to the provinces and agreed on ambitions to halt the decline biodiversity and increase social engagement with nature. Part of the agreement was a transdisciplinary policy research program geared at mutual learning and increased nature policy impact: the Natuurpact program. The program is conducted by a team of researchers from the PBL Netherlands Environmental Assessment Agency, (*Planbureau voor de Leefomgeving,* PBL), a government expert organization, and Wageningen Environmental Research (WER), a university research department, with support of the authors (Athena Institute) as both facilitators and evaluators. It is planned to run until 2028 and comprises sequential and generative three-year research periods. The findings we present are derived from its first research period (2014–2017), which we refer to as 'project' or 'case' for the remainder of this paper (Fig. 1).

An interdisciplinary core team of six researchers from the PBL and the WER conducted the project (including two project leaders). Few had prior experience with TDR. The project team met twice-weekly to discuss progress and plan research activities. The participants were primarily provincial policy actors who are responsible for the development and execution of nature policy since the decentralization, and national policy actors who are responsible for international obligatory

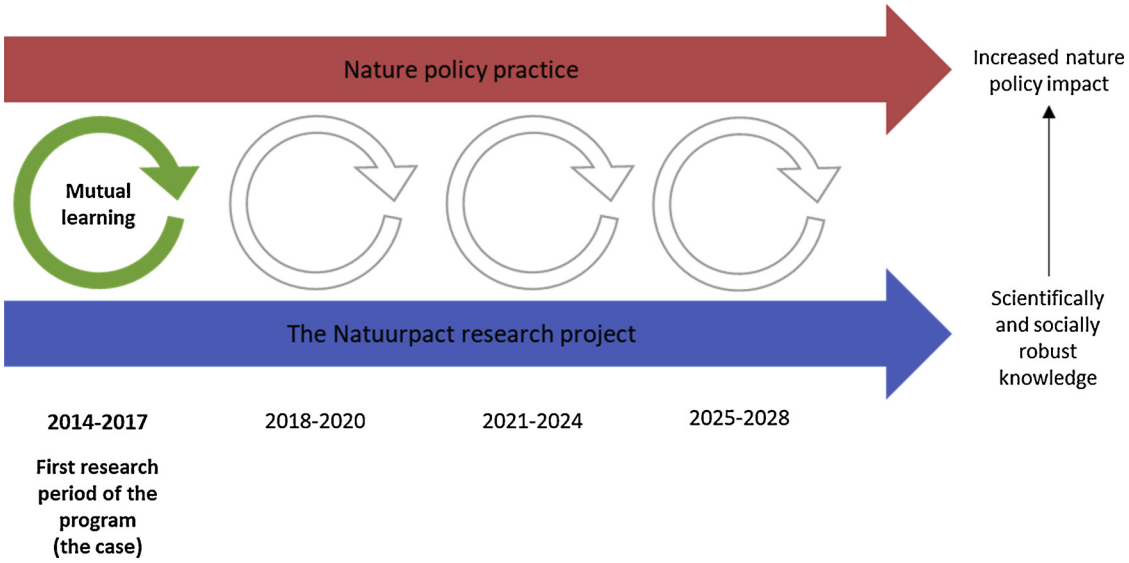

Fig. 1. Timeline of the entire Natuurpact research program (2014–2028). The case presented in this paper concerns the program's first period (2014–2017).

biodiversity goals. Societal actors such as nature organizations were consulted, but not intensively engaged.

The project comprised three phases: 1) developing a joint research design, 2) conducting research and shared sense-making of findings, and 3) joint formulation of action and dissemination.

The main interaction between the team and participants occurred through bi-monthly meetings with a formal working group of 12 representatives from provincial government and eight multi-stakeholder workshops that took place throughout the project and whose purpose differed in accordance with the respective research phase. Interaction also took place with the program commissioners (administrators from national and provincial government) twice a year, to check that the project was still on track. Fig. 2 depicts a schema of the project's design.

3.2. Roles, material and methods

In this section we describe our research design according to the two roles we combined: facilitator and evaluator.

3.2.1. Facilitators

Authors A1, A3 and A5 were (variably) part of the project team and supported the team with their first TDR endeavor. As facilitators, they encouraged joint reflection on the challenges posed by developing and practicing the four key TDR features. Due to TDR's contextual specificity, the coordinating team is required to operationalize these more generic TDR features into a bespoke design that adequately corresponds to the issue at hand. Challenges surrounding this operationalization are therefore considered inherent part of the conduct of TDR; the right

