ORIGINAL ARTICLE

Improving Scaling Performance in Research for Development: Learning from a Realist Evaluation of the Scaling Readiness Approach

Elias Damtew[1,2] · Boru Douthwaite[3] · Marc Schut[1,4] · Murat Sartas[1,5,6] ·
Cees Leeuwis[1]

Abstract

Complexity-sensitive decision support approaches (CSDSA) have gained prominence in the research for development (R4D) sector. However, limited attention has been given to critically examining the underlying causal assumptions of CSDSAs and their overall effectiveness in navigating complexity and achieving desired outcomes. Scaling Readiness has emerged as a novel CSDSA that is increasingly applied in R4D programs in low- and middle-income countries to improve the scaling of innovation. This study offers theory-based explanations on the extent to which Scaling Readiness supports evidence-based design, implementation and monitoring of scaling strategies in two R4D interventions. The contribution of Scaling Readiness is influenced by various contextual factors, including pre-existing partnerships and established institutional intervention project and performance management practices. The findings underscore the significance of investing in broader institutional impact culture growth. This includes critical evaluation of how funding, incentive, and performance mechanisms enable or constrain evidence-based decision-making and adaptive management at intervention and organizational level towards achieving Sustainable Development Goals.

✉ Marc Schut
    marc.schut@wur.nl

[1] Knowledge, Technology and Innovation Group, Wageningen University, Wageningen, The Netherlands

[2] Sociotechnical Systems Research Center, Local Innovation Group, Massachusetts Institute of Technology (MIT), Cambridge, USA

[3] Selkie Consulting Ltd., Westport, Ireland

[4] Portfolio Performance Unit, CGIAR System Organization, Kigali, Rwanda

[5] West and Central Africa, International Institute of Tropical Agriculture, Kigali, Rwanda

[6] Climate Action, The Alliance of Bioversity and CIAT, Rome, Italy

Résumé

Les approches d'aide à la décision sensibles à la complexité (CSDSA) sont de plus en plus utilisées dans la recherche pour le développement (R4D). L'un des défis est le manque de réflexion critique sur la question de savoir si et comment ces CSDSA contribuent à naviguer dans la complexité et à atteindre les résultats souhaités. Scaling Readiness est un nouveau CSDSA développé et largement utilisé dans les programmes de R4D dans les pays à revenu faible et intermédiaire. Cette étude fournit des explications théoriques sur la question de savoir si et comment Scaling Readiness a appuyé la mise à l'échelle des processus décisionnels dans deux interventions R4D. L'étude montre que le fait de rendre explicite la logique de l'hypothèse causale derrière la CSDSA est une première étape précieuse dans leur évaluation. En ce qui concerne la mise en œuvre de Scaling Readiness, différents facteurs contextuels tels que les partenariats existants et les pratiques institutionnelles enracinées de gestion des projets et des performances modèrent sa contribution à la mise à l'échelle de processus décisionnels et à l'amélioration des performances de mise à l'échelle. Enfin, l'étude souligne l'importance d'investir dans une capacité institutionnelle de façon élargie, en particulier dans le changement de mentalité et de culture, pour réaliser le plein potentiel de l'approche de Scaling Readiness. Cela inclut une réflexion critique sur les mécanismes de financement, d'incitation et de performance et si ceux-ci favorisent la prise de décision et la gestion adaptative le long du cheminement lié à l'innovation et à la mise à l'échelle.

Introduction

Over the years, the emergence of complexity-sensitive decision support approaches (CSDSA) have aimed to help research for development (R4D) interventions navigate complexity and achieve broader impact (Schut et al. 2020; Wigboldus et al. 2016; Swaans et al. 2014). The approaches intend to support innovation processes and create opportunities to continuously improve the evidence base for collective decision-making and action (Rose et al. 2016). Examples of popular approaches to support innovation and scaling performance include Scaling Readiness (Sartas et al. 2020b), PRactice-Oriented Multi-level perspective on Innovation and Scaling (PROMIS) (Wigboldus et al. 2016), Scaling Scan (Woltering et al. 2019), SCALE (FHI 360 2004), the IFAD Scaling Up Framework and MSI Approach (Cooley and Linn 2014). The approaches are informed by systems thinking and complexity perspectives with often implicit change theories and assumption logics (Rogers 2008). Much has been explored about the value of the approaches when intervening in complex natural, social and economic systems; far fewer studies have focused on the actual outcomes (Douthwaite and Hoffecker 2017). Despite the considerable interest and use of CSDSA to support R4D decision-making and strategy design processes,

it is not self-evident that they serve their intended purpose (Falleti and Lynch 2009). Since CSDSAs are usually embedded in an already existing trajectory of (institutional) thinking and modes of operation, their contribution to decision-making always stands the test of many foreseen and unforeseen circumstances (Rogers 2008). R4D interventions often find themselves in a web of tensions, while continuously changing internal and external dynamics (Blundo-Canto et al. 2019; Horton and Mackay 2003) that have implications on how decision support approaches are appreciated and used. Hence, from an evaluation viewpoint, the validity of CSDSA assumptions regarding its usefulness cannot be taken for granted and require sound empirical investigation. This helps to understand whether, how and why their application leads to intended outcomes, including the role (facilitative or disruptive) that contextual issues play in their contribution to intervention decisions and outcomes. This can also generate important lessons for improving the approaches and revisiting their underlying change theories and assumptions.

CSDSA evaluations often become complicated by the generic and unspecified change theory models on how and/or why such approaches work—due to little articulation of plausible causal mechanisms or logic models to enable systematic evaluation (Pawson and Tilley 1997). As Rogers (2008) noted, without an explicit logical model laying out the specific activities and anticipated outcomes, it is impossible to carry out a meaningful evaluation. Without an 'off-the-shelf' evaluation framework, there is a need to define and analyze a set of hypotheses or mechanisms to help evaluators accrue learning, identify patterns and test relationships within real-world application of the CSDSA (Meyfroidt 2016; Pawson 2013; Della Porta and Keating 2008). Notwithstanding, this is rarely performed in R4D contexts where CSDSA evaluation is largely based on an *ex-post* descriptive assessment. It is also rarely applied to randomized controlled experiments where salient causal processes that explain, for instance, how or why the same input could lead to multiple outcomes in different situations are black-boxed (Cieslik and Leeuwis 2021; Schmitt and Beach 2015). Accordingly, an essential step in the systematic evaluation task of CSDSA is explicating theories of change with testable causal mechanisms on how such approaches influence decision-making processes and outcomes of R4D interventions. The Theory of change, most commonly in the form of causal diagrams showing expected links from activities to outputs to outcomes to impact, has been widely used in the development and assessment of impact pathways of R4D program interventions (Wigboldus et al. 2020; Maru et al. 2018). However, these diagrams rarely consider the often-non-linear dynamics of programs operating in complex environments, which are normal for agricultural R4D programs (Douthwaite and Hoffecker 2017).

This study uses a theory-based approach to evaluate Scaling Readiness, a specific CSDSA that aims to support evidence-based design, implementation and monitoring of strategies for scaling innovations. For this study, we define innovation scaling as the complex, multi-faceted and only partially manageable process through which an innovation becomes embedded and used in a given societal context at a scale that generates intended and/or unintended development impacts (Wigboldus et al. 2016). This embedding and use involves the innovation becoming part of a package of social and technical components that is supported and used by a network or coalition

of interdependent stakeholders, working towards the replacement or improvement of pre-existing technical, social and institutional practices (see Schut et al. 2020; Sartas et al. 2020b). Our definition of scaling acknowledges that scaling often requires simultaneous processes of scaling up (impacting through law and policy), scaling out (impacting through deliberate replication/ spread), scaling down (impacting through replacing existing practices) and scaling deep (impacting through changing underlying norms, values and cultures) (Moore et al. 2015).

This study aims to achieve two objectives. The first objective is to empirically assess whether and how Scaling Readiness, as CSDSA case, works and can deal with the complexity of the R4D contexts in which it is applied. The second objective is to draw lessons relevant for the theory, evaluation and implementation of CSDSA in general and that of Scaling Readiness in particular. The study raises conceptual, methodological and practical issues that need to be considered when applying or evaluating CSDSA in complex R4D contexts. As further expounded in the methodological framework, our assessment goes through three phases:

I.   Elicit a theory of change for the Scaling Readiness approach;
II. Empirically evaluate whether and how Scaling Readiness contributes to scaling decision-making processes and improved scaling performance;
III. Draw lessons relevant for the theory, evaluation, and implementation of CSDSA and Scaling Readiness in the R4D context.

Methods

Eliciting a Theory of Change and Causal Mechanisms

Proponents of theory-based evaluations have advanced the use of a change theory as an explanatory account of how programs, approaches or interventions works. The analysis guiding the evaluation of causal mechanisms as a tool to enable stronger causal inferences (Astbury and Leeuw 2010; Pawson and Tilley 1997). The causal mechanisms are sets of interrelated statements about the nature, agents and objects of change. They provide more fine-grained explanations (Checkel 2006) and increase the theory's credibility.

