fteval JOURNAL
for Research and
Technology Policy Evaluation

APRIL 2022, Vol. 53, pp. 43-56
DOI: 10.22163/fteval.2022.541

DEVELOPING CRITERIA OF SUCCESSFUL PROCESSES IN CO-CREATIVE RESEARCH
A FORMATIVE EVALUATION SCHEME FOR CLIMATE SERVICES

SUSANNE SCHUCK-ZÖLLER, SEBASTIAN BATHIANY, MARKUS DRESSEL, JULIANE EL ZOHBI, ELKE KEUP-THIEL, DIANA RECHID AND MIRKO SUHARI

DOI: 10.22163/fteval.2022.541

ABSTRACT

Climate change and its socio-ecological impacts affect all sectors of society. To tackle the multiple risks of climate change the field of climate services evolved during the last decades. In this scientific field products to be applied in practice are developed in constant interaction between climate service providers and users. To judge the effectiveness of these co-creation endeavours, evaluation is crucial. At present, output and outcome assessments are conducted occasionally in this research field. However, the summative evaluation does not help to adjust the ongoing process of co-creation. Thus, our work focuses on the formative evaluation of co-creative development of science-based climate service products.

As the first step, main characteristics of the product development process were identified empirically. Secondly, we determined the six sub-processes of climate service product development and related process steps. Thirdly, we selected the questions for the formative evaluation relevant to all the sub-processes and process-steps. Then, a literature review delivered the theoretical background for further work and revealed further quality aspects. These aspects from literature were brought together with our results from the empirical work. In the end, we created a new scheme of quality criteria and related assessment questions for the different sub-processes in climate services, based on both, empirical and theoretical work.

As the authors take into account the process of co-production in a real-life case, the criteria and assessment questions proposed are operational and hands-on. The quality aspects refer to the five principles of applicability, theoretical and empirical foundation, professionalism, transparency of processes and the disclosure of preconditions. They are elaborated comprehensively in our study. The resulting formative evaluation scheme is novel in climate service science and practice and useful in improving the co-creation processes in climate services and beyond.

1. INTRODUCTION

Negative impacts of climate change are becoming more apparent. As climate change affects all sectors of society, the cooperation between civil society, political, economic and scientific actors is a key element for the development of climate mitigation and adaptation strategies. For more than fifteen years, the concept of climate services has been developed and regarded as an answer to climate change challenges (Brasseur & Gallardo, 2016). Climate service institutions have been established in many countries.

Climate services are defined as "the transformation of climate-related data — together with other relevant information — into customised products such as projections, forecasts, information, trends, economic analysis, assessments (including technology assessment), counselling on best practices, development and evaluation of solutions and any other service in relation to climate that may be of use for the society at large." (European Commission, 2015, p.10)

Researchers from very different research fields, such as physics, meteorology, biology, agricultural research, social science, economics and others are working together in climate services. They apply the participatory and interactive modes of scientific knowledge production in a highly interdisciplinary research agenda (European Commission, 2015). This agenda comprises approaches of collaboration that includes practitioners. In order to enhance the adaptive capacity of society to climate change, knowledge integration is crucial. Researchers of all relevant fields, partners from practice, and users of climate services need to share their knowledge and learn from each other.

We are aware of multiple and diverse approaches of such interactive modes of scientific knowledge production (for an overview see Newig et al., 2019; Bremer & Meisch, 2018; Brinkmann et al., 2015). Mauser et al. (2013) proposed "co-creation" as an overarching term for the different phases of transdisciplinary research processes in the field of Earth sciences. We follow this proposition and use the terms "co-creation" and "co-creative" research synonymously.

Who might be involved in co-creative projects? On one hand, there are interdisciplinary experts from different research fields, here called "researchers", who are part of a project consortium. They work together in a "scientific team" and form the "scientific party" of the co-creation partnership. The institutions, the researchers belong to, are called climate service "providers". They aim for appropriate information, products and tools to support decision-making in terms of climate change and

impacts. If the researchers need scientific knowledge or data to work with, they might look for "scientific cooperation partners", who are not yet part of the project consortium. They might also involve experts from practice, called "practitioners", to benefit from their experience. They are very often "users" of climate services, who are highly interested in the future products and, therefore, form the "party of practitioners". Beyond the party of practitioners there might be future "external users" as well.

To make co-creation processes effective, their quality and evaluation have been studied for years. In the literature, a broad range of exemplary evaluations assessing whole projects can be found – e.g. in environmental research (Jahn & Keil, 2015), hydrology (Maag et al., 2018) or especially related to weather and climate services (Wall et al., 2017). More in-depth ex post evaluations of single products are still rare, above all in climate services (Körner & Lieberum, 2014; Haße & Kind, 2019). Belcher et al. (2021) only recently have contributed with a quality assessment framework for social innovation impacts in co-creative research.

As evaluation research is increasingly indicating a relation between good co-creation processes and their effectiveness and success (Lux et al., 2020; Maag et al., 2018; Wolf et al., 2013), the adjustment and improvement of these processes become a key issue, above all in climate services, reacting to the urging problem of climate change. Formative evaluation delivers the chance to adapt the processes over the course of an ongoing project and restructure it.

In consideration of the fact that research processes in co-creation are crucial for societal impacts (Maag et al., 2018) the project NorQuATrans (Normativity, Objectivity and Quality of Transdisciplinary Processes) was initiated and implemented[1] in the Helmholtz-Institute HICSS where the scientists from the Climate Service Center Germany (GERICS) and Hamburg University cooperate. The objective of the NorQuATrans is – amongst others – to empirically analyse challenges and limitations of climate services and to examine closely the quality of the co-creation processes. As accompanying research to other projects, the NorQuATrans aims at developing a concept for a formative evaluation scheme of co-creation in climate services as well as a suitable set of quality criteria and methods.

The work described here uses the case study approach by accompanying a co-creative research project in the field of agriculture. The ADAPTER (ADAPT tERrestrial systems) project[2] delivers the innovative simulation-based products to support optimal adaptation to both short-term weather variability and weather extremes, as well as to long-term regional climate change. To do so, ADAPTER involves the practitioners from different areas of agricultural practice. The co-creation processes between ADAPTER scientists and practitioners were used as a showcase to analyse their quality and develop possible evaluation criteria. Hence, this contribution does not present a concrete evaluation activity. It describes instead the development of a methodological approach, leading to a formative evaluation framework aiming at the provision of criteria and related assessment questions for upcoming process evaluations.