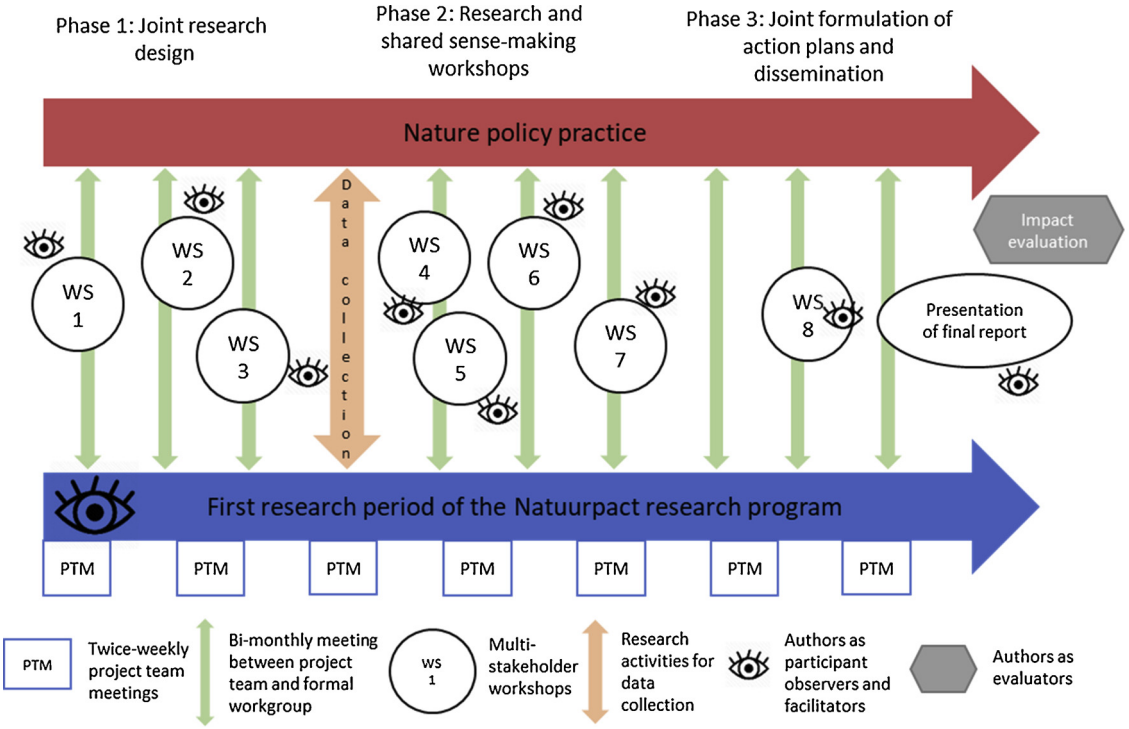

Fig. 2. Schematic overview of our case, the first period (2014–2017) of the Natuurpact research program.

conditions, knowledge, competences, circumstances are never in place from the start. The four TDR key features that were to be operationalized are: 1) participation by relevant stakeholders; 2) knowledge integration for change; and 3) responsive and emergent design; and 4) effective boundary management (Hoffmann et al., 2017; Regeer and Bunders, 2009; Scholz and Steiner, 2015).

The facilitators recorded the developments in the team regarding the challenges associated with these features and other internal and external project developments in field notes, as well as participant observations made during the multi-stakeholder workshops. The facilitators also supported the design (and facilitation) of these workshops, and the operationalization of the TDR features featured heavily in the discussions around their design. Interviews were held with individual team members at the start and halfway through the project on their views on the project's successes or failings, which were audio-recorded and transcribed. The insights from these data on the project's progress were used to inform future sessions of joint reflection and served as feedback mechanism.

To identify the development of the team's TDR capacities, we retrospectively analyzed these data using content analysis (Hsieh and Shannon, 2005) and focused on the team's encountered challenges regarding the TDR features, their actions to overcome these, as well changes in the team's overall narrative regarding these challenges.

3.2.2. Evaluators

A1, A2, A4 and A5 evaluated the project after its conclusion through an exclusively qualitative approach (including semi-structured interviews followed by a focus group discussion (FGD). Here, the facilitators' earlier involvement provided important contextual background and familiarization with the project and its participants that informed the evaluation design, such as the respondent selection and the design of the interview and FGD guidelines. The interviewees included seven provincial and three national policy actors who were selected on the basis of diversity in their degrees of participation (ranging from two workshops attended, to being a member of the working group) and levels of enthusiasm for the TDR approach (to ensure critical appraisal). As we were interested in the participants' views on what constituted impact and attributable process elements, we adopted Hellström's action-value attribution framework (2015; Hansson and Polk, 2018). During the interviews, the participants were asked in which ways the project had been of value to their policy practice and to what particular elements of the process they attributed this value. A printed timeline of the project's main events and development was produced based on the facilitators' knowledge combined with project publications, meeting minutes and a guided team self-evaluation (similar to Fig. 2). This timeline worked to structure the interviews (which were otherwise flexible and open-ended) and aided potential memory distortion. All interviews were audio-recorded and transcribed. All data were analyzed using content analysis, focused on categorizing different types of effects and attributed process elements.

The preliminary findings from the interviews were presented and discussed during an FDG with the entire working group, for validation through member checks and joint sense-making. We frequently reminded the participants of our dual role as facilitators and evaluators of the TDR project for full disclosure of our research purposes.

4. Results

In this section, we first present the findings obtained through our role as facilitators and discuss the four key TDR features and highlight actions of the team to develop these and the corresponding capacities they built.

4.1. Key TDR features and corresponding capacities

From the outset, it was evident to the team that they would conduct a long-term transdisciplinary study aimed at mutual learning and improving nature policy impact. Much, however, remained unclear: how to develop particular TDR features, such as stakeholder participation, were matters the team had to learn along the way.

4.1.1. Participation by relevant stakeholders

The first TDR key feature is the participation of relevant stakeholders to address real-world problems and to access their knowledges for socially robust solutions to these issues. During the first project phase, the team approached provincial policy actors to develop a joint research design. They were faced with limited willingness to participate in the research, despite the provincial agreement given by signing the Natuurpact. The team explored the concerns the provinces had for participation through informal conversations, through which they learned about the levels of mistrust between national and provincial government, which had intensified during the decentralization. Some provinces suspected the research was a strategic move by national government to retain control over nature policy, despite the recent decentralization. This made the provinces hesitant to open up their policy processes to the team. Stakeholder participation was the first TDR feature the team had to develop.