Our formulation of plausible causal mechanism is based on a review of conceptual perspectives that underpinned the development of the Scaling Readiness approach (Sartas et al. 2020b), in-depth discussion with the team that developed Scaling Readiness and insights from previous preliminary pilot testing of the approach in different R4D contexts (Sartas et al. 2017). Accordingly, six overarching causal mechanisms as building blocks of the Scaling Readiness theory of change were articulated for further evaluation (see Results and Analysis section and Fig. 2). Each causal mechanism was formulated as a hypothesis including an activity (AC), a mechanism (M), and activity outcome (AO) as its interrelated subcomponents

(Beach and Pedersen 2011). The overall causal logic is based on a particular Scaling Readiness activity generating information that will contribute to a decision or outcome through one or more mechanisms which, in turn, provides a clearer conceptual footing for our evaluation. The causal mechanisms are laid out as relatively broad pathways whereby AC contributes to producing AO through M (Beach and Pedersen 2019).

Empirical Testing

R4D Context

Two CGIAR-RTB-program funded scaling projects[1] employed the Scaling Readiness approach to guide their overall scaling activities and decisions. *Scaling approach for flash drying of cassava starch and flour at small scale* was a two-year project launched in Colombia, DR Congo and Nigeria, with the objective to increase energy efficiency of small-scale cassava processing thereby reduce production costs of cassava starch and flour, two main industrial products of cassava. *Orange Fleshed Sweet potato (OFSP) Puree for Safe and Nutritious Food Products and Economic Opportunities for Women and Youths* was another 2-year Scaling Fund project that aimed to increase the use of OFSP puree in fried and baked products in Kenya, Uganda and Malawi. These two projects served as R4D contexts to test the causal mechanisms and theory of change of the Scaling Readiness approach. The projects are described in more detail in Schut et al. (2022).

Empirical Data

This study was based on information from written secondary sources and in-depth interviews. The secondary sources included all accessible documents produced by the projects, including Scaling Fund project proposals, capacity development workshop reports, intervention characterization documents, Scaling Readiness diagnosis survey reports, stakeholder engagement plans, scaling strategy/activity plan documents and quarterly and annual project reports. After an initial review of the written sources that aimed to develop a timeline of events, thirteen in-depth, 90-min interviews were conducted with intervention managers, researchers and implementers of the Scaling Fund projects between December 2020 and June 2021. For the interviews, a set of semi-structured questions and discussion topics were developed that corresponded with the timeline of events and the different components of the causal mechanisms (see Supplementary Material).

---

[1] The CGIAR Research Program on Roots, Tubers and Banana (RTB) had a so-called 'Scaling Fund.' The Scaling Fund supported researchers and public/ private scaling partners to design, implement and monitor strategies for scaling innovation with a high scalability potential, using the Scaling Readiness approach.

Analysis

The study had a systematic way for identifying and utilizing evidence. Specifically, an examination of the documents and interviews to unearth three types of evidence was made (Schmitt and Beach 2015; Beach and Pedersen 2019):

- *Sequence evidence* Temporal chronology of activities and outcomes as antici-pated by a specific hypothesized causal mechanism; meaning, we expect to see the triggering factor before an anticipated outcome.
- *Account evidence* The content of empirical material, such as an oral account of what took place or meeting minutes that detail what was discussed regarding a causal mechanism; for example, the execution or non-execution of a Scaling Readiness activity or how it led to the expected or a different outcome.
- *Trace evidence* Evidence whose mere existence provides proof for the presence of (part of) a causal mechanism; for example, a resource invested for the execu-tion of a Scaling Readiness activity.

A stepwise workflow employed the three evidence types for the empirical testing. Figure 1 shows a simplified depiction of our causal analysis for easier understanding of the analytical logic. As this section will show, the actual analytical process was not as linear and straightforward as shown in the figure.

The sequence-evidencing processes involved developing a timeline of Scaling Readiness activities and outcomes as they unfolded in the Scaling Fund projects. Next, in-depth analysis of documents and interviews was carried out, guided by exploring 'what informed the event' and 'what was the outcome of the event.' In this regard, any explanation found in documents (*Trace and Account evidence*) or given by interviewees (*Account evidence)* on the causes and consequences of events was used to build a causal story around the events. It is important to note that the types of evidence used in the analysis are not mutually exclusive and what is used as sequence evidence in most cases is also trace evidence.

While the timelines were being developed, there was a corresponding task of determining whether and how the sequence of AC and AO in the elicited theory of change align or not with the hypothesized causal mechanism. This alignment informed whether the causal mechanism was valid. Finding sequential fit strength-ens the validity of the casual assumption, whereas a mismatch in the temporal sequence of events between the observed and the assumed was taken as an indica-tion for a different causal mechanism.

The account and trace evidencee helped further strengthen empirical testing by showing whether and how the causal mechanisms manifested in the case fitted with what was assumed in the theory. A key aspect of this causal inferencing task is decerning mismatch between the assumed and the observed then the likely reasons, mechanisms, and/or contextual conditions that could explain the difference. Such information was central in our discussion of important contextual factors that should be considered and potential adaptations that might be needed to improve the useful-ness of CSDSA in general, and Scaling Readiness in particular, in supporting R4D interventions to navigate complexity and make evidence-based decisions.
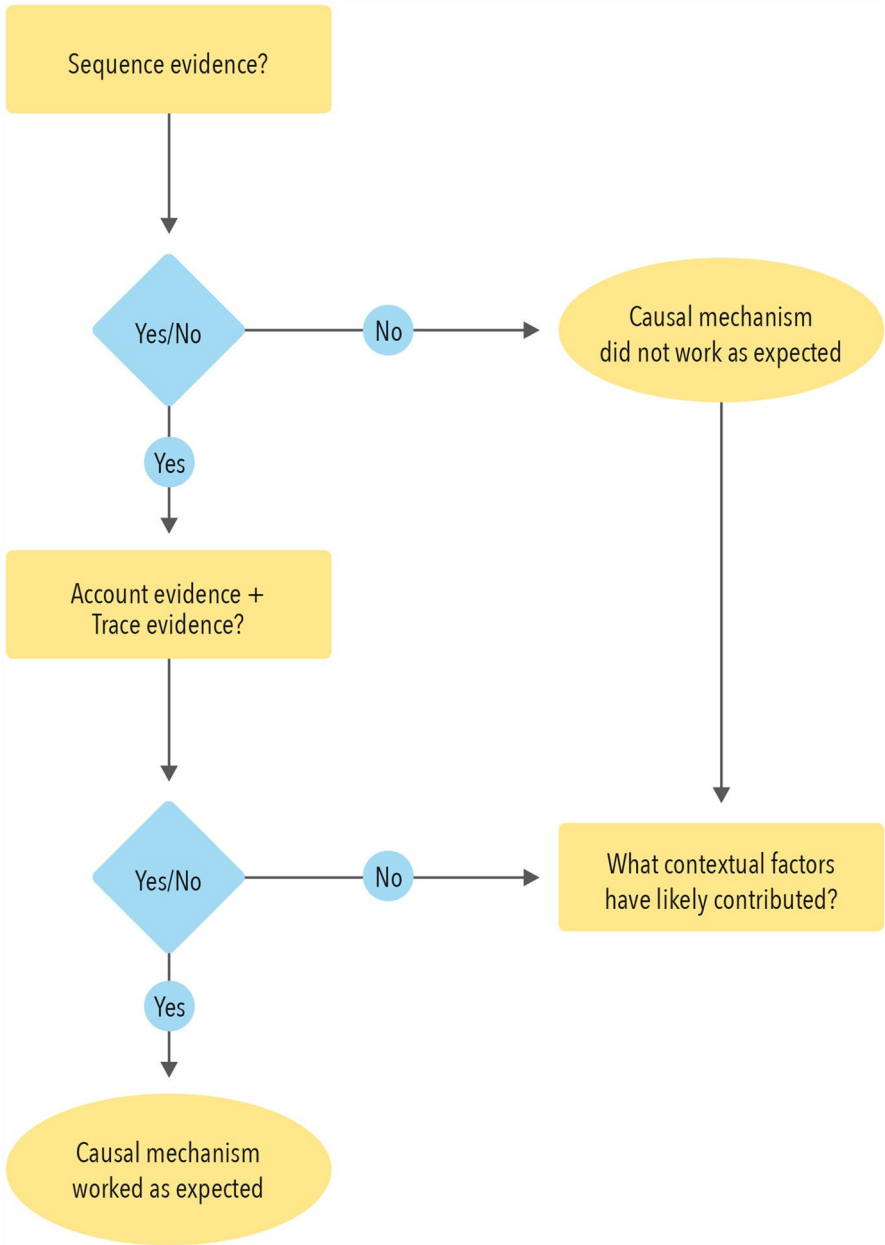
Fig. 1 A simplified depiction of the causal analysis logic