To investigate this in more detail, we focus on the joint production of single climate service products. Within the research projects such single product development endeavours are usually just one part of the whole project. Still, our work described here focuses on a consequent close-up view of one single product development process.

The two main questions we aim to answer in this paper are:

- How can climate service providers ensure quality and success of the co-creative processes?
- What are appropriate questions to reflect and evaluate the ongoing processes throughout different project phases?

This is the first study to have developed a formative evaluation scheme dedicated to the field of climate services. Related to the process of climate service product development we apply a method consisting of three steps: a) developing evaluation questions based on newly identified sub-processes of the whole product development endeavour; b) collecting criteria and indicators from a literature overview; c) bringing these deliverables together and adjoining five main principles of co-creative research in climate services.

On this basis, we introduce a process-oriented and formative evaluation scheme encompassing evaluation criteria and assessment questions. The aim is to raise awareness of the importance of formative evaluation amongst researchers or managers in scientific co-creation. It is intended to trigger self-reflection and help the co-creating parties to reflect on the collaboration processes in climate services. Therefore, we adopt and extend existing evaluation schemes for this particular research field.

In the second section, we will present the three-step methodological approach. The results will be presented in the third section, including the process-oriented, formative evaluation scheme. In the fourth section, the results are summarised and discussed in context of the existing literature followed by the limitations of the methodology. Finally, we provide an outlook on open questions and possible further activities in terms of formative evaluation research in the fifth section.

2. METHODOLOGY

The work described here was performed as an accompanying research to the project ADAPTER. Within this project researchers are in dialogue with key agricultural practitioners in Germany. The practitioners' party includes, for example, the farmers' associations on the local level as well as seed grower companies. The collaboration between science and practice within the project has been characterised by intense and regular correspondence by email, by phone and in person. In addition, several workshops took place and at the moment of writing the first product is nearly finished and in the last phase of testing.

THE EMPIRICAL PART

Together with the scientific party in ADAPTER, the NorQuATrans scienific team looked back on the past project phases and identified the steps and sub-steps of their co-creative product development. As a preliminary result, this empirical work firstly showed a workflow containing steps in different grades of details. This workflow – evolved from the ADAPTER project – consisted of a large amount of complexity and details. Therefore, additional experienced climate service researchers were gathered for jointly re-condensing the workflow. As a result, a list of six

1 https://www.hicss-hamburg.de/projects/NorQuATrans/index.php.en
2 www.adapter-projekt.de

sub-processes is summarised. This final list was again revised after the discussion with peer groups (scientists from climate services) and made more coherent.

Then, the scientific parties of ADAPTER and NorQuATrans reflected together how high quality can be ensured during the different process-steps and sub-processes. We adjusted the related evaluation indicators and questions. Particularly in those cases where collaboration had not worked perfectly, we could well identify missing process steps and deduce respective evaluation questions. Intentionally, the six sub-processes and the related questions were developed only from experience.

THE THEORETICAL PART

In parallel, but properly separated, the scientific team of NorQuATrans reviewed the existing literature on formative evaluation. In total, 25 articles from the peer-reviewed journals were identified, which illustrate quality criteria and indicators related to co-creative research. These publications were then analysed for their content and only those which complied with the requirements below were selected:

- focus on the processes – those which focused on output or outcome (OECD, 2002) were dismissed, and
- explicit discussion of co-creative research.

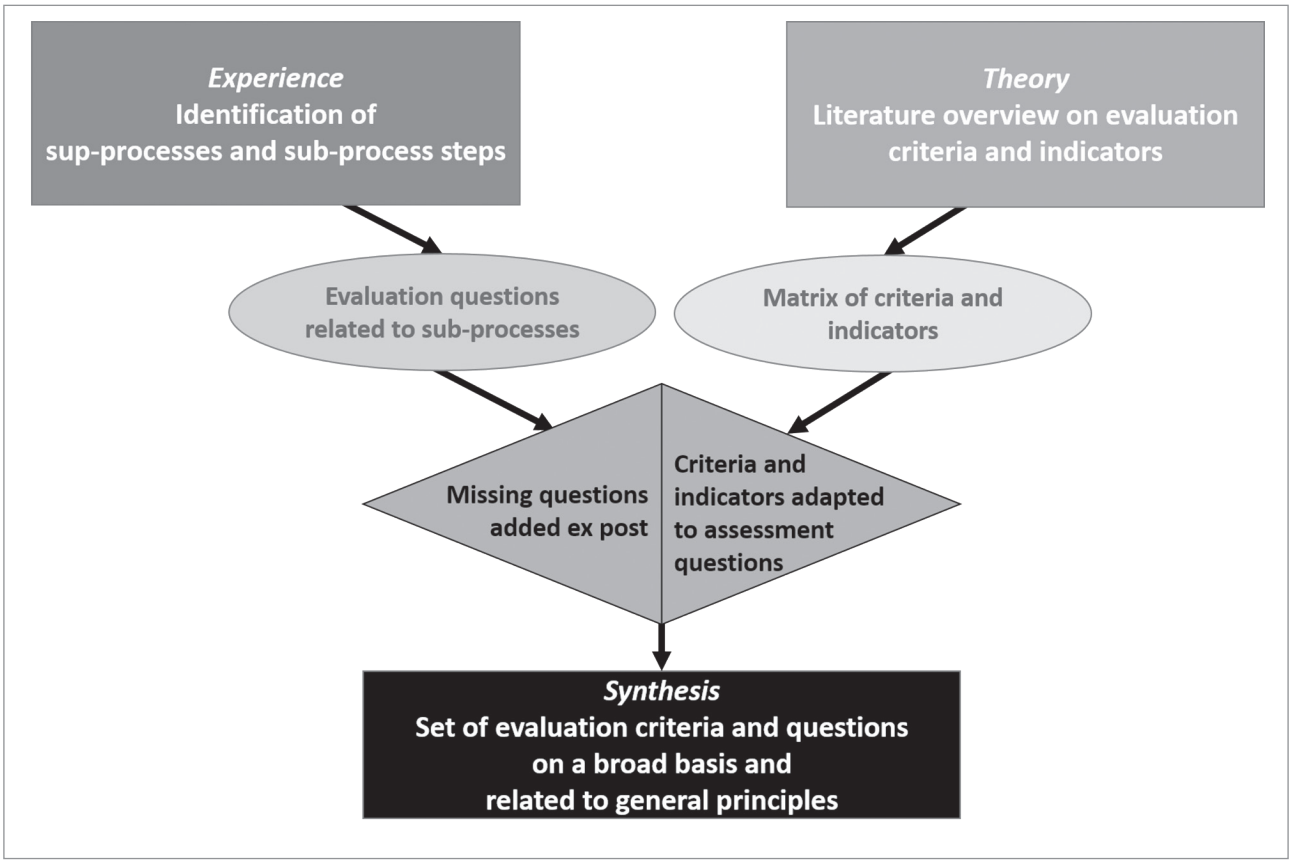
The selected 16 articles presented concrete quality criteria of co-creative research processes – of course with different grades of details – and included different viewpoints. Most of the criteria and indicators were deduced from real cases in practice (e.g. Maag et al., 2018; Bergmann et al., 2005; Wall et al., 2017). A few of the publications analyse huge numbers of projects (e.g. Lux et al., 2020; Newig et al., 2019). No distinction was made during the review whether process criteria and indicators had been identified ex post or during the respective processes. The ADAPTER project team was not involved in this theoretical work.