The team decided to focus their efforts on encouraging the provinces' buy-in to the project and its transdisciplinary ambitions. Three multi-stakeholder workshops were organized to come to a joint research design. During the first two, primarily provincial representatives were invited (approximately 50 versus five from national government) to stimulate their ownership over the project. Nevertheless, the provinces' limited willingness to participate continued to create difficulties in the second phase, during the actual research and shared sense-making of the findings. When the team approached the provinces to collect data on policy plans some withheld information and questioned the legitimacy of the team. The team decided to visit each province individually to explain face to face their intended co-partnership, by which they learned that some provinces strongly felt nature policy was their prerogative and experienced the research as invasive, illustrating their strong sense of ownership over nature policy. By visiting personally, the team came to understand the provinces' point of view, through which rapport was built and access to provincial policy plans for analysis permitted.

In parallel, the team also used other strategies to further incentivize participation. For instance, the team sought provincial government officials from a high strategic level who functioned as ambassadors of the TDR approach. These officials underlined the importance and urgency for the Natuurpact program and their endorsement also encouraged provincial participation.

These diverse strategies to encourage stakeholder participation eventually proved successful: the provinces started to share their policy processes. We observed that by the time the workshops for shared sense-making were organized near the end of phase two, provincial participation had dropped off the team's list of challenges. We identify the respective capacity that the team built was *developing TDR ownership* for stakeholder participation.

4.1.2. Knowledge integration for change

The second key TDR feature is knowledge integration for change. In light of this, the team made an inventory of the participants' TDR needs during the first three workshops (phase one) for the joint research design. These needs, however, proved difficult for the team to translate into feasible research questions because of their largely operational character. Furthermore, the needs were highly diverse; not just among provinces, but also between government tiers there was large diversity. In addition, some team members had their own 'expert view' on what would be relevant research questions. For the final design the team decided to build on the inventoried needs, but also communicated that feasibility of the research design would be a leading criterion.

The workshops in phase two focused on knowledge integration by

means of shared sense-making of the research findings. This time it was the team who had to share details of their research process just as the participants had to share their policy plans. As facilitators, we observed how unnerving this was to the team: what if the participants did not recognize the findings? Would their expertise be questioned? Despite feeling vulnerable, the team decided that for mutual learning and knowledge integration, equal footing between them and the participants was paramount. They decided to open up the 'black box' of their analyses and explain their work in way that would allow for deliberation and joint interpretation, by using visualizations and steering clear of jargon. While the team was initially nervous, this approach proved an important success for the entire project, as we will discuss in Section 4.2. The mode of working initiated by this approach was continued during the project, and led to the development of three shared principles for collaboration: openness, transparency and being upfront about decisions, referred to as 'working without surprises' (*verrassingsvrij werken*).

However, the workshops in phase two also highlighted a mismatch between the scale at which some findings were presented (national) and the scale at which provinces sought policy recommendations (provincial). We observed that to the team it had been self-evident that their models would not produce scientifically sound findings at such a local scale, while the provinces felt the national scale held little relevance to their practice. It appeared that the research design had not been communicated with this level of detail because of a failure to understand the need for it. While the mismatch was addressed (which is discussed next), this was an important lesson in expectation management for the team.

We observed the team gradually became skilled in integrating knowledge in a way that ensured the research addressed the participants' divergent needs. Relevant to this was their mode of 'working without surprises' that embodied the equal footing between researchers and participants, as well as expectation management. The capacity that we saw built to develop the feature of knowledge integration for change, was that of *openness and transparency for integrating divergent TDR needs.*

4.1.3. Responsive and emergent design

The third key TDR feature is its responsive and emergent design, which allows the research to develop as insights increase and TDR needs develop. As discussed previously, a mismatch was identified regarding the scale at which the findings had relevance. With intent of being responsive to the participants' needs, the team allocated resources to resolve the mismatch, which resulted in a significantly greater workload that in turn compromised research feasibility. The team experienced a tension between a responsive and emergent design on the one hand, and institutional realities of available time and budget on the other. The lessons that were drawn concerned the need to build in budgetary space and capacity to allow for contingencies, and for critical consideration of which needs the project should be responsive to, namely those that enhance the usability of the findings to contribute to real-world problems.

In line with TDR ideology, as the conclusion of the project approached, the team discussed options for co-authoring the final report with the participants to underline their mutual investments and co-partnership. To the team's surprise, the participants were opposed to this idea. They made clear that they preferred an independent research report with policy recommendations, as these would be more effective for public legitimization of their policy decisions. As a compromise, the final report was published by the PBL and WER, but featured text boxes with stories from the participants on their experiences with the project.

We observed how familiarization with the participants' points of view allowed for the team to develop the research in a way that optimized its relevance and usability, while taking the project's feasibility into account. The capacity the team built corresponding to a responsive and emergent design was that of *purposeful responsiveness to emergent TDR needs.*

4.1.4. Effective boundary management

The final key TDR feature is effective boundary management. This concerns the boundaries between the different worlds and institutional backgrounds – with often different rules and expectations – of the actors that are brought together in TDR. In the project these boundaries were, for example, especially tangible when the team was confronted with the realities from their home organizations. For both the PBL and the WER, TDR was a novel approach and, particularly for the PBL, technocratic conventions on sound policy research prevailed. Consequently, the team frequently had to account for the scientific rigor of the TDR approach. The team managed this boundary, and ensured institutional support for the project, by framing the TDR approach in terms that adhered to PBL's mission statement: as 'a method to enhance policy impact'. Here, we were also mobilized; as TDR 'experts' we were presented by the team as tasked with guarding the project's scientific quality, appealing to the organizations' technocratic rationales.