Role and Position of the Researchers

The (co)authors of this contribution played different roles in the research journey. The three last mentioned authors conceptualised and developed Scaling Readiness

(see Sartas et al. 2020a, b), while the third and fourth authors trained and supported the team that used Scaling Readiness in the two cases selected for this study. The three last mentioned authors and the CGIAR programme (that funded the development of Scaling Readiness) proposed to conduct a study that would rigorously and critically examine the contribution of Scaling Readiness to scaling decision-making processes and improved scaling performance. The first and second authors were selected to conduct the study after terms of reference and a tendering procedure were developed. In their selection, criteria such as research experience, affinity with the study subject and independence played an important role. The first author did most of the empirical work and was advised by the second author on issues pertaining theory (of change) and methodology. The last three authors were discussion partners in developing the methodological design, served as respondents for Phase I of the study (*elicit a theory of change for the Scaling Readiness approach*) and gave various rounds of feedback on the case reports and contributed to the drafting of this paper, which were produced by the first and second authors. This feedback was oriented to sharpening and enhancing the quality and rigour of the text, while respecting the independent interpretation, judgment and conclusions of the first and second authors.

Result and Analysis

Scaling Readiness is an approach developed to support R4D organizations, projects and programs in scaling innovations and achieving impact. In addition to providing decision support in a management sense, it encourages critical reflection on how ready innovations are for scaling and what appropriate actions could accelerate or enhance scaling (Schut et al. 2020; Sartas et al. 2020a, b). Anchored on conceptual and methodological perspectives and approaches from various fields—such as innovation system science, technology studies, social network studies—the approach proposes a structured set of activities that generate information for making evidence-based scaling decisions. Since the focus of this study is on identifying and evaluating the theoretical assumptions, we will not go into the specifics of the activities, measures and/or implementation steps of Scaling Readiness, however, detailed information can be found in published works on the approach (Sartas et al. 2020a,b).

Elicited Scaling Readiness Theory of Change

Given the rather dominant technology-centered approach to scaling generally taken by R4D interventions (Pfotenhauer et al. 2022; Wigboldus et al. 2016), Scaling Readiness offers capacity development around the underpinning concepts and principles of system-oriented approaches to innovation and scaling as a key undertaking to shape practices (Sartas et al. 2020a). This is expected to catalyze critical reflection within R4D intervention teams on why scaling of innovation is challenging, why many of the current approaches have not resulted in the desired impact and what needs to be done differently to be more successful. Through expert-facilitated

capacity development workshops, intervention teams and partners make informed decisions about which strategies, partnerships and investments have the highest likelihood to achieve their scaling ambitions.

Figure 2 provides a schematic representation of the generic Scaling Readiness theory of change, its key steps and causal mechanisms. These were derived from the literature and based on discussion with the Scaling Readiness development team. Section 3.2 interrogates the theory of change and its causal assumptions based on case data.

Causal mechanism 1: "Capacity development within scaling interventions (AC1)" leads to "higher willingness to invest time and resources in developing, implementing and monitoring evidence-based scaling strategies (AO1)" through "a better understanding of the key principles and concepts underlying scaling of innovation (M1)."

Once there is a willingness to invest time and resources in the development, implementation and use of scaling strategies, this approach provides a specific decision support that identifies critical bottlenecks for scaling of innovation in a specific context (Sartas et al. 2020a). This is anchored in the rationale that despite the focus of many R4D interventions to scale a specific technology or innovation, innovations scale as part of an innovation package (Barrett et al. 2020; Sartas et al. 2020b). The scaling of any core innovation is influenced by interactions with other complementary innovations that can either enable or constrain, the latter as bottleneck innovations. Through the definition of innovations as a package and the systematic valuation of the level of maturity (innovation readiness) and level of support by networks surrounding the innovation (innovation use), Scaling Readiness is assumed to support R4D interventions by prioritizing systemic bottlenecks that limit scaling potential and, hence, need to be addressed.

Causal mechanism 2: "Context-specific innovation packages and the assessment of their Scaling Readiness (AC2)" facilitates "the prioritization of bottleneck innovations" through "a greater awareness of interdependencies between core and complementary innovations and their current innovation readiness and use to achieve societal outcomes (M2)."

Following the broader perspective on scaling innovations and (re)defining an intervention as a package of interdependent innovations, identifying bottlenecks is expected to serve as key diagnostic evidence for initiating the design of a scaling strategy (Sartas et al. 2020b). It is suggested that R4D interventions can pursue different options to overcome the bottlenecks and increase the impact potential of the innovation package. To this end, Scaling Readiness is anticipated to facilitate critical reflection and discussion on available options and to make strategic decisions to overcome the bottleneck innovations given available time, human and financial resources.

Causal mechanism 3: "The systematic exploration of strategic options to overcome bottleneck innovations by the intervention team (AC3)" results in "better/different decisions regarding proposed investments and actions as part of a
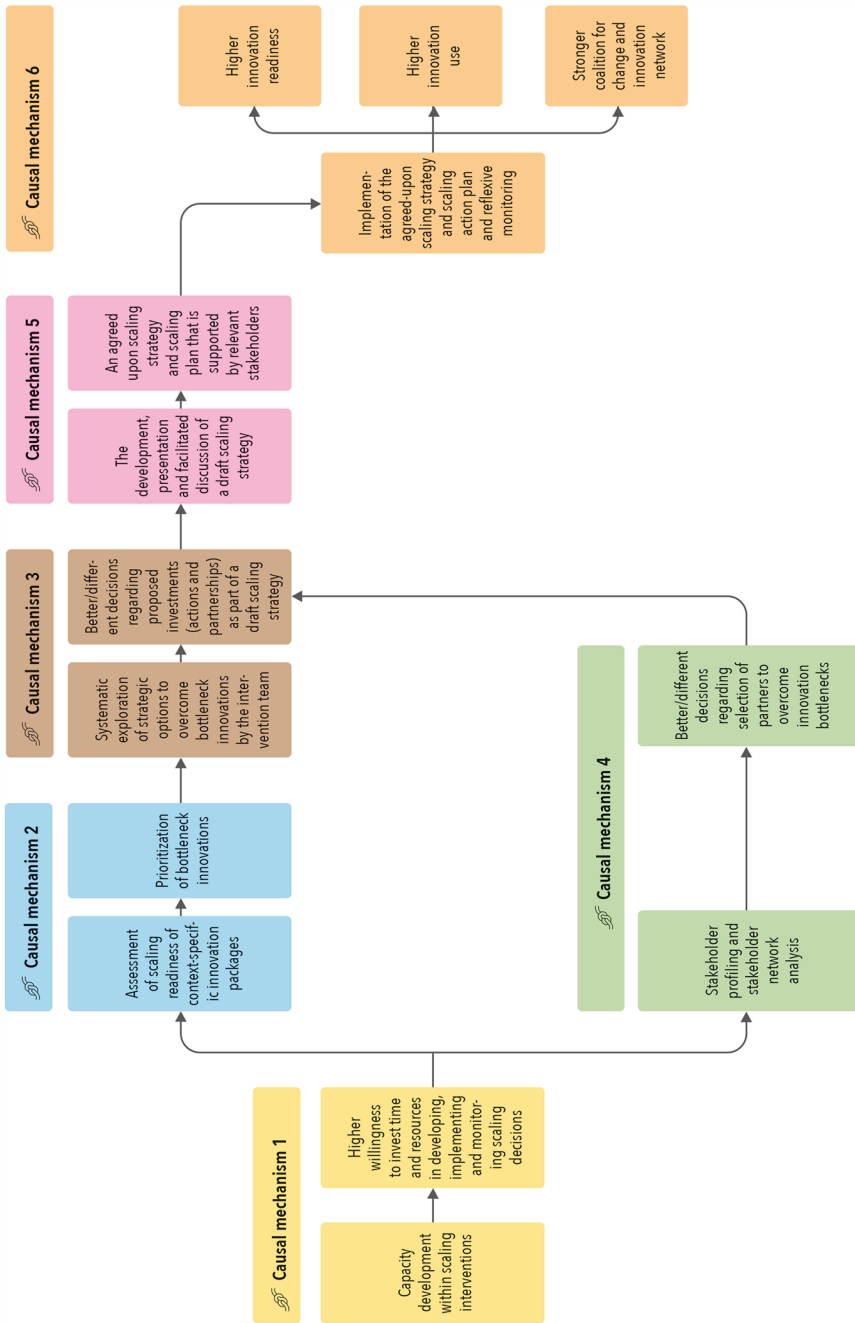
Fig. 2 Schematic representation of the Scaling Readiness theory of change

draft scaling strategy (AO3)" through "a greater/novel awareness of available options for enhancing the Scaling Readiness of the innovation packages that are realistic within limitations of the scaling intervention (M3)."