SYNTHESIS

In the end, the process-related evaluation questions gained by practical experience in climate services and the matrix of quality criteria and indicators from the literature review were compared and synthesized (Fig. 1). Many of the evaluation questions could be directly related to quality criteria and indicators from the theoretical discussion.

If criteria from the literature (here labelled as "theory") were not yet covered by assessment questions from our empirical study ("experience"), suitable questions were added. The criteria and indicators from theory were reformulated if they were not tangible enough to cover our assessment question.

Figure 1: Synthesis of the empirical insights from experiences in the ADAPTER project and theory resulting in a set of evaluation criteria and questions for formative evaluation



The final list of evaluation criteria and questions was sorted and the criteria were aligned with more general categories (here called "principles"), which were also derived from the literature overview.

The two scientific teams a) the ADAPTER co-creative team and b) the NorQuATrans team of accompanying research had different tasks in the course of this development work: Namely the identification of co-creative processes (paragraph a) substantially originated from ADAPTER, the literature review and synthesis work were carried out by the NorQuA-Trans team. A common discussion did not happen until after the synthesis draft was developed and had to be validated.

The development work, presented here, was performed by scientific parties. To pursue the idea of co-creation consequently, practitioners get the chance to contribute and are asked for ex post validation (see section 5).

3. RESULTS

In the following, the results derived from the exercises described above are presented.

SIX SUB-PROCESSES FORM THE CO-CREATION PROCESS IN CLIMATE SERVICES

The empirical work in ADAPTER allowed for the identification of six sub-processes in co-creation (Fig. 2). These sub-processes can be generalised in the direction of a common co-creative process of climate service product development, as they underwent different iteration steps as described above.

RESULTS FROM THEORY AND PRACTICE COMPLEMENT EACH OTHER

The assessment questions gained from the six sub-processes and related process steps were brought together with process evaluation criteria identified through the literature review. This synthesis made clear that the aspects of good quality from theory or practice show many consistencies but complement each other as well. Thus, the collection of criteria resulting from the theoretical review showed gaps in our set of empirical questions. And vice versa, we missed criteria in literature that are seen crucial in practice.

FIVE MAIN PRINCIPLES AND SEVENTEEN QUALITY CRITERIA

To better structure the criteria and related assessment questions, we looked into the literature to find more general terms for categorisation. However, the terms for categories tend to differ. For instance, it is common to see "principles" (Norström et al., 2020; Krause & Schupp, 2019; Belcher et al., 2016) and "dimensions" (Jahn & Keil, 2015), alongside the more normative "recommendations" (Lux et al., 2020; Nagy et al., 2020) or even "elements of success" (Garard et al., 2018) and "areas of improvement" (Jagannathan et al., 2019). We adopted the term "principles", which means here "characteristic principles of co-creation".

Five principles were defined: (1) *common ground*, (2) *transparency*, (3) *professionalism*, (4) *enhancement of applicability* as well as (5) *theoretical and empirical foundation* (table 1). All principles are of equal importance and therefore, the sequence is arbitrary. The second column of table 1 contains the seventeen criteria that we propose. They show what is behind the principles and make them more tangible. Some have already been defined in previous studies. However, we have in some cases changed the wording to make the criterion more suitable for practical use. The four criteria in bold font are newly added by us. As the criteria are overlapping in a few cases, the assignment presented here is not the only sensible possibility.

The corresponding assessment questions will be presented later in the tables 2 to 8. Often one criterion covers different questions, therefore the criteria show up several times.

Figure 2: Six sub-processes of co-creating a climate service product

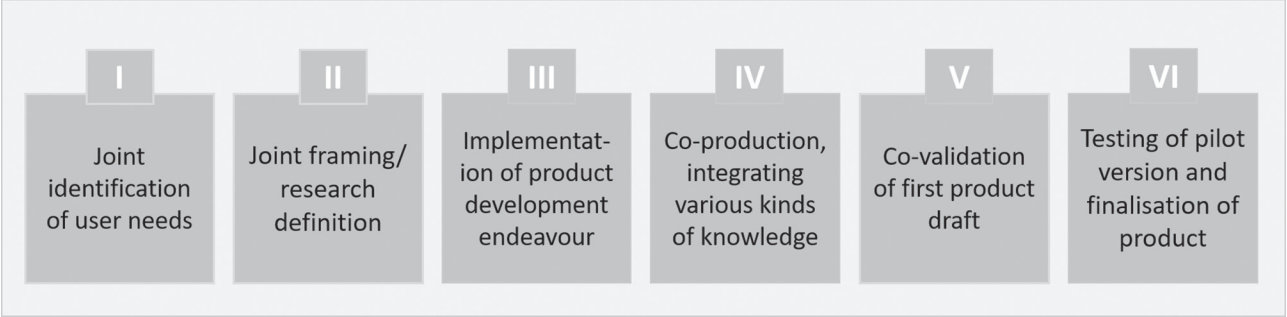


Table 1: Principles and criteria for co-creation processes (by the authors newly added criteria in bold font)

Principles	Criteria	Criterion inspired by	Application in all sub-processes
Common Ground	Equitable opportunities to participate	Belcher et al., 2016; Norström et al., 2020	X
	Trust building	Schuck-Zöller et al., 2018; Norström et al., 2020	
	Joint problem ownership	Schuck-Zöller et al., 2018	
Transparency	Overall development process documentation	Schuck-Zöller et al., 2018	X
	Open exchange on all preconditions		
Professionalism	Clear management and integration concept	Bergmann et al., 2005	X
	Accountability	Krause & Schupp, 2019	X
Enhancement of applicability	Ongoing monitoring and reflection	Bergmann et al., 2005; Belcher et al., 2016	X
	Knowledge integration	Newig et al., 2019; Lux et al., 2020	X
	Benefit of diversity	Maag et al., 2018; Norström et al., 2020	
	Sustainable use		
Theoretical and empirical foundation	Clear research problem definition	Bergmann et al., 2005	
	Scientific soundness	Jahn & Keil, 2015	X
	Use of state-of-the-art knowledge	Belcher et al., 2016; Maag et al., 2018	X
	Scientific peer reviews		
	Coherence	Schuck-Zöller et al. 2018	X
	Impact on science	div.	