As facilitators we perceived how the team navigated the different institutional realities of their home organizations by reframing TDR in a way that stroked with the dominant frames within these realities, without compromising TDR's purpose. For effective boundary management, we identified *navigating institutional realities and TDR ambitions* as the final built capacity of the team. Fig. 3 shows all four capacities.

Each capacity that was built proved relevant throughout the entire project: developing TDR features was not just a task at the start of the project, but required constant awareness and anticipation. As the team became more skilled, we observed their demeanor and corresponding narrative developed in concert. For instance, rather than expressing concern when provinces were uncooperative, the team's response became more relaxed as they learned to recognize it as strategic play. It became increasingly second nature to the team to consider the participants' points of view in any research decision, and their confidence with the TDR approach grew as the project progressed.

4.2. Impact assessment

We now turn to the project's effects in terms of outputs, outcomes and impact (as suggested by Wiek et al., 2014) (Fig. 4), followed by process elements that were found to be attributable to these effects. Where applicable, we highlight how the role of facilitator complemented our role of evaluator.

4.2.1. Societal effects

4.2.1.1. Outputs. The outputs of the Natuurpact project comprise its final products: reports and a multitude of presentations, including shared action plans. The participants perceived these as an important prerequisite for societal effects and considered the 'deliverables' useful for attaining or contributing to other outcomes.

4.2.1.2. Outcomes. Wiek et al. (2014) distinguish *network effects* and *enhanced capacities* (which we termed *cognitive effects*), to which we added two effect categories: *affective effects* and *legitimizing effects.*

Network effects

The participants agreed that the increased frequency of their interactions during the project helped expand professional networks and strengthened existing relationships. We could corroborate this with the facilitators' observations from the multi-stakeholder workshops; for example, provinces were seen to talk enthusiastically and change information for future contact. Also illustrative of this effect is the initiation of a provincial 'learning policy network', a platform that focuses on mutual learning on nature policy topics that are outside the scope of the Natuurpact program.

Affective effects

For the following category our role as facilitators was of value in two ways: first, we knew to ask about the participants' relations because

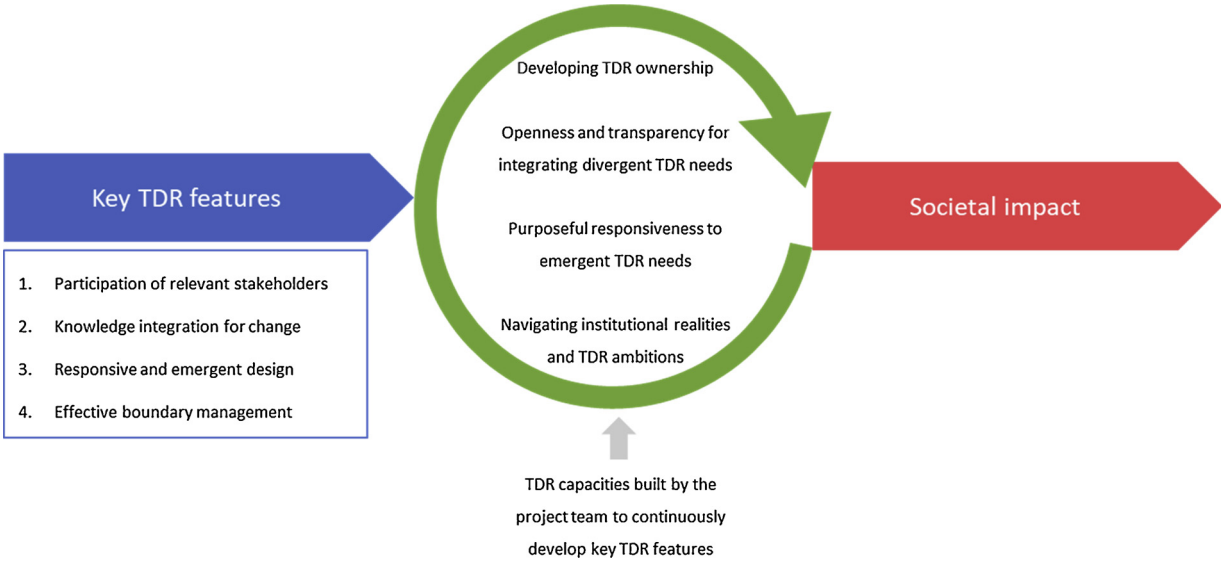

Fig. 3. Four key TDR features and the respective TDR capacities that the project team built.

we were aware that their tense history had affected their participation. Second, as we had become familiar faces to the respondents, they seemed to be at ease with sharing their concerns and reliefs. It was through this that we found that, more so than network effects, the participants valued the project's *affective effects*: interacting with interprovincial colleagues instilled a sense of relief and reassurance through learning that they face similar issues with nature policy, thereby validating their own experiences. The subsequent sense of belonging corresponds with Wiek et al.'s community identity (as part of *network effects*). We also consider the increased levels of trust between national and provincial government and the project team an affective effect, the significance of which we were able to grasp through our knowledge of their history. We single out affective effects as separate outcome category as our findings show it was vital for sustained