 Partnerships and collaborations are central to scaling innovations, regardless of sector, approach or pathway (Kohl 2021). In other words, the scaling potential of innovations is largely shaped by the social networks in which they are embedded, supported and used. Overcoming scaling bottlenecks requires the identification of stakeholders who are well-placed to contribute to increasing the bottlenecks' innovation readiness and use (Sartas et al. 2020b). An essential part of scaling strategy development is, therefore, having a good understanding of the stakeholder context in a particular intervention setting. With an emphasis on characterizing the broader system in which a scaling intervention operates, Scaling Readiness stakeholder profiling and network analysis are anticipated to influence decisions on the selection of potential partners. This can help overcome the bottlenecks as the intervention team becomes better aware of stakeholders, their mandates and involvements in activities relevant for the scaling of innovations.

> *Causal mechanism 4:* "Stakeholder profiling and stakeholder network analysis (AC4)" leads to "better/different decisions regarding selection of partners to overcome the innovation bottlenecks (AO4)" through "a greater/novel awareness of gaps in the competencies that are required for scaling (M4)."

 Relatedly, R4D interventions cannot realize their scaling ambitions independently and depend on other innovation system stakeholders (Schut et al. 2020; Eastwood et al. 2017). Stakeholders can provide diverse insights into the technological, organizational and institutional characteristics of the problem, the innovations to be scaled and the scaling context or the innovation system (Sartas et al. 2020b). They may pursue different and potentially even conflicting goals and interests, directly affecting the successful implementation of any proposed scaling strategy (Sartas et al. 2020b; Wigboldus et al. 2016). Scaling Readiness is anticipated to catalyze this important process by engaging relevant stakeholders in the deliberation and negotiation process of proposed scaling strategies (Sartas et al. 2020a).

> *Causal mechanism 5:* "The development, presentation and facilitated discussion of a systematically underpinned draft scaling strategy (AC5)" leads to "an agreed-upon scaling strategy and scaling action plan that is supported by relevant stakeholders (AO5)" through "a better understanding of the scaling strategy building blocks and a greater motivation to collaborate towards overlapping objectives (M5)."

 R4D interventions operate in real and uncontrolled settings, which implies that stakeholders and intervention implementers are likely to face unforeseen developments and activities that give rise to unintended consequences and outcomes (Geels and Schot 2007). With the scaling context changing continuously, interventions require mechanisms to capture and navigate the dynamic situation in

which they operate by remaining flexible and ready to embark on new paths (Sartas et al. 2020b; Klerkx et al. 2012). In this regard, Scaling Readiness is deemed to facilitate and monitor the scaling strategy implementation through reflexive learning (Sartas et al. 2020b; Van Mierlo et al. 2010). The approach assists intervention teams in periodically reflecting on the implementation of the scaling strategy and, if necessary, updating it to reach the desired scaling objective. Monitoring can be based on short-term feedback loops that guide the scaling strategy implementation. They can also rely on long-term feedback loops with a second innovation system analysis (innovation package reconfiguration and stakeholder profiling) and Scaling Readiness assessment to see whether the scaling strategy has had the desired effect on improving the Scaling Readiness of the innovations (Sartas et al. 2020a).

> *Causal mechanism 6:* "The reflexive monitoring of the implementation of the agreed-upon scaling strategy and scaling action plan (AC6)" leads to "improved scaling performance (AO6)" through "overcoming bottleneck innovations and greater enthusiasm, energy, and synergy in the partnership (M6)."

Testing the Scaling Readiness Theory of Change

This section unpacks the causal processes that unfolded in the two R4D contexts using the constructed causal mechanisms as theoretical scaffolding. The sequence evidence is subscripted with numbers (1, 2, 3, etc.) in chronological order cutting across the six causal mechanisms. The trace and account evidence are subscripted with letters **'a'** and **'b'** within a specific causal mechanism serving as casual narratives establishing and/or strengthening the causal analysis. We intentionally focus on unpacking *whether* and *how* the *CSDSA* (Scaling Readiness) informed scaling *activities* and *outcomes* and less about detailing the content of those activities and outcomes. A more detailed account of the cases' activities and outcomes can be found in Schut et al. (2022).

Cassava Flash Dryer Case

Between early 2019 and the end of 2020, the Cassava flash dryer Scaling Fund project was jointly implemented by different partnering organizations in Nigeria, Colombia and DRC to promote cost-effective Cassava drying solutions (Schut et al. 2022).

Causal mechanism 1 [1]The intervention team was first exposed to the Scaling Readiness approach and underpinning concepts and principles during the Scaling Fund proposal development process. [b]After the project launch, seminars on Scaling of Innovations and the Scaling Readiness approach were organized that shaped subsequent scaling practices **(AC1)**. [2]The project invested time and resources to redefine its scaling intervention as packages of innovations that eventually guided the design of new scaling strategies for the different intervention locations. [b]It was pointed out that the new scaling approach came at a good time as the intervention designers were

searching for a decision support approach to scale the Flash Dryer, which had gone through a few years of technical experimentation **(AO1)**. [b]Interviews and document reviews demonstrate that there is a novel appreciation of important concepts in scaling of innovations (e.g., investment in key scaling bottlenecks, contextual approach to scaling, reflexive monitoring). [b]A key member of the intervention team elaborated his learning about the stark contrast between the new perspective to scaling of innovations and the existing technology-centered practices that overlooked the decisive role of the enabling environment for the scaling of the technologies **(M1)**.

Causal Mechanism 2 [2]With a follow-up innovation profiling, 16 potential components were identified as complementary innovations around the core innovation (*Efficient Small Scale Flash Dryer for Cassava Starch and Flour*). This was a significant modification to the intervention, considering only 5 components were put forward around the start of the project. [3]Assessment on the degree of use and level of readiness led to the complementary innovations being bundled as a package **(AC2)**. [a]The readiness assessment identified three key scaling bottlenecks specific to the three intervention locations. While the key bottlenecks identified in DRC were technological, the bottlenecks in Colombia and Nigeria have additional demand-side market problems (Colombia and Nigeria) and political clout (Nigeria) **(AO2)**. [b]Through the innovation profiling and Scaling Readiness assessment, the intervention team made sense of the different innovation components as one comprehensive innovation package. [a]The intervention was redefined as a more structured and interconnected set of component innovations and was systematically categorized as products, services, practices and institutional arrangements deemed necessary for the scaling of the technology. The Scaling Readiness assessment shed light on the most pressing bottlenecks that fall under the radar at the initial stage of the intervention **(M2)**.

Causal Mechanism 4[2] [4]With the emphasis on a better understanding of the broader scaling context, a stakeholder profiling and network analysis were conducted before a formal selection of partners. [a]In view of the initial stakeholder engagement plan of the scaling project, the stakeholder profiling and network analysis provided a coherent and detailed account of the stakeholder context **(AC4)**. [a]Partnerships were forged with some of the mapped-out equipment manufacturers and cassava processors as part of the work plan to address the scaling bottlenecks **(AO4)**. [a]The activity generated rich information on new stakeholders, their networks, mandates and level of involvement in the Cassava Flash Dryer system. [b]Its contribution was highlighted by the intervention team in the further screening and engagement of operational equipment manufacturers and Cassava processors **(M4)**.

Causal Mechanism 3 [5]The exploration of strategic scaling options focused on overcoming prioritized bottlenecks in the different intervention locations **(AC3)**. [a]It was decided to stop investments in cassava drying in Nigeria and Colombia as a result of