Maag et al. (2018) are the first to introduce the term **common ground**, which extends the generally identified need for the category "trust" (Norström et al., 2020; Krause & Schupp, 2019; Schuck-Zöller et al., 2018). Comparing to the term "trust" the notion of *common ground* is more precise to describe communication at eye level, "openness as an attitude" (Garard et al., 2018) and "mutual understanding" (Maag et al., 2018). We understand that this aspect covers the different challenges of communication and collaboration. The related criteria can already be found in Norström et al. (2020), Schuck-Zöller et al. (2018) and Belcher at al. (2016).

Transparency in all collaboration issues is quite generally demanded (i.e. Garard et al., 2018; Schuck-Zöller et al., 2018; Belcher et al., 2016; Jahn & Keil, 2015; Bergmann et al., 2005). This aspect seems a key issue in looking on processes. Thus, we made it a principle in our scheme. *Overall process documentation* is firstly recommended by Schuck-Zöller et al. (2018), the *open exchange on all preconditions* is for the first time taken into account in this paper here. Lux at al. (2019) already pointed out on the clarification of the roles, which is one aspect of the criterion proposed here.

The importance of "good management" is addressed in some contributions (i.e. Wooten et al., 2014; Bergmann et al. 2005). Bergmann et al. (2005) describe in detail how to realise it. Most of the studies agree explicitly that good management is crucial (Schuck-Zöller et al., 2018) or implicitly by noting with other similar criteria (i.e. Wall et al., 2017; Jahn & Keil, 2015; Garard et al., 2018). Still, the importance is not always stressed sufficiently. We decided, that **professionalism** is an appropriate principle to underline the overarching character of this quality aspect and to prevent it from being neglected in practice. Krause & Schupp (2019) point to an aspect which might be underestimated so far: *accountability* of all co-creation participants, above all, of the managing team leads to mutual trust. We see it as a very important aspect of *professionalism*.

The main objective of co-creation endeavours is without any doubt the applicability of research results. While this can be reviewed best ex post, we demand the **enhancement of applicability** for the co-creation process. Very common here is the demand for *ongoing monitoring and reflection* (i.e. Bergmann et al., 2005; Belcher et al., 2016) and *knowledge integration* (i.e. Newig et al., 2019; Lux et al., 2020), whereas the *benefit of diversity* is not highlighted very often (except for Maag et al., 2018; Norström et al., 2020). *Sustainable use* is made a criterion of climate service product development here for the first time. We want to stress how essential it is to provide for long-lasting use of products.

The **theoretical and empirical foundation** (Belcher et al., 2016) alludes to the soundness of both, research as well as experiential knowledge and all integration activities. It is strongly related to *professionalism* in research and the facilitation of *knowledge integration*. We follow several papers with the aligned criteria (Jahn & Keil, 2015; Belcher et al., 2016; Maag et al., 2018; Schuck-Zöller et al., 2018). The special demand for *scientific peer reviews* is a special concern of ours, based on the experience that thorough discussions with scientific peers and critical reviews of the ongoing product development often get out of sight. *Impact on science* is adopted from traditional evaluation of basic research. We apply it to co-creative endeavours as well.

Which gaps did we identify by allocating the assessment questions from experience with process evaluation criteria from the literature review? We missed in literature, for instance, informative criteria related to *theoretical and empirical foundation*. This principle was rarely made an issus in terms of co-creation processes quality. Furthermore, the criterion *open exchange on all preconditions* to the upcoming processes was not addressed in this general meaning. In the end, four criteria seemed to be missing and essential enough to be added by us (in bold font, see table 1).

FINDINGS CONDENSED IN A FORMATIVE EVALUATION SCHEME

The synthesis described above leads to a scheme of evaluation criteria and questions, aligned to the different principles of co-creation covering the process from the first idea of a climate service product to the finalised version. The collection of questions is broad and the questions are general enough to be usable for different kinds of products.

How to use the scheme? Our aim is to provide a set of reasonable questions to researchers and co-creation facilitators – regardless whether these two tasks have to be taken over by one person or not – who want to monitor and evaluate a co-creation process. The scheme can either be used by the co-creating researchers or the facilitators of the co-creation endeavour (self-evaluation), by colleagues from their institution, who are not involved in the product development (in-house evaluation), or by external evaluation specialists (external evaluation) who might look neutrally on the development process.

The type of questions and their addressees differ: Some are to be answered by the co-creation facilitating team themselves. Others allow for learning about the perception of either all participants or just the practitioners or researchers. The questions are recommended to be operationalised by a five-point Likert-scale, or as simple as a binary question with Yes/No-answer. The possibility to choose between different gradations can increase the motivation to answer the questions and allows for comparing the perception of different participating parties. If there are time constraints, we recommend not to use open questions, as they require more resources for answering and interpreting. However, open questions reveal more in-depth information. In general, we decided to design the set of questions incoherently in form (i.e. open or closed): The change of different survey techniques may give a more nuanced picture and make the participation more interesting.

In general, the sequence of the questions in the tables follow the course of the co-creation process.

Some of the questions are quite similar to each other on purpose. Different principles and criteria mirror different aspects of quality and might suit for triangulation. The question *Have all practitioners been included in the previous step of reflection?*, for example, focuses on the possibility to participate, whereas *How many practitioners have taken part in the previous step of reflection?* concentrates on the readiness of practitioners to appreciate this possibility. But in general, some of the alignments are propositions and not compelling. The challenging of these alignments could be a first step to the self-reflection we aim to trigger.

The assessment questions in tables 2 to 8 define the underlying indicators. These indicators are presented in bold font. As we want to present a scheme sufficiently general for different kinds of products, some questions might not fit the objective of that very special product development endeavour. We invite the project leaders to select those questions that apply to their development process and state reasons why others do not.

Some of the evaluation questions came up in every sub-process in similar shape, due to the demand for ongoing reflection. To simplify the scheme, we extracted these evaluation criteria and the related 24 evaluation questions and generalised them. We propose to use them in every sub-process (table 2).