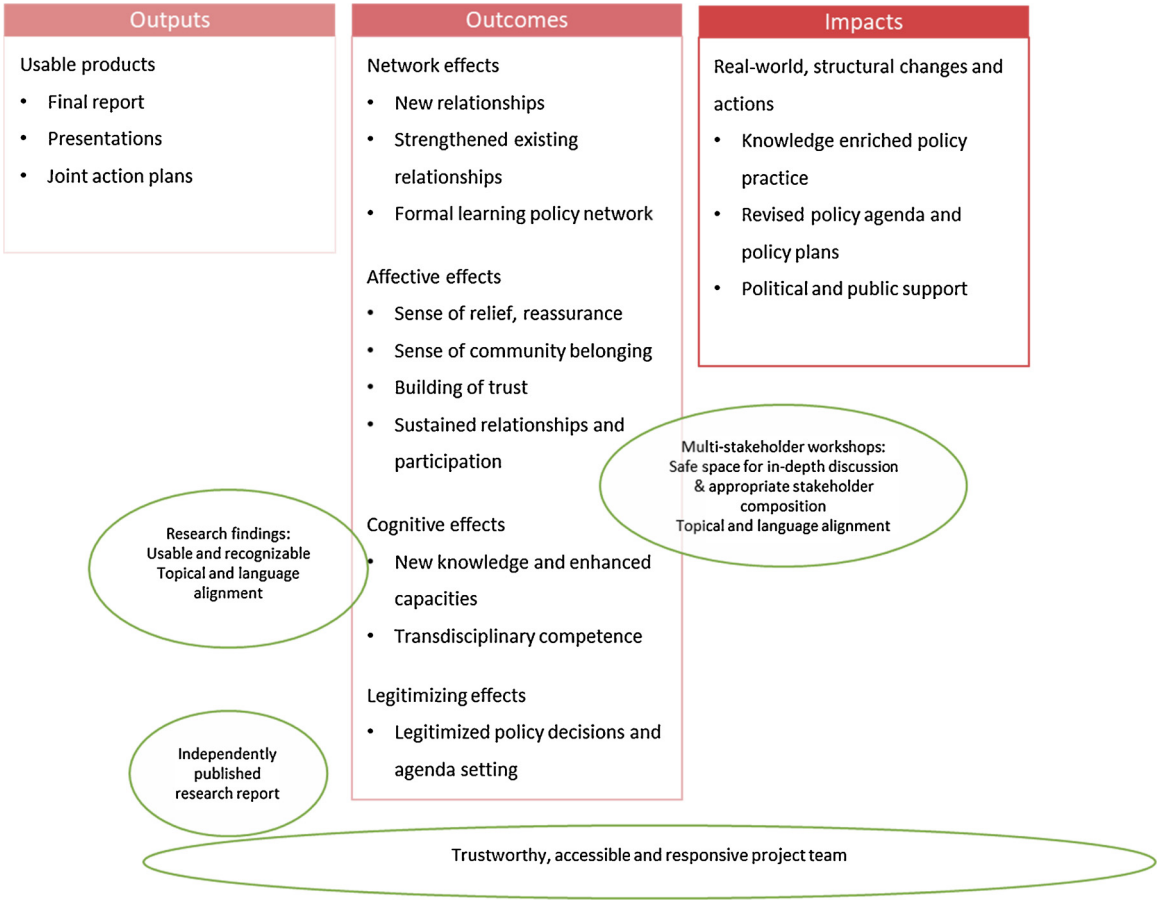

Fig. 4. Societal effects of the Natuurpact project presented as outputs, outcomes and impacts. Process elements that were identified as attributable by the participants are indicated in green ovals.

stakeholder participation.

Cognitive effects

Most participants said they had learned from the project: *cognitive effects*. With regard to what they had learned, we identify two categories. The first comprises the newly acquired knowledge that was produced by the project. This knowledge has been formalized in the project's reports and encompassed system, goal and transformation knowledge (e.g. Walter et al., 2007). Part of this new knowledge was also the shared language that we as facilitators observed had developed between participants and the team. In a similar vein, we also observed conceptual alignment and alignment of purposes with nature policy among the participants. We consider this first category *new knowledge and enhanced understanding.*

The second category concerns knowledge of a more implicit guise. The participants discussed the value of the project for instilling deeper understanding with respect to their own and each other's perspectives and worldviews. They highlighted how this affected their interactions regarding nature policy in a manner congruent with what we understand as 'anticipatory competence' in relation to stakeholder perspectives – we consider this second category *transdisciplinary competence.*

Legitimizing effects

As a final outcome we add the research's *legitimizing effects* to the framework. Knowing the participants had declined co-publication, we inquired whether the anticipated legitimizing effect of the independent report had been a success, and how. They confirmed that it politically and publicly legitimized and justified their policy agendas, while still enjoying the TDR benefits of enhanced understanding and usability of the findings.

4.2.1.3. ImpactsStructural changes and actions

Finally, we turn to *impacts* – effects that are the ultimate goal of TDR, namely *structural changes and actions* (Wiek et al., 2014). The participants said that the project's outputs and legitimizing effects had been instrumental to their practice: the provinces had used the knowledge to inform nature policy programs and set their policy agenda, for which the legitimizing effects issued political and public support. Together with the other effects, which encompass a more social dimension, the development of a *knowledge enriched policy practice* was identified.

4.3. Attributed process elements

In order to understand how the TDR project helped to create these effects, we asked the participants which specific elements of the TDR process were attributable. It was here we experienced our preceding role of facilitators as especially pertinent: knowing the team's actions to develop the four key TDR features, how the project in itself had developed and how workshops and other interactions had passed, guided our inquiry.