---

[2] Note to clarify that in the Flash Dryer case Causal Mechanism 4 occurred before Causal Mechanism 3 as is also shown in Fig. 3.
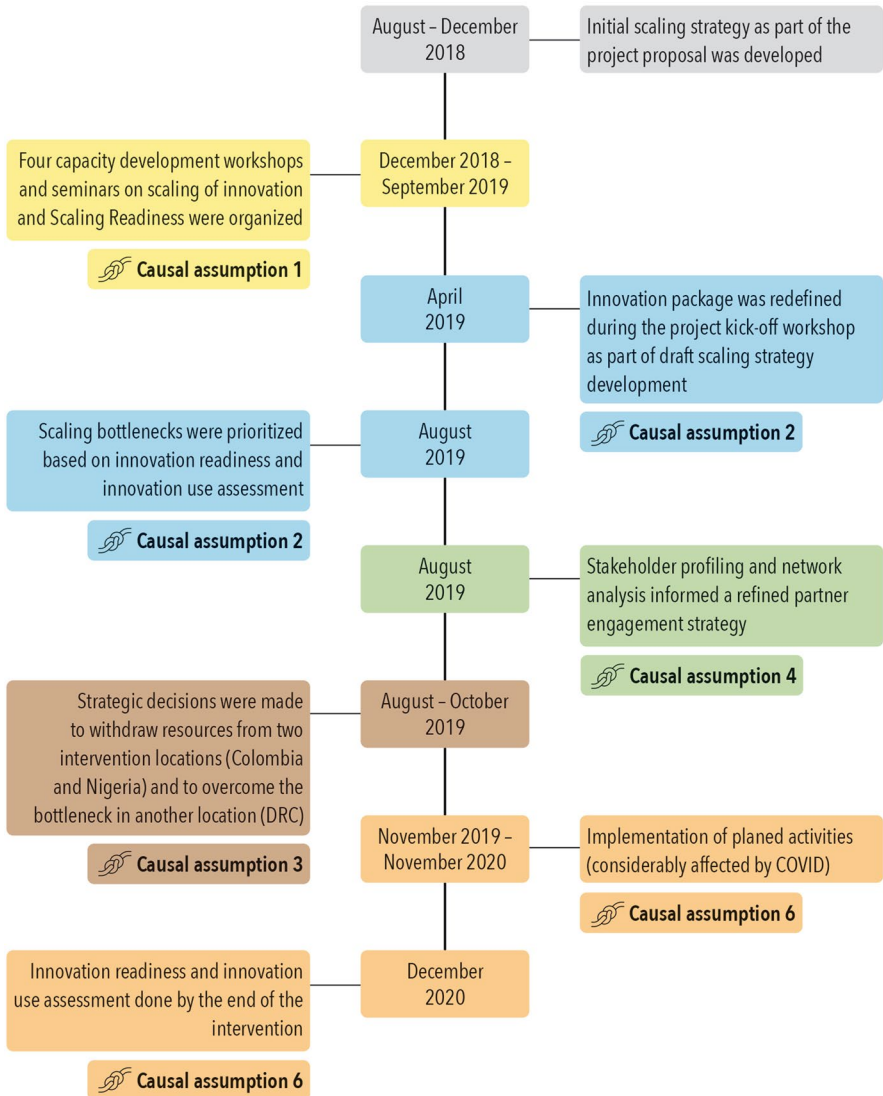
Fig. 3 Timeline of major events and outcomes Cassava Flash Dryer Case

market and policy issues that were beyond the scope of the project. In the DRC, the decision was made to continue to work on the bottlenecks **(AO3)**. [b] Scaling Readiness supported the identification of structural issues that demanded solutions beyond the scaling project's capacity and informed decisions to relocate and substitute interventions in Nigeria and Colombia **(M3)**.

Causal Mechanism 5[3] [b]Both document reviews and interviews indicated the lack of multi-stakeholder consultation and agreement processes in validating the different strategic decisions made by the intervention team **(AC5)**. [b]It was mainly through deliberations with implementing partners that agreements were signed to work together to improve the technical efficiency of operational Flash Dryers **(AO5)**. [6]The prioritization of technological bottlenecks and the subsequent technical improvement strategy seemed to have encouraged the intervention team to engage only with the partners that were equipment manufacturers and Cassava processors. This was justified by a shared belief amongst the intervention team and partners on a stable cassava market environment and the pressing need to improve the identified bottlenecks around the energy and production efficiency of existing Flash Dryers **(M5)**.

Causal Mechanism 6 [7]Apart from joint field visits and follow-up discussions, an online support network including the intervention team and implementing partners facilitated visual and textual information exchange, discussion and technical backstopping during the implementation of planned activities **(AC6)**. This online network was setup before the covid crisis, and proved all the more valuable as travel restrictions prevented the in-person interventions planned initially. [a]The implementation led to improvements in the readiness of some technological bottleneck innovations (e.g., pipe length, heat exchanger, air blower and feed system) **(AO6)**. [b]Interviews showed that the online collaborative platform was considered as a continuous learning and peer-support system that assisted processors to improve the efficiency of the technology. [a]This was further demonstrated by the continued collaborative engagement with the online space after the official closing of the project **(M6)**.

Case Summary In line with the theory of change, the empirical evidence on the Cassava Flash Dryer case provides support that the capacity development activities have catalyzed learning around concepts and principles on scaling of innovations and Scaling Readiness that informed scaling decisions in the development, implementation and monitoring of scaling strategies. The innovation context characterization and bottleneck prioritization has usefully contributed to the design of evidence-based scaling strategies in the different intervention locations. This is particularly the case in Nigeria and Colombia where planned scaling activities and associated resources were shifted to different locations and value chains to scale the Flash Dryer technology. Decisions were made to relocate and reorient interventions in Colombia and Nigeria, and to continue working on the technological bottlenecks in DRC. Scaling Readiness, through its proposed stakeholder engagement strategy, fostered reflexive monitoring and learning around the implementation of scaling activities that led to improvements in the Scaling Readiness of some of the bottleneck innovations. The observed changes in capacity, decision-making

---

[3] Due to an important Scaling Readiness informed decision to withdraw scaling investments in two of the intervention locations (Nigeria and Colombia) of this case study, findings presented under *Causal mechanism 5* and *6* point out to only scaling activities in DR Congo.

and resource allocation indicate that Scaling Readiness had kickstarted a deeper process of culture change and learning (e.g. Moore et al. 2015) in this case.

A follow-up visit to three of the private sector partners in DRC in November 2022 showed that improvements to flash dryers continued beyond the formal end of the Scaling fund project (2019–2020). After the initial training and online support provided throughout the project, the partners took ownership of the innovations and successfully applied them to reduce their energy consumption and production costs, and reach market viability. The Scaling Readiness-informed identification and prioritization of key bottlenecks played an essential role in focusing tasks and guiding partners through their investments and construction work (personal communication Flash Dryer project leader, March 2023).

OFSP Puree Case

The OFSP Puree Scaling Fund project was implemented from 2019 to 2020 in Kenya, Malawi and Uganda. Different international and national implementing partners from research, government and the private sector were involved to promote the use of OFSP puree in popular baked and fried goods in the different countries (Moyo et al. 2022) (Fig. 4).

Causal Mechanism 1 [2]A key component of the capacity building activity was a three-day workshop during the project inception on concepts and principles of scaling of innovations and the Scaling Readiness approach **(AC1)**. [a]By modifying the original project proposal developed before the introduction of Scaling Readiness, the approach redefined the intervention as a package of innovations that directly progressed scaling strategy design. **(AO1)**. [b]Scaling Readiness is appreciated as an approach that creates greater opportunities for different implementing partners to work on various intervention components of the package. [b]A key intervention team member understood the approach as a set of activities that can benefit the project without necessarily adhering to all its recommended activities. Another one questions the emphasis placed by the approach on *process management* rather than *delivery,* the later explained as 'reaching as many beneficiaries as possible with the technology.' **(M1)**.

Causal Mechanism 2 [b]Even though the intervention was redefined as a package of innovations, Scaling Readiness assessment of the new innovation packages was not done at this particular stage of the project. [b]Consequently, no systematic prioritization of bottlenecks was made to gauge if some bottlenecks were more critical than others before moving into partnerships or the design of scaling activities.