Table 2: SUB-PROCESSES I to VI **– Common criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Common Ground	Equitable opportunities to participate	*Have all participants perceived **balanced opportunities** to participate?*
		*Has a **balanced influence** between all project partners (from science and practice) been assured in this sub-process?*
		*Have **all practitioners been involved in** the previous step of reflection?*
Enhancement of applicability	Knowledge integration	***How many practitioners participated** in the previous step of reflection?*
Theoretical and empirical foundation	Use of state-of-the-art knowledge	*Have the a) **recent data base** and b) **recent analyse methods been used** in the previous research step?*
	Scientific soundness	*Have the tasks in the respective sub-process been executed in a **scientifically sound** manner?*
	Coherence	*Have **contradicting viewpoints** of single project partners been a) **handled constructively** and b) **made coherent**?*
Professionalism	Accountability	*Have the **management methods** in the respective subprocess been **applied appropriately**?*

Transparency	Overall development process documentation	*Has the respective **sub-process** been* *a) **transparent** to all participants and* *b) **properly** documented?*
Enhancement of applicability	Ongoing monitoring and reflection	*Has the **format for reflection** during the upcoming step been chosen **appropriately**?*
	Knowledge integration	*To which extent have in the previous reflection step been **considered*** *a) the original **contributions of knowledge** (e.g. local, experiential) **from practitioners** and* *b) the **feedback of practitioners**?*
		*Have **consequences been taken from the feedback** gained in the former reflection step?*
		*Have all **participating parties been satisfied** with the former reflection step concerning* *a) format,* *b) method and* *c) result?*
		*Has the **application of integration formats and methods** been **successful**?*
Professionalism	Clear management and integration concept	*Have the process steps of the past sub-process been **executed in line with the different plans**, i.e.* *a) the time schedule in detail (mile stones),* *b) the integration concept and* *c) the documentation concept?*
		*Do parts of the **plans need adaptation**?*

These questions should be taken into account in each of the six sub-processes discussed here. Listing different aspects of the questions (e.g. a, b, c, etc.) might help to realise different facets and/or illustrate the criterion.

All other criteria and evaluation questions are presented in the context of the different sub-processes (fig. 2) and follow the workflow within. Sub-process I starts with the idea of a climate service product be it expressed by practitioners, researchers or funding institutions. The most important steps are related to the identification and recruitment of key experts from practice (practitioners) and the specification of their needs. Therefore, the questions mainly focus on these aspects (table 3).

Table 3: SUB-PROCESS I – Joint identification of user needs – **Criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Enhancement of applicability	Benefit of diversity	*Has the **selection of practitioners** been conducted in a **systematic way** concerning the project content and goals of co-creation?*
	Ongoing monitoring and reflection	*Has the **selection process been reflected** within the project consortium?*
	Benefit of diversity	*Are the **targeted user groups appropriately represented** by the selected practice partners?*
		*Has the **analysis of user needs** been executed **methodologically sound**, i.e.* *a) open-ended,* *b) supported by balanced and appropriate communication and/or* *c) providing balanced opportunities to utter needs?*
	Ongoing monitoring and reflection	*Has the result of the **analysis of user needs been shared** with the practitioners and commonly reflected?*
Common Ground	Trust building	*How far have **trust and motivation been established** during the contact phase with the practitioners?*
		***How many of the desired practitioners** could successfully **be recruited**?*

The eligible disclosure of *all preconditions* for the co-creation endeavour and their *open exchange* follows in sub-process II (table 4). Several questions are to illuminate the different aspects of this newly introduced criterion. Assessing *scientific soundness* in sub-process II leads to a further reflection on the feasibility and methodological limits. To finalise the sub-process, we propose to look back on the *knowledge integration* in general: *How well does the scientific research question cover the needs from practice?*

Table 4: SUB-PROCESS II – Joint framing/research definition – **Criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Transparency	Open exchange on all preconditions	*To what extent heave **general preconditions**, such as* *a) mutual expectations,* *b) potential benefits and* *c) potential risks* *been **shared between all participants**?*
		*Have all **formal or external preconditions** for a co-creation endeavour been **shared between all participants**, i.e.* *a) readiness for open communication versus restriction as for strategic or competition issues,* *b) timely resources on all parties' sides and/or* *c) financial resources and conditions?*
Professionalism	Clear management and integration concept	Have ***formal or external conditions been taken into account*** for the management and integration concept?
Theoretical and empirical foundation	Use of state-of-the-art knowledge	*Has the **scientific feasibility** been proven, i.e.* *a) are appropriate climate (model)data or model ensembles available?* *b) have scientific state-of-the-art methods already been developed?* *c) will scientific state-of-the-art methods be able to be applied?* *d) have scientific cooperation partners – if needed – been found?* *e) could identified gaps be filled in?*
Transparency	Open exchange on all preconditions	*Has the **proof of scientific feasibility** as well as the potentials and limitations of research methods (i.e. bandwidths in climate simulations) been **shared*** *a) within the scientific team,* *b) with project partners and* *c) with practitioners?*
Common Ground	Trust building	Have ***all practitioners been included*** in the discussion on ***scientific methodology***?
Enhancement of applicability	Ongoing monitoring and reflection	*To which extent have **practitioners** been able to **accept limits of research** methods or other external conditions?*
Theoretical and empirical foundation	Clear research problem definition	*Have* *a) temporal and* *spatial **scales of the scientific answers** aimed for been clearly **defined**?*
	Scientific peer reviews	*Have common **discussions on the formulation of the research question with scientific peers**, i.e. colleagues taken place?*
Enhancement of applicability	Ongoing monitoring and reflection	*Has the **research question** been* *a) thoroughly **discussed with practitioners** and* ***a) formulated jointly**?*
Common Ground	Joint problem ownership	*To which extent is the research question identified with by* *a) practitioners and* *b) scientists?*
Theoretical and empirical foundation	Knowledge integration	*How well does the **scientific research question cover the needs** from practice?*

The most distinctive sub-process is the third one (table 5). The facilitators of the co-creation endeavour have to set up fair and realistic concepts for any sort of management and knowledge integration and, thus, determine the character of the further co-creation. The researchers have to contribute with detailed descriptions of their tasks and the identification of related milestones. Many of the questions and criteria refer to the principle of *professionalism*. Key for this rather challenging sub-process is to include practitioners and all those participating in the decision-making process that will shape the upcoming co-creation.