The participants identified four main elements that were attributable to the project's impacts. To start with, they attributed networks and affective effects, and parts of the cognitive effects, to the various multi-stakeholder workshops. The provinces mentioned the 'safe space' the workshops had provided, within which they could discuss more sensitive matters, such as their assumptions and opinions on nature policy. We then inquired what specifically had produced this safe space, which was ascribed to the stakeholder composition of workshops: the provinces felt more at ease to discuss potentially sensitive topics when the stakeholder ratio gravitated towards them, as a consequence of the initial distrust between the governments.

Second, the participants attributed cognitive effects and policy actions to the usability and recognizability of the project's findings. Despite the initial mismatch regarding the scale at which the findings were applicable, the research findings had predominantly met the participants' research needs. We also asked about the role of the workshops in light of usability, to which most participants reflected that the alignment of the content of the workshops to the participants' frames of reference – both regarding the workshops' topics and the language used by the team – as crucial element.

The third factor was that the final report was published independently, and therefore attributable to the legitimizing effects of the project. The fact that the participants could say *'this is what science advises'* helped them argue for certain policy decisions. The provinces reflected that their own institutional realities did not allow them to diverge too far from a traditional science–policy relationship, and that the responsiveness of the team by providing a compromise had effectively navigated this tension.

Finally, the participants attributed the project's overall success and quality to the project team. Most spoke highly of the team, in particular regarding their transparency, their accessibility to answer questions and their responsiveness to concerns. Participants explained that their trust in the team grew throughout the project, and that this was an important factor in their motivation for active participation in the project (Fig. 4).

5. Discussion and conclusion

It has been argued that general agreement on how to evaluate TDR societal impact and quality is viewed as the final phase of TDR development as a research discipline (Carew and Wickson, 2010). This article has sought to address the difficulties that have been identified in literature for conducting meaningful evaluation of TDR impact and quality. We have argued that some of these difficulties derive from decoupled evaluation, in which access to participants and comprehensive understanding of a project's intricacies are nearly impossible to attain. We also argued that the complexity of TDR warrants evaluation approaches that support teams in terms developing key TDR features from a project's outset in a way that improves its quality. We have shared the results from a combined approach in which we acted both as facilitators and as evaluators of a TDR project, to conduct an embedded, meaningful evaluation.

We did so by outlining this dual role during the evaluation of the first period of the Natuurpact program, which allowed us to identify several transdisciplinary capacities that the team built to enhance the quality of their research project. The capacities we found are consistent with previous work on TDR researchers' skills and associated challenges. To explain researchers' success in addressing these challenges Sarkki et al. (2013) use the metaphor of sensitivity: researchers' ability to be open to the needs and problem framings of stakeholders, to respect different worldviews, perspectives and forms of knowledge, and to understand biases and power relations. To this we add detail on the various ways in which researchers may subsequently act on the knowledge gathered through their sensitivity. Indeed, the capacities we identified go beyond greater insights and suggest a developed know-how and confidence in dealing with stakeholders and their diverse knowledges and perspectives, demonstrating how personal interactions are vital for initiating contact and encouraging sustained engagement (Woltersdorf et al., 2019). Our findings confirm that developing key TDR features requires a constant effort throughout the project (Di Iacovo et al., 2016; Ribeiro et al., 2019). To better understand the link between transdisciplinary capacities, the research process and its impacts, it would be interesting to explore how these capacities relate to the concepts of credibility, legitimacy and relevance as suggested by Belcher et al. (2016), which we intend to do in the following Natuurpact research program period.

In the results we have highlighted moments where we experienced that our involvement with the project as facilitators of transdisciplinary learning processes proved especially complementary to our work as evaluators. We identify four (interrelated) benefits. First, it increased our access to participants, an issue previously identified by others (Wiek et al., 2014). We enjoyed the support of the project leadership and had become familiar faces to the participants over the course of the

project. For example, we were allowed to use an entire meeting of the formal working group for our FDG, a quite exceptional occurrence due to their demanding schedules. Second, we had obtained a 'lived through' perspective of the project's internal and external context and how this, in complex interplay, had shaped the research. For instance, the tense relationship between national and provincial governments influenced provincial willingness to participate, which in turn affected how the workshops were designed. It also influenced which outcomes were attained: the preceding power dynamics increased the importance of the legitimizing effects of the research report for the provinces. Although legitimizing effects of TDR output in itself are not new, the need for an independent report in our case was a direct consequence of the tense relationships between the parties involved. It is a perfect example of why evaluating TDR according to pre-determined criteria (e.g. 'co-created output') is a poor measure for meaningful evaluation and stresses the importance for evaluators to understand complex contextual factors, such as the political arena in which participants are acting, in order to conduct an adequate impact assessment (Rau et al., 2018). The third benefit we experienced was that our involvement with the project allowed us to conduct better interviews. It made us more sensitive to topics about which to inquire and to carefully probe. In addition, the established rapport with the participants allowed us to discuss impacts that went beyond the direct use of findings and also brought to light impacts that '*are far more intangible but considered just as important by participants*' (Bracken et al., 2014:10). These included affective effects such as a sense of belonging and increased trust between governments, topics that might not have been explored in such depth had we not been aware of the preceding tense relations or if we had not built rapport. Such insights have been suggested as vital for a meaningful and comprehensive evaluation (Hansson and Polk, 2018). Finally, and in line with the previous point, the fourth benefit was that our combined role allowed us to corroborate – or triangulate (Creswell and Miller, 2000) – our interpretations of the project from our facilitator perspective during our evaluation work. This allowed for a more rigorous analysis of different 'impact pathways' that linked the team's capacities to process elements and, finally, to societal effects.