Causal Mechanism 3 [3]A work plan meeting was held with intervention partners whereby potential scaling activities in line with the different complementary innovations were presented and discussed—[a]a process principally hinged on draft work plans (activities, budget, timelines) put forward by the implementing partners **(AC3)**. [4]Initially proposed scaling activities and associated funds on the initial proposal were
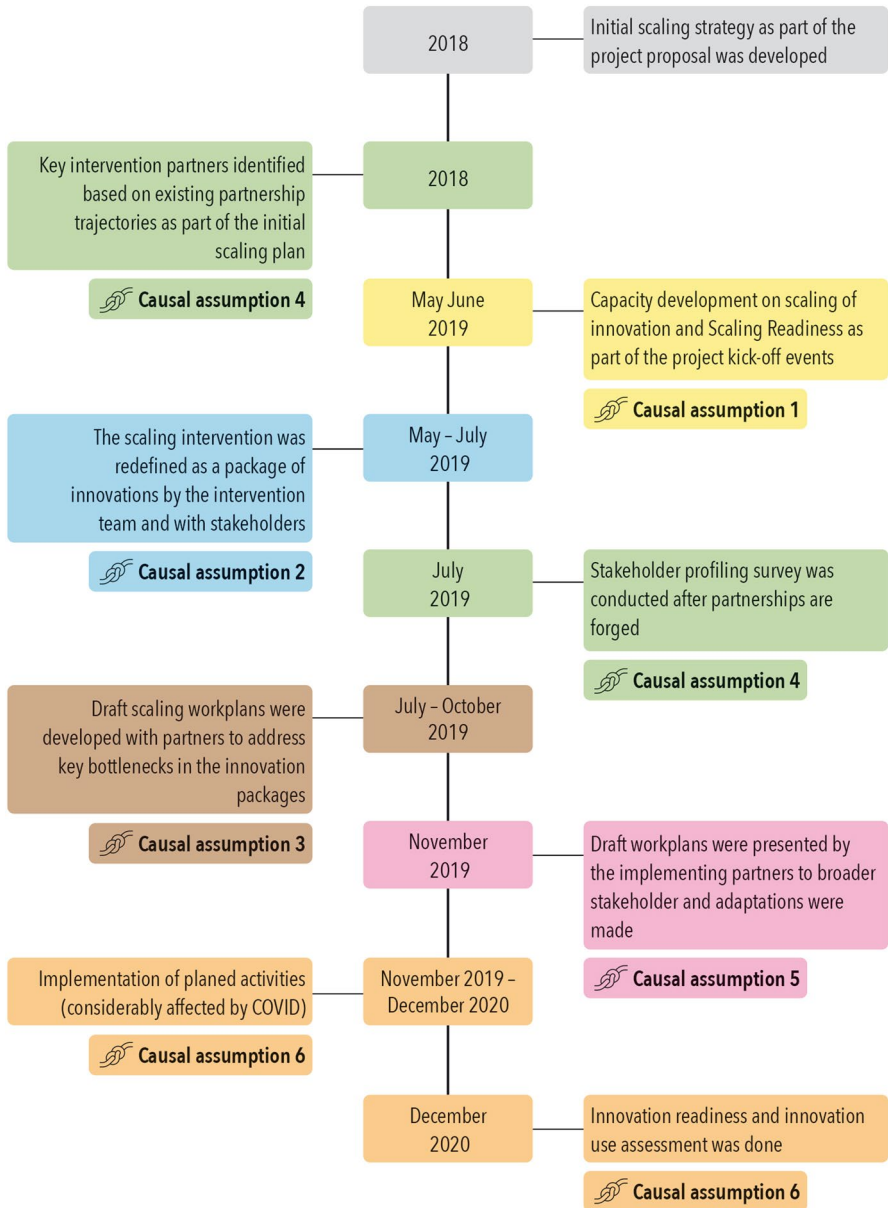
Fig. 4 Timeline of major events and outcomes OFSP Puree Case

adjusted as per the new scaling work plans developed around the complementary innovations (**AO3**). [b]With the establishment of partnerships before bottlenecks were prioritized, capitalizing on existing work and networks of implementing partners around the different complementary innovations was applied as strategy (**M3**).

Causal Mechanism 4 [5]Even though stakeholder profiling and stakeholder network analysis were done producing information on the type, mandate and levels of involvement of different system actors working around OFSP **(AC4)**. [1]Most partnerships were already established before scaling work plans were proposed **(AO4)**. [b]The key implementing partners have a longer history of working and existing partnerships reproduce themselves in the absence of a major disruption. This was observed in the explanations of intervention team members and implementing partners how previous and active working relationships in other projects were crucial in engaging in the OFSP Puree scaling project. [a]The key implementing partners were already designated as 'output leaders' in the project proposal or before Scaling Readiness activities were initiated **(M4)**.

Causal Mechanism 5 [6]Partners-championed scaling work plans around the different complementary innovations were presented and discussed with stakeholders working in the system. [b]Given that intervention partnerships were already formed, there was little room for deliberation on the partnership aspect of the scaling strategy or work plan **(AC5)**. [b]The development of the draft scaling work plans and the deliberation processes on the content of the work plan went concurrently whereby on the stakeholder consultation meeting an agreement was reached to hold a follow-up bilateral meeting with each implementing partner **(AO5)**. [a]The consultation process went through an engaging discussion whereby detailed feedback was provided around improvements needed on the action plan that implementing partners took into account before moving into implementation **(M5)**.

Causal Mechanism 6 During the implementation phase, there is little evidence of a reflective type of monitoring and learning processes. Internal update meetings were happening within the intervention team based on partners' quarterly written reports. A Scaling Readiness diagnosis (long-loop monitoring) that was conducted by the end of the project revealed changes in the innovation readiness and use of some of the complementary innovations at the different locations. However, this could not evidence any changes in innovation readiness and innovation use since the initial Scaling Readiness assessment was not conducted. For lack of a reflexive monitoring and learning process during the implementation process, interviews revealed a perceived disconnect between organizational monitoring and evaluation system and reflexive monitoring approach of Scaling Readiness. With the priority given to conventional monitoring and evaluation approach that they have been enacting, a 'second' type of monitoring was felt as an additional burden.

Case Summary The OFSP Puree case shed light on how capacity development may not always lead to the type of understanding or shift in thinking Scaling Readiness envisages to catalyze. The approach was mainly appreciated as a stakeholder engagement tool. Contrary to Scaling Readiness capacity development assumptions, the intervention team questioned the added value of focusing on (innovation) processes rather than on technology delivery. However, Scaling Readiness contributed to the design of a scaling strategy for the intervention through the

inclusion of new supporting innovation components relevant for the scaling of the technology. Previously established partnerships (which came first in the sequence of events of our causal analysis) seemed to have limited the prospect of prioritizing and investing in addressing key bottlenecks for scaling OFSP. Strong working ties or partnership trajectories that transcend the scaling project timeframe had implications for the approach's contribution to the scaling decision-making process. Scaling Readiness had limited influence on activity monitoring with observed difficulty in reconciling accountability and learning objectives. Contrary to the Flash Dryer case, the OFSP case shows less clear signs of deeper scaling culture and practice change (e.g. Moore et al. 2015).

Discussion

The discussion section "Reflections on the Scaling Readiness Theory of Change, Causal Mechanisms and Implementation" discusses whether or not the causal mechanisms worked as expected and suggests specific improvements to the Scaling Readiness approach theory and implementation. The section "Reflections on the Theory of Change of Complexity-Sensitive Decision Support Approaches" provides a broader reflection relevant to change theories of similar CSDSA.

Reflections on the Scaling Readiness Theory of Change, Causal Mechanisms and Implementation

Scaling Readiness Causal Mechanisms

The causal mechanisms were observed differently among the two Scaling Fund cases. Most of the mechanisms were supported in the Flash Dryer case except for causal mechanism 5. On the contrary, only Mechanism 5 was fully supported in the OFSP case, with mechanisms 2, 3 and 4 only partially supported (Table 1).

While the hypothesized *causal mechanism 1* was supported by findings from Cassava Flash Dryer case, the same could not be said for the OFSP case. In the OFSP case, a stakeholder-focused understanding of Scaling Readiness contributed to directing scaling investments in an already established working relationships. This led to fully committing time and resources to prioritize bottleneck innovations and hence, to develop, implement and monitor evidence-based scaling strategies.

Causal mechanism 2 was supported by the Cassava Flash Dryer case since the definition of context-specific innovation packages and their Scaling Readiness assessment set the stage for the prioritization of key bottlenecks from the newly included complementary innovations. The causal mechanism is partially manifested in the OFSP case as the definition of context-specific innovation packages facilitated a greater appreciation of relevant complementary innovations needed for the scaling

Table 1 Summary of empirical testing of the causal mechanisms and reflections on non-supported causal mechanisms

Causal mechanism	Flash dryer case	Reflection on the flash dryer case	OFSP puree case	Reflection on the OFSP case
(1) Capacity building leads to investments in Scaling Readiness implementation	Supported	N/a	Not Supported	Established project routines prevented allocation of sufficient resources to the effective implementation of Scaling Readiness
(2) Scaling Readiness assessment guides investment on scaling bottlenecks	Supported	N/a	Partially supported	Innovation packages design and bottleneck identification did not inform (re-)investment of project resources
(3) Scaling Readiness strategic options lead to better scaling decisions	Supported	N/a	Partially supported	Established working relationships inhibited the realization of identified strategic options
(4) Scaling Readiness stakeholder tools guide fit-for-purpose partnerships	Supported	N/a	Partially supported	New partnerships were established, but existing partnerships were not changed
(5) Scaling Readiness supports agreement on scaling strategies	N/a	Existing evidence was inconclusive	Supported	N/a
(6) Implementation of agreed strategies improves scaling performance	Supported	N/a	N/a	Existing evidence was inconclusive

of OFSP Puree but could not directly be related to the prioritization of bottleneck innovations.

Similarly, the Flash Dryer case lends support to *causal mechanism 3* whereby the exploration of options to overcome the prioritized bottlenecks informed decisions to work on improving the readiness of the bottleneck innovations in DRC and withdrawal of planned scaling activities in Nigeria and Colombia. The fact that proposed scaling action plans in the OFSP were based on the newly defined innovation packages partially supports *causal mechanism 3*. However, existing partnerships, which came early in the sequence evidencing, continued to exert their influence in shaping decisions regarding proposed scaling investments and actions.