Table 5: SUB-PROCESS III – Implementation of product development endeavour – **Criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Professionalism	Clear management and integration concept	*Has a **knowledge integration** concept been established describing* *a) internal communication,* *a) regular reflection after every sub-process,* *b) how the integration of different kinds of knowledge can be supported and* *c) how different methods of co-creative research can be reflected?*
		*Has a **management plan** been set up containing* *a) a time schedule in detail (mile stones) and* *b) evaluation criteria?*
		*Does the **timeframe** of the project **meet all project partners**'* *a) needs,* *b) constraints and* *c) goals?*
		*To which degree does the planning of the product development enable to **react to the results of the different iteration** processes?*
Transparency	Overall development process documentation	*Has a **concept for the documentation** of process steps been established?*
Enhancement of applicability	Ongoing monitoring and reflection	*Have the **scientific team and the practitioners agreed upon** the different **concepts**, i.e.* *a) the time schedule in detail (mile stones),* *b) the integration concept,* *c) the documentation concept,* *d) an external communication concept of outputs and/or* *e) the evaluation criteria?*
Professionalism	Clear management and integration concept	*How **realistic** is the **product development schedule** in general, i.e. including phases of internal communication, reflection and iteration?*
		*Have the different steps of the **scientific process** been **planned thoroughly**, i.e. related to external preparative work?*
Common Ground	Equitable opportunities to participate	*To which extent are the* *a) researchers and* *b) practitioners* *s**atisfied with the level of engagement**?*

Sub-process IV (table 6) is the most complex one and will take the most time. It entails the process of co-producing research results and often takes several months to even years. Though *knowledge integration* is an overarching task of the whole product development process, it is a key aspect here to assure the *applicability* of the climate service product. Therefore, we explicitly stress its importance. Sub-process IV leads to a first draft of the product.

Table 6: SUB-PROCESS IV – Co-production, integrating various kinds of knowledge – **Criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Theoretical and empirical foundation	Coherence	*To which extent has it been possible to combine **scientific excellence with the aim of solving real-world problems**?*
	Scientific peer reviews	*How far have the different product development steps and their results been **discussed with peers from science**, i.e. by presentations on scientific meetings/conferences?*
	Coherence	*To which extent are the a) researchers and b) practitioners **satisfied with the joint research** process?*
Enhancement of applicability	Ongoing monitoring and reflection	*Has the the **first draft of the product been reflected** jointly with the researchers and practitioners?*
	Knowledge integration	*To which extent has the development team succeeded in **meeting the problems and objections resulting from the common reflection** of the first draft of the product?*

This first product draft has to undergo a thorough testing by the target group (sub-process V, table 7). The conditions for this validation steps have to be designed close to real ones. Consideration should be given to expanding the group of test users to include external potential users. A validation by peers from science is recommended as well. After the different revisions, a pilot version of the product is created.

Sub-process VI (testing of pilot version and finalisation of product) shows similar steps (table 8) like sub-process V: After tests of the pilot version by different user groups, the pilot is revised and brought to application. The first application phase delivers the chance of last revisions before the product will be finalised. In this phase, it is crucial to ensure sustainability by providing an easy-to-use manual. A long-term support might further enhance the chance of *sustainable use*.

Table 7: SUB-PROCESS V – Co-validation of first product draft – **Criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Enhancement of applicability	Ongoing monitoring and reflection	*Has the **format for the different steps of co-validation been chosen appropriately**?*
		*Is the format for the different steps of **co-validation of the first product draft close to reality**?*
	Knowledge integration	*Is the **first product draft easy to use**?*
	Ongoing monitoring and reflection	*Has **the first product draft** been a) **tested**, b) **revised** and c) **finalised** **with practitioners**?*
Theoretical and empirical foundation	Scientific soundness	*Has the **first product draft** been **validated by peers** from science and **revised** accordingly?*
Enhancement of applicability	Knowledge integration	*Has the project team succeeded in meeting the **problems and objections** resulting from **the co-validation of the first product draft**?*

Table 8: SUB-PROCESS VI – Testing of pilot version and finalisation of product – **Criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Enhancement of applicability	Ongoing monitoring and reflection	*Is the format for the different steps of **co-validation of the pilot version close to reality**?*
	Benefit of diversity	*Has an appropriate point at which **external target audiences** enter been made an issue?*
	Sustainable use	*Have the needs a) for a **manual and** b) for a **long-term support concept** been **considered**?*
		*Has a l**ong-term support concept** – if needed – been **assured**?*
		*Is there **staff provided to do a long-term support**?*
	Ongoing monitoring and reflection	*Has the **manual** – if needed – been a) **tested**, b) **revised and** c) **finalised**?*
		*Has the **service concept** – if needed – been a) **proven**, b) **revised and** c) **finalised**?*
Professionalism	Accountability	*Is the **pilot version easy to use**?*
Theoretical and empirical foundation	Scientific soundness	*Is the **group to test the pilot version big** and **diverse** enough?*
		*Has the **pilot version** been **tested within a scientific peer group**?*
	Impact on science	*To what extent have the **findings** of the product development endeavour and the **research results contributed** to **the scientific community**?*
Enhancement of applicability	Ongoing monitoring and reflection	*Has the **pilot version** been a) **tested**, b) **revised and** c) **finalised** **with practitioners**?*
	Knowledge integration	*Has the project team succeeded **in meeting the problems and objections** resulting from **the co-validation of the pilot version**?*

It is obvious which criteria are the most important ones: *Ongoing monitoring and reflection* and *knowledge integration* are named most frequently in our tables. They are to be assessed by questions applicable in the same shape over all sub-processes but, in addition, scrutinised by more special questions and indicators during each single sub-process.

4. SUMMARY AND REFLECTIONS

The projects NorQuATrans and ADAPTER identify six sub-processes of climate service product development. Five principles of co-creation as well as seventeen criteria allow for assessing the quality of co-creative development processes. By covering the quality criteria by assessment questions and indicators, easy-to-use application is provided. A framework is presented for climate service researchers, managers and other participating parties to thoroughly reflect on.

We show that all sub-processes of the product development workflow can be addressed by a specific set of evaluation questions and underlying indicators.

We started with an experience-based analysis and then – like Maag et al. (2018), Wall et al. (2017) and Belcher et al. (2016) did – widened the perspective by a literature review that delivered a more general point of view. The synthesis of both, empirical analysis and theoretical background leads to criteria and indicators, which target on very concrete product development processes and should be applicable in practice.