This is not to say the approach was beyond reproach. One shortcoming is its limited potential for capturing unintended effects that go beyond the more obvious 'outcome spaces' of research (Mitchell et al., 2015). Our 'insiders'' view on impacts may have narrowed down our perspective on possible effects and attributable process features. We observe that awareness of this risk and openness to alternative signals are fundamental characteristics of facilitators/evaluators to guard against research bias. In light of this, additional researchers who have not been involved as facilitators (such as A2 and A4 in our case) play a crucial role. Another strategy may be to expand the selection of interviewees to non-participant actors, although then similar difficulties as with decoupled evaluation may jeopardize their commitment to participation. To this, there is no obvious solution.

Furthermore, as facilitators we sought to encourage joint reflection on challenges with the team to formulate collective action. However, it was often difficult to reserve time for reflection, as other matters were perceived as more urgent, an issue also identified by De Wildt-Liesveld et al. (2015). This may also be partly explained by our other role as evaluators. As mentioned previously, to ensure institutional support we were sometimes presented as TDR experts responsible for the scientific quality of the TDR approach. While this appealed to the technocratic culture of their organizations, it may have overshadowed our role as facilitators of transdisciplinary learning: implicitly, the team had outsourced the responsibility for the TDR quality of their project to us. Interestingly, we experienced similar challenges as the team members had faced themselves: just as they had to develop TDR ownership among the project participants, we had to develop the team's ownership over 'our' evaluation. As we could not always compel the team to pause and reflect, it appears we were not completely successful. Much has been written on evaluations that seek to reconcile purposes of learning and impact assessment (Botha et al., 2016; Regeer et al., 2016; Van der Meer and Edelenbos, 2006). It is challenging to combine the two foci because they serve different needs (practitioners want a learn-by-doing approach while managers seek insights into cost efficiency) and require a different evaluation approach. Although the difficulties we encountered did not concern tensions regarding our study design, it did lead to confusion about our role and the externalization of ownership over the evaluation.

Although the issue of linking impact to transdisciplinary research processes is not definitively resolved, our approach implies that combining the roles of facilitator and evaluator results in an evaluation that is better matched to the project under scrutiny. We found these roles are complementary: they allow for in-depth understanding of a TDR project's intricacies and access to practitioners' experiences and their views on the project's impacts, while supporting the TDR team with developing key TDR features in the face of complex internal and external project dynamics. With national and international research funding for solution-oriented research on the rise (e.g. Mazzucato, 2018), TDR popularity is likely to increase. The call for assessing its impact will not decline, nor is it likely that TDR complexity will diminish. This underlines the need to reconcile TDR impact evaluation with promoting 'learning-by-doing' and transdisciplinary capacity building to accommodate TDR's inherent contingencies. The approach presented in this paper may serve as stepping stone for the TDR community to further the conversation on (the impact of) inclusive, reflexive and responsive research.

Funding

The evaluation of the Natuurpact project was commissioned by the PBL Netherlands Environmental Assessment Agency.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the project team and the participants of the Natuurpact project for their commitment to the study, the reviewers for their valuable feedback and Tara White and Deborah Eade for editing.

References

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, 1–17. https://doi.org/10.1093/reseval/rvv025.

Bergmann, M., Brohmann, B., Hoffmann, E., Loibl, M.C., Rehaag, R., Schramm, E., Voss, J.-P., 2005. Quality criteria of transdisciplinary research. A guide for the formative evaluation of research projects. Inst. fuer sozial-oekologische Forsch. 5–75.

Botha, N., Coutts, J., Turner, J.A., White, T., Williams, T., 2016. Evaluating for Learning and Accountability in System Innovation: Incorporating Reflexivity in a Logical Framework.

Bracken, L.J., Bulkeley, H.A., Whitman, G., 2014. Transdisciplinary research: understanding the stakeholder perspective. J. Environ. Plan. Manag. 1–18. https://doi.org/10.1080/09640568.2014.921596.

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

Creswell, J.W., Miller, D.L., 2000. Determining validity in qualitative inquiry. Theory Pract. 39, 124–130. https://doi.org/10.1207/s15430421tip3903_2.

De Jong, S.P.L., Wardenaar, T., Horlings, E., 2016. Exploring the promises of transdisciplinary research: a quantitative study of two climate research programmes. Res. Policy 45, 1397–1409. https://doi.org/10.1016/j.respol.2016.04.008.

De Wildt-Liesveld, R., Bunders, J.F.G., Regeer, B.J., 2015. Governance strategies to enhance the adaptive capacity of niche experiments. Environ. Innov. Soc. Transitions 16, 154–172. https://doi.org/10.1016/j.eist.2015.04.001.