Findings regarding partnerships in the Flash Dryer case aligned with *causal mechanism 4* as the stakeholder profiling and network analysis supported the characterization and enlisting of broader stakeholders within the implementing partners. The OFSP case offers partial support to the causal mechanism there was a strong focus on the continuation of existing partnership trajectories which influenced the overall scaling investments and action plan design.

The scaling strategy agreement process of the OFSP case supports the validity of *causal mechanism 5* as deliberations between stakeholders facilitated agreements on the proposed activity plans. The Flash Dryer case offers less support to the presence of the hypothesized causal mechanism as deliberations only involved few partners.

In the Flash Dryer case, observed improvements in the innovation readiness and use of the prioritized bottlenecks through a reflexive type of monitoring of scaling activities appear to support *causal mechanism 6*. The same mechanism was could not be observed in the OFSP case because of non-conformity between the assumed (learning-based monitoring) and the actual (accountability-based) monitoring practice.

The differences reflected in Table 1 can be explained in two related ways. First, capacity development on innovation and scaling alone may not be sufficient to revise and redirect the country-focus, committed partnerships and activity plans. This supports the literature arguing that changes in knowledge alone are often insufficient for changes in action (Hanisch and Wald 2011; Strohhecker 2016). One may know that a partnership is not entirely fit-for-purpose, but terminating the actual partnership agreement comes with broader considerations and implications. Second, partial implementation of the Scaling Readiness activities in the OFSP project may have limited to contribution of the approach to influencing the design and implementation of the OFSP scaling strategies (Engwall and Jerbrant 2003; Hendriks et al. 1999). Scaling Readiness was mainly used as a stakeholder engagement tool, and did not contribute to (re-)investing project resources in tackling identified scaling bottlenecks in the different countries; a key element of the Scaling Readiness approach.

Given the crucial importance of the first causal principle ("capacity building leads to investments in Scaling Readiness implementation"), more attention needs to be given to understanding what other elements (beyond capacity building) influence whether or not new knowledge, evidence or insights actually lead to change in scaling decisions and action that Scaling Readiness intends to support. The next section tries to answer this and other questions.

Suggestions for Improving the Scaling Readiness Approach

Increased Attention for Scaling Mindset and Impact Culture Growth What was evident in one of the R4D intervention cases was that the Scaling Readiness capacity development approach does not necessarily lead to the type of *systemic* view it intends to foster. Scaling Readiness was appreciated as a stakeholder engagement approach with the intervention team having unanswered questions on the added value of focusing on broader scaling processes rather than project deliverables. '*Understanding the key principles and concepts underlying scaling of innovation*' requires a shift in perspective on innovation and change processes, including change in (organizational) impact culture (Woltering et al. 2019; Leeuwis et al. 2018). This type of change in mindset and culture is much more demanding and sometimes threatening since it involves questioning and perhaps letting go of the basic certainties, strategies, goals and values that one acted upon previously (Van Mierlo et al. 2010). In this regard, changing established practices and structures that underly (and stimulate) certain routines, strategies, and dominant ways of doing things should accompany capacity development of individuals and teams (Woltering et al. 2019; Smith 2007; Argyris and Schön 1996). Moore et al. (2015) refer to this as scaling deep, where there is increased attention for questioning the norms, values and cultures that underly problem and solution framing (e.g. how more meaningful innovation scaling can be supported?).

Dealing with Path-Dependency in Partnerships and Coalition Formation Existing partnerships are influenced by the factors related to the intervention and other factors independent of the intervention. They have a significant influence on strategy design and coalition formation, with a marked tendency to dwell on existing working ties in one of the intervention case contexts. Lamberg et al. (2008) noted that stakeholders' interests, identities, demands, power and structural relations create the boundaries for partnerships, while limiting operational and strategic options for interventions. Given the likely strong influence of broader partnership trajectories on intervention decisions, CSDSA can provide a greater service by supporting transparent deliberation and negotiation processes on partnerships. This occurs not only from the perspective of using evidence to support decision-making on a specific intervention agenda, but also on established stakeholder interaction pattern and underlying motives, interests and incentives governing those interactions. For instance, how should Scaling Readiness support coalition formation when introduced into a new intervention context where existing relationships might be costly to change? More sensitivity of the approach to these issues would create the opportunity for R4D interventions to critically assess *whether* and *how* they can continue leveraging some partnerships or break away from others without necessarily risking long-term relationships or compromising their commitment to achieving impact at scale.

Tension Between Existing and Desired Project and Performance Management Practices Apart from the largely dominant accountability-focused project perfor-

mance, incentive and monitoring practices, many CSDSAs emphasize the need for new incentive mechanisms, monitoring and evaluation to support adaptive management and learning (Kohl 2021; Regeer et al. 2016). This may include changing countries, revising activity plans and partnerships and re-allocating budgets based on progressive insights and a changing project implementation context (e.g., as a result of COVID). Tension is likely to arise based on discordance between the prevailing project performance, incentive and monitoring systems and practices (often provided by funders) and the desired monitoring, reflexive learning and adaptive management approach that Scaling Readiness promotes to achieve impact. The value of decision support approaches in reflexive monitoring, learning and adaptive management could be particularly challenged if introduced in an R4D context where there is limited flexibility to amend activity plans, re-allocate budget and change partnerships (Regeer et al. 2016, Connell and Kubisch 1998). This calls for exploring ways to influence institutionalized project and performance management practices, a process that can be seen as an aspect of the broader (long-term) capacity, mindset and culture shift in views on how change happens and/or how it should be incentivized and monitored. To reap the (short-term) benefits of approaches like Scaling Readiness and to reduce tensions for projects using the approach, there needs to be up-front clarity on the space for adaptive management and change. In situations where such space is limited, Scaling Readiness may not be the lead to tangible outcomes.

Reflections on the Theory of Change of Complexity-Sensitive Decision Support Approaches

Explicating and Testing a Theory of Change of CSDSAs

With their focus on R4D programs, different social theory-informed studies have pointed out the challenge of meaningfully evaluating program interventions when there is an absence of a theory of change that describes how they are expected to work (Cieslik and Leeuwis 2021; Douthwaite and Hoffecker 2017). This is also valid for CSDSA whereby articulating a change theory on how the approaches is imperative to help interventions navigate through complexity and/or make evidence-based decisions. This is central for assessing and improving such approaches to better respond to the R4D context in which they are applied and expected to contribute to specific objectives and outcomes (Swaans et al. 2014). In this regard, the process of explicating a theory of change in the form of causal mechanisms has not only allowed a more structured empirical assessment of the Scaling Readiness approach under study but also uncovered causal mechanisms that otherwise remained implicit or unexplained, providing a basis for improving the CSDSA. Surfacing theories that identify 'what works in which circumstances, why and for whom?', rather than merely 'does it work?' are rarely applied in CSDSA. Yet, this established notion for conducting meaningful evaluations is useful for decision-makers operating in complex and frequently place-based interventions (Pawson and Tilley 1997).

Context in 'Complex-Sensitivity' of Decision Support Approaches

The relevance of context (e.g., temporal, spatial, institutional) in causal mechanism-based assessment of theories has been emphasized in realist evaluation literature (Falleti and Lynch 2009; Pawson 2000; Bunge 1997). Our study demonstrated that various established views and practices of agents and institutions (e.g. users' change perspective, M&E approaches, partnership trajectories) influenced the performance of the Scaling Readiness approach under study despite not necessarily being considered in our construction of the theoretical mechanisms. Since it would be impossible to predict all contextual factors in the development of a testable change theory, the evaluation and eventual modification of theoretical assumptions could usefully target the articulation of broader contextual issues that are likely to be relevant for a decision support approach to function as expected. As also highlighted in political science studies, the bounds of applicability of causal mechanisms could be explicitly posited by defining the context or the relevant aspects of the surroundings where the mechanism is expected to operate (Falleti and Lynch 2009; Pawson 2002). In this regard, the testable CSDSA theories should not only describe how their use is expected to lead to decision outcomes but also under which conditions. This could prompt evaluators and implementers of such approaches not to lose sight of the context or be attentive to possible interactions between causal mechanisms and the context that jointly explain the performance of decision-support approaches.

Conclusion

From a conceptual standpoint, eliciting the underlying assumptions of CSDSA into testable theories of change with explicit causal mechanisms proved to be effective evaluation method. Perhaps even more importantly, it also generated valuable insights on how to improve Scaling Readiness and how to enable institutional context to unlock the full potential of the approach. With the growing popularity of CSDSA in the R4D context, empirical assessment of their added value in supporting effective innovation and scaling decision-making is imperative to further refining their conceptual logic and implementation. Taking Scaling Readiness as a case study, our empirical paper unravelled how the approach contributed to intervention decision-making and outcomes, and shaped and was shaped by the intervention context in which it was applied. The approach has, as expected, contributed to making key strategic decisions around innovation packages design, scaling bottleneck prioritization and partnerships and scaling investment. Additionally, contextual factors—including existing partnerships and institutional factors such as project and performance management systems and existing scaling capacity, mindsets and culture—were found to influence its performance and contribution. In the short term, this could be a challenge for approaches like Scaling Readiness to realize their full potential. However, it also presents an opportunity to influence broader institutional and sectoral change, challenge the way R4D is funded and how successful project execution is perceived (e.g. stick to the plan or adaptively manage) and finally, design and implement more

meaningful performance, incentive and monitoring mechanisms and practices that promote reflexive learning and adaptive management.