The case study approach is, as well, used by many of the forerunners, like Bergmann et al. (2005), Wall et al. (2017), Maag et al. (2018). We proceeded similarly to Maag et al. (2018), but identified only six sub-processes ending with the finished product, i.e. excluding implementation and dissemination activities. Like Jahn & Keil (2015), we revealed rather concrete aspects. However, in contrast to most of the existing literature, we only investigated one specific development process in detail.

We experienced that the merging of the empirical development work with the discussions in literature generated particularly interesting results by identifying gaps in our set of empirical questions and vice versa. For example, the quality of iteration and reflection steps of both co-creation parties did not seem to be satisfyingly defined so far. Recommendations were missing, how to proceed in details. Thus, we integrated numerous related assessment questions in every single sub-process to stress the need for repeated common reflection.

Our study aims at the development of a methodological approach and leads to a formative evaluation framework. This is the first time that such a scheme is developed for the dedicated use in the field of climate services.

Especially the principle *of theoretical and empirical foundation* is well elaborated by six different criteria – two of them firstly introduced in this context. These six criteria are not only about using state-of-the-art methodology, but about sharing experiences with the scientific community and inspiring other co-creative projects. The fact, that scientific quality is rather rarely made an issue in literature, might originate from the focus on the co-creation dialogue process phases in previous publications.

What we also introduce for the first time, is the highlighting of different kinds of *preconditions* and the trial to grasp them at the beginning of the development process. Furthermore, for the last sub-process the idea of *sustainable use* is emphasized and backed by assessment questions.

The evaluation framework aims at enhancing the readiness for formative evaluation in co-creation processes. This might be an external, an internal or a self-evaluation. Therefore, we do not see a problem or role conflict in having involved the co-creative scientists in this framework development because co-creative researchers have to cover both roles in self-evaluation activities, as well.

For time constraints, the point of view of practitioners could not yet be directly incorporated into the work described. Still, the co-creative scientists contributed with their experience from practitioner dialogues. Thus, the practitioners' view was represented indirectly. A direct reaction by practice parties to the criteria defined is pending.

One could argue that a single project in the focus impedes generalisation. We tried to meet this challenge by repeatedly discussing the development stage with peers from climate services and comparing the experiences from ADAPTER with other product development processes. Therefore, we are convinced that the six sub-processes allow for generalisation in the direction of a common co-creative process of climate service product development. However, future projects of climate services are to validate the workflow and adjust as well as generalise it. A further step in the direction of generalisation will be the validation of criteria by surveys (see section 5). As the literature review covers all fields of co-creative research, it does not only deliver a theoretical confirmation of the empirical questions and underlying indicators, but also enhances their potential to be transferred to other transdisciplinary research areas.

As we present a framework sufficiently general for different kinds of products, some questions might not fit the objective of every special product development endeavour. We invite the project leaders to select those questions that apply to their development process and perhaps add new criteria and assessment questions. Furthermore, before a new development endeavour is initiated, the set of quality criteria should be discussed. If criteria are to be left out, the reasons should be stated and new criteria could be argued for. Thus, the set of questions is adjustable to the objective of the specific product development endeavour.

5. OUTLOOK

To integrate practitioners' view, we are going to have key criteria of this scheme validated by experts from practice. For this purpose, an empirical study is being prepared in NorQuATrans. Hence, the criteria presented here will be further reviewed by experts from practice beyond the ADAPTER project and the agricultural sector. Thus, we will gain more general information on practitioners' understanding of criteria and their priorities. An additional survey with scientists of different fields of co-creation would further enhance the potential for generalisation. Still missing is an application test of the whole evaluation scheme. This has to be performed, once the scheme will be further validated.

Another still widely open field is the issue of "co-evaluation" (Lux et al., 2020) of co-creation processes and beyond. In general, this aspect has not yet been sufficiently studied – at least in climate services, except for Restrepo et al. (2020). In applying this framework, an open discussion on the evaluation concepts, results and possible re-adjustments should be performed by all participants of the co-creation endeavour from the very beginning. Thus, the scheme could ensure *transparency* for all involved actors and throughout all co-creation phases. A consequent disclosure of all perspectives in assessing quality could make a difference for future work in this respect.

Our framework can be expanded into guidelines for formative evaluation in the future. A comprehensive manual on formative evaluation in climate services is to be generated and presented to research organisations and communities of transdisciplinarity and climate research. Finally, the results can benefit other co-creative research fields. We assume that the scheme for formative evaluation presented here, as well as the resulting guidelines, might be transferable – at least in parts – to other fields of co-creative research.

ACKNOWLEGEMENTS

This work was conducted and financed within the framework of the Helmholtz Institute for Climate Service Science (HICSS), a coperation between Climate Service Center Germany (GERICS) and Universität Hamburg, Germany.

This work was supported by the project ADAPTER, which is funded by the Helmholtz Association under agreement number WT-0104.

We would like to thank Laura Schmidt for a comprehensive internal review and Jo-Ting Huang-Lachmann for her very thorough English language check.

REFERENCES

(Publications included in the literature overview are marked by: L)

Belcher, B., Claus, R., Davel, R., Jones, S.M. (2021). Evaluating and improving the contributions of university research to social innovation. Social Enterprise Journal 1750-8614

Belcher, B., Rasmussen K.E., Kemshaw, R.M., Zornes, D. (2016). Defining and assessing research quality in a transdisciplinary context. Research Evaluation 25(1) – L

Bergmann, M., Brohmann, B., Hofmann, 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. ISOE-Studientexte 13. Frankfurt am Main – L

Brasseur, G.P., Gallardo, L. (2016). Climate Services: Lessons Learned and Future Prospects. Earth's Future 4(3)

Bremer, S., Meisch, S. (2017). Co-production in climate change research – reviewing different perspectives. WIREs Climate Change 8 (6)

Brinkmann, C., Bergmann, M., Huang-Lachmann, J.T., Rödder, S., Schuck-Zöller, S. (2015). Zur Integration von Wissenschaft und Praxis als Forschungsmodus. Climate Service Center Germany. Report 23, Hamburg

European Commission (2015). A European Research and Innovation Roadmap for Climate Services. https://ec.europa.eu/programmes/horizon2020/en/news/european-research-and-innovation-roadmap-climate-services (31.10.2021)

Garard, J., Koch, L., Kowarsch, M. (2018). Elements of success in multi-stakeholder deliberation platforms. Palgrave communications 4:129 – L

Haße, C., Kind, C. (2019). Updating an existing online adaptation support tool: insights from an evaluation. Climatic Change 153