Di Iacovo, F., Moruzzo, R., Rossignoli, C.M., Scarpellini, P., 2016. Measuring the effects of transdisciplinary research: the case of a social farming project. Futures 75, 24–35.

https://doi.org/10.1016/j.futures.2015.10.009.
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.
Hansson, S., Polk, M., 2018. Assessing the impact of transdisciplinary research: the usefulness of relevance, credibility, and legitimacy for understanding the link between process and impact. Res. Eval. 27, 132–144. https://doi.org/10.1093/reseval/rvy004.
Hellström, T., 2015. Formative evaluation at a transdisciplinary research center. In: Polk, M. (Ed.), Co-Producing Knowledge for Sustainable Cities - Joining Forces for Change. Routledge, London, pp. 162–181.
Hoffmann, S., Pohl, C., Hering, J.G., 2017. Exploring transdisciplinary integration within a large research program: empirical lessons from four thematic synthesis processes. Res. Policy 46, 678–692. https://doi.org/10.1016/J.RESPOL.2017.01.004.
Hsieh, H.-F., Shannon, S.E., 2005. Three approaches to qualitative content analysis. Qual. Health Res. 15, 1277–1288. https://doi.org/10.1177/1049732305276687.
Klaassen, P., Verwoerd, L., Kupper, F., Regeer, B.J., 2019. Reflexive monitoring in action as a methodology for learning and enacting responsible research and innovation. in press In: Yaghmaei, E., Van de Poel, I. (Eds.), Assessment of Responsible Innovation: Methods and Practices. Routledge.
Mazzucato, M., 2018. MISSIONS: Mission-oriented Research & Innovation in the European Union. Brussels.
van Economische Zaken, Ministerie, 2013. Natuurpact: ontwikkeling en beheer van natuur in Nederland. Ministerie van Economische Zaken, Den Haag. https://doi.org/10.1007/s13398-014-0173-7.2.
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.
Pohl, C., 2011. What is progress in transdisciplinary research? Futures 43, 618–626. https://doi.org/10.1016/j.futures.2011.03.001.
Pohl, C., Rist, S., Zimmermann, A., Fry, P., Gurung, G.S., Schneider, F., Speranza, C.I., Kiteme, B., Boillat, S., Serrano, E., Hadorn, G.H., Wiesmann, U., 2010. Researchers' roles in knowledge co-production: experience from sustainability research in Kenya, Switzerland, Bolivia and Nepal. Sci. Public Policy 37, 267–281. https://doi.org/10.3152/030234210X496628.
Popa, F., Guillermin, M., 2017. Reflexive methodological pluralism. J. Mix. Methods Res. 11, 19–35. https://doi.org/10.1177/1558689815610250.
Rau, H., Goggins, G., Fahy, F., 2018. From invisibility to impact: recognising the scientific and societal relevance of interdisciplinary sustainability research. Res. Policy 47, 266–276. https://doi.org/10.1016/j.respol.2017.11.005.
Regeer, B.J., Bunders, J.F.G., 2009. Knowledge Co-creation: Interaction Between Science and Society. Advisory Council for Spatial Planning, Nature and the Environment (RMNO), The Hague.
Regeer, B.J., de Wildt-Liesveld, R., van Mierlo, B., Bunders, J.F.G., 2016. Exploring ways to reconcile accountability and learning in the evaluation of niche experiments. Evaluation 22, 6–28. https://doi.org/10.1177/1356389015623659.
Regeer, B.J., Hoes, A.-C., van Amstel-van Saane, M., Caron-Flinterman, F.F., Bunders, J.F.G., 2009. Six guiding principles for evaluating Mode-2 strategies for sustainable development. Am. J. Eval. 30, 515–537. https://doi.org/10.1177/1098214009344618.
Ribeiro, C.D.S., van de Burgwal, L.H.M., Regeer, B.J., 2019. Overcoming challenges for designing and implementing the One Health approach: a systematic review of the literature. One Heal. 7, 1–19. https://doi.org/10.1016/J.ONEHLT.2019.100085.
Roux, D.J., Stirzaker, R.J., Breen, C.M., Lefroy, E.C., Cresswell, H.P., 2010. Framework for participative reflection on the accomplishment of transdisciplinary research programs. Environ. Sci. Policy 13, 733–741. https://doi.org/10.1016/j.envsci.2010.08.002.
Sarkki, S., Heikkinen, H.I., Karjalainen, T.P., 2013. Sensitivity in transdisciplinary projects: a case of reindeer management in Finland. Land Use Policy 34, 183–192. https://doi.org/10.1016/j.landusepol.2013.03.004.
Scholz, R.W., Steiner, G., 2015. The real type and ideal type of transdisciplinary processes: part I—theoretical foundations. Sustain. Sci. 10. https://doi.org/10.1007/s11625-015-0326-4.
Van der Meer, F.-B., Edelenbos, J., 2006. Evaluation in multi-actor policy processes: accountability, learning and Co-operation. Evaluation 12, 201–218. https://doi.org/10.1177/1356389006066972.
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, 325–338. https://doi.org/10.1016/j.evalprogplan.2007.08.002.
Wickson, F., Carew, A.L., Russell, A.W., 2006. Transdisciplinary research: characteristics, quandaries and quality. Futures 38, 1046–1059. https://doi.org/10.1016/j.futures.2006.02.011.
Wiek, A., Talwar, S., O'Shea, M., Robinson, J., 2014. Toward a methodological scheme for capturing societal effects of participatory sustainability research. Res. Eval. 23. https://doi.org/10.1093/reseval/rvt031.
Woltersdorf, L., Lang, P., Döll, P., 2019. How to set up a transdisciplinary research project in Central Asia: description and evaluation. Sustain. Sci. 14, 697–711. https://doi.org/10.1007/s11625-018-0625-7.
Zscheischler, J., Rogga, S., Lange, A., 2018. The success of transdisciplinary research for sustainable land use: individual perceptions and assessments. Sustain. Sci. 13, 1061–1074. https://doi.org/10.1007/s11625-018-0556-3.