Jahn, T., Keil, F. (2015). An actor-specific guideline for quality assurance in transdisciplinary research. Futures 65 – L

Jagannathan, K., Arnott, J.C, Wyborn, C., Klenk, N., Mach, K.J., Moss, R.H., Sjostrom, K.D. (2019). Great expectations? Reconciling the aspiration, outcome, and possibility of co-production. Current Opinion in Environmental Sustainability 42 – L

Körner, C., Lieberum, A. (2014). Instrumente der Anpassungskommunikation in nordwest2050. Evaluation der Online-Medien. Beese, K., Katz, C., Körner, C., Molitor, H. (eds.) Anpassung an regionale Klimafolgen kommunizieren. Konzepte, Herausforderungen und Perspektiven. München, Oekom

Krause, G., Schupp, M. (2019). Evaluating knowledge transfer at the interface between science and society. GAIA 28/3 – L

Lux, A., Schäfer, M., Bergmann, M., Jahn, T., Marg, O., Nagy, E., Ransiek, A.C., Theiler, L. (2020). Societal effects of transdisciplinary sustainability research. How can they be strengthened during the research process? Environmental Science & Policy 101 – L

Maag, S., Alexander, T.J., Kase, R., Hoffmann, S. (2018). Indicators for measuring the contributions of individual knowledge brokers. Environmental Science & Policy 89 – L

Mauser, W., Klepper, G., Rice, M., Schmalzbauer, B.S., Hackmann, H., Leemans, R., Moore, H. (2013). Transdisciplinary global change research: the co-creation of knowledge for sustainability. Current Opinion in Environmental Sustainability (5)

Nagy, E., Ransieka, A., Schäfer, M., Lux, A., Bergmann, M., Jahn, T., Marg, O., Theiler, L. (2020). Transfer as a reciprocal process: How to foster receptivity to results of transdisciplinary research. Environmental Science & Policy 104 – L

Newig, J., Jahn, S., Lang, D.J., Kahle, J., Bergmann, M. (2019). Linking modes of research to their scientific and societal outcomes. Evidence from 81 sustainability-oriented research papers. Environmental Science & Policy 101 – L

Norström, A.V., Cvitanovic, C., Löf, M.F., West, S., Wyborn, C., Balvanera, P., Bednarek, A.T., Bennett, E.M., Biggs, R., de Bremond, S., Campbell, B.M., Canadell, J.G., Carpenter, S.R., Folke, C., Fulton, E.A., Gaffney, O., Gelcich, S., Jouffray, J.B., Leach, M., Le Tissier, M., Martín-López, B., Louder, E., Loutre, M.V., Meadow, A.M., Nagendra, H., Payne, C., Peterson, G.C., Reyers, B., Scholes, R., Speranza, C.I., Spierenburg, M., Stafford-Smith, M., Tengö, M., van der Hel, S., van Putten, I., Österblom, H. (2020). Principles for knowledge co-production in sustainability research. Nature Sustainability 3 – L

OECD (2002). Glossary of key Terms in Evaluation and results based management. http://www.oecd.org/development/peer-reviews/2754804.pdf (31.10.2021)

Pohl, C., Truffer, B., Hirsch Hadorn, G. (2017). Addressing Wicked Problems through Transdisciplinary Research. Robert Frodeman (ed). The Oxford Handbook of Interdisciplinarity (2nd). Oxford University Press

Restrepo, M.J., Lelea, M.A., Kaufmann, B.A. (2020). Assessing the quality of collaboration in transdisciplinary sustainability research. Farmers' enthusiasm to work together for the reduction of postharvest dairy losses in Kenya. Environmental Science & Policy 105

Schuck-Zöller, S., Brinkmann, C., Rödder, S. (2018). Integrating Research and Practice in Emerging Climate Services: Lessons from Other Transdisciplinary Dialogues. Serrao-Neumann, S., Coudrain, A., Coulter, L. (eds). Communicating Climate Change Information for Decision-Making. Springer International – L

Vaughan, C., Dessai S. (2014). Climate services for society: origins, institutional arrangements, and design elements for an evaluation framework. Wiley Interdisciplinary Reviews. Climate Change 5 – L

Wall, T.U., Meadow, A.M., Horganic, A. (2017). Developing evaluation indicators to improve the process of coproducing usable climate science. Weather, Climate, and Society 9 – L

Wolf, B., Lindenthal, T., Szerenscits, M., Holbrook, J.B., Heß, J. (2013). Evaluating Research beyond Scientific Impact. GAIA 22(2) – L

Wooten, K.C., Rose, R.M., Ostir, G.V., Calhoun, W.J., Ameredes, B.T., Brasier, A.R (2014). Assessing and Evaluating Multidisciplinary Translational Teams. A Mixed Methods Approach. Evaluation & the Health Professions 37(1) – L

AUTHORS

SUSANNE SCHUCK-ZÖLLER
Climate Service Center Germany (GERICS)/Helmholtz-Zentrum Hereon
Fischertwiete 1, 20095 Hamburg, Germany
E: susanne.schuck@hereon.de
ORCID ID: 0000-0003-0992-8334

SEBASTIAN BATHIANY
Climate Service Center Germany (GERICS)/Helmholtz-Zentrum Hereon
Fischertwiete 1, 20095 Hamburg, Germany
E: sebastian.bathiany@hereon.de

MARKUS DRESSEL
Climate Service Center Germany (GERICS)/Helmholtz-Zentrum Hereon
Fischertwiete 1, 20095 Hamburg, Germany
E: markus.dressel@hereon.de

JULIANE EL ZOHBI
Climate Service Center Germany (GERICS)/Helmholtz-Zentrum Hereon
Fischertwiete 1, 20095 Hamburg, Germany
E: juliane.el_zohbi@hereon.de
ORCID ID: 0000-0003-0329-0552

ELKE KEUP-THIEL
Climate Service Center Germany (GERICS)/Helmholtz-Zentrum Hereon
Fischertwiete 1, 20095 Hamburg, Germany
E: elke.keup-thiel@hereon.de

DIANA RECHID
Climate Service Center Germany (GERICS)/Helmholtz-Zentrum Hereon
Fischertwiete 1, 20095 Hamburg, Germany
E: diana.rechid@hereon.de
ORCID ID: 0000-0002-6035-2935

MIRKO SUHARI
Leuphana University
21335 Lüneburg, Germany
E: mirko.suhari@leuphana.de

KEYWORDS: (Formative) Evaluation, co-creation, science-practice engagement, climate change, criteria