References

Argyris, C., and D. Schön. 1996. Organizational learning II: Theory, method and practice. *Reading: Addison Wesley* 305 (2): 107–120.

Astbury, B., and F.L. Leeuw. 2010. Unpacking black boxes: Mechanisms and theory building in evaluation. *American Journal of Evaluation* 31 (3): 363–381.

Barrett, C.B., T.G. Benton, K.A. Cooper, J. Fanzo, R. Gandhi, M. Herrero, S. James, M. Kahn, D. Mason-D'Croz, A. Mathys, and R.J. Nelson. 2020. Bundling innovations to transform agri-food systems. *Nature Sustainability* 3 (12): 974–976.

Beach, D. and R.B. Pedersen. 2011. *What is process tracing actually rracing? the three variants of process tracing methods and their uses and limitations', paper prepared for presentation at the American political science association annual meeting* (pp. 1–4). Washington: Seattle.

Beach, D., and R.B. Pedersen. 2019. *Process-tracing methods: Foundations and guidelines*. Michigan: University of Michigan Press.

Blundo-Canto, G., B. Triomphe, G. Faure, D. Barret, A. De Romemont, and E. Hainzelin. 2019. Building a culture of impact in an international agricultural research organization: Process and reflective learning. *Research Evaluation* 28 (2): 136–144.

Bunge, M. 1997. Mechanism and explanation. *Philosophy of the Social Sciences* 27 (4): 410–465.

Cieslik, K., and C. Leeuwis. 2021. Theorising theories of change in international development: What counts as evidence? In *The Politics of Knowledge in Inclusive Development and Innovation*, ed. D. Ludwig and B. Boogaard, 165–180. Milton Park: Routledge.

Checkel, J.T. 2006. Tracing causal mechanisms. *International Studies Review* 8 (2): 362–370.

Cooley, L., and J. F. Linn. 2014. Taking Innovations to Scale. Methods, applications and lessons. https://www.usaid.gov/sites/default/files/documents/1865/v5web_R4D_MSI-BrookingsSynthPaper0914-3.pdf. Accessed Jan 2023.

Connell, J.P., and A.C. Kubisch. 1998. Applying a theory of change approach to the evaluation of comprehensive community initiatives: Progress, prospects, and problems. *New Approaches to Evaluating Community Initiatives* 2 (15–44): 1–16.

Della Porta, D., and M. Keating, M. eds. 2008. *Approaches and methodologies in the social sciences: A pluralist perspective*. Cambridge University Press.

Douthwaite, B., and E. Hoffecker. 2017. Towards a complexity-aware theory of change for participatory research programs working within agricultural innovation systems. *Agricultural Systems* 155: 88–102.

Eastwood, C., L. Klerkx, and R. Nettle. 2017. Dynamics and distribution of public and private research and extension roles for technological innovation and diffusion: Case studies of the implementation and adaptation of precision farming technologies. *Journal of Rural Studies* 49 : 1–12.

Engwall, M., and A. Jerbrant. 2003. The resource allocation syndrome: The prime challenge of multi-project management? *International Journal of Project Management* 21 (6): 403–409.

Falleti, T.G., and J.F. Lynch. 2009. Context and causal mechanisms in political analysis. *Comparative Political Studies* 42 (9): 1143–1166.

FHI 360. 2004. Going to SCALE—System-wide Collaborative Action for Livelihoods and the Environment. https://www.fhi360.org/sites/default/files/media/documents/Going%20to%20SCALE.pdf. Accessed Jan 2023.

Geels, F.W. and J. Schot. 2007. Typology of sociotechnical transition pathways. *Research Policy* 36 (3): 399–417.

Hanisch, B., and A. Wald. 2011. A project management research framework integrating multiple theoretical perspectives and influencing factors. *Project Management Journal* 42: 4–22.

Hendriks, M.H.A., B. Voeten, and L. Kroep. 1999. Human resource allocation in a multi-project R&D environment: Resource capacity allocation and project portfolio planning in practice. *International Journal of Project Management* 17 (3): 181–188.

Horton, D., and R. Mackay. 2003. Using evaluation to enhance institutional learning and change: Recent experiences with agricultural research and development. *Agricultural Systems* 78 (2): 127–142.

Klerkx, L., B. Van Mierlo, and C. Leeuwis. 2012. Evolution of systems approaches to agricultural innovation:concepts, analysis and interventions. *Farming Systems Research into the 21st century: The new dynamic* 457–483.

Kohl, R. 2021. *Crosscutting issues affecting scaling: a review and appraisal of scaling in international development*. Washington DC: Scaling Community of Practice.

Lamberg, J.A., K. Pajunen, P. Parvinen, and G.T. Savage. 2008. Stakeholder management and path dependence in organizational transitions. *Management Decision* 46 (6): 846–863.

Leeuwis, C., L. Klerkx, and M. Schut. 2018. Reforming the research policy and impact culture in the CGIAR: Integrating science and systemic capacity development. *Global Food Security* 16: 17–21.

Maru, Y., A. Sparrow, R. Stirzaker, and J. Davies. 2018. Integrated agricultural research for development (IAR4D) from a theory of change perspective. *Agricultural Systems* 165: 310–320.

Meyfroidt, P. 2016. Approaches and terminology for causal analysis in land systems science. *Journal of Land Use Science* 11 (5): 501–522.

Moore, M.-L., D. Riddell, and D. Vocisano. 2015. Scaling out, scaling up, scaling deep: Strategies of non-profits in advancing systemic social innovation. *The Journal of Corporate Citizenship* 58: 67–84.

Moyo, M., V.D. Truong, J. Simunovic, J. Pankuku, G.O. Abong, F.K. Amagloh, R. Fuchs, A. Magnaghi, S. Rajendran, F. Grant, and T. Muzhingi. 2022. Orange-Fleshed Sweet potato Puree: A Breakthrough Product for the Bakery Sector in Africa. In *Root, Tuber and Banana Food System Innovations*, ed. G. Thiele and M. Friedmann, 145–172. Cham: Springer.

Pawson, R. 2000. Middle-range realism. *European Journal of Sociology/archives Européennes De Sociologie* 41 (2): 283–325.

Pawson, R. 2002. Evidence-based policy: The promise of realist synthesis. *Evaluation* 8 (3): 340–358.

Pawson, R. 2013. *The science of evaluation: A realist manifesto*. London: Sage.

Pawson, R., and N. Tilley. 1997. *Realistic Evaluation*. Thousand Oaks, CA: SAGE Publications.

Pfotenhauer, S., B. Laurent, K. Papageorgiou, and A.J. Stilgoe. 2022. The politics of scaling. *Social Studies of Science* 52 (1): 3–34.

Regeer, B.J., R. de Wildt-Liesveld, B. van Mierlo, and J.F. Bunders. 2016. Exploring ways to reconcile accountability and learning in the evaluation of niche experiments. *Evaluation* 22 (1): 6–28.

Rogers, P.J. 2008. Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation* 14 (1): 29–48.

Rose, D.C., W.J. Sutherland, C. Parker, M. Lobley, M. Winter, C. Morris, S. Twining, C. Ffoulkes, T. Amano, and L.V. Dicks. 2016. Decision support tools for agriculture: Towards effective design and delivery. *Agricultural Systems* 149: 165–174.

Sartas, M., M. Schut, C. Proietti, G. Thiele, and C. Leeuwis. 2020b. Scaling Readiness: Science and practice of an approach to enhance impact of research for development. *Agricultural Systems* 183: 102874.

Sartas, M., M. Schut, P. Pypers, B. van Schagen, J. Newby, C. Velasco, G. Thiele, and C. Leeuwis. 2017. Scaling Readiness: Accelerating the Scaling of RTB Innovations. Four Case Studies That are Piloting the Scaling Readiness Approach. CGIAR Research program on Roots, Tubers and Bananas (RTB). https://cgspace.cgiar.org/handle/10568/90666. Accessed Jan 2023.

Sartas, M., M. Schut, B. van Schagen, C. Velasco, G. Thiele, C. Proietti, and C. Leeuwis. 2020a. Scaling Readiness: Concepts, Practices, and Implementation. CGIAR Research Program on Roots, Tubers and Bananas (RTB). www.scalingreadiness.org and www.rtb.cgiar.org.

Schmitt, J., and D. Beach. 2015. The contribution of process tracing to theory-based evaluations of complex aid instruments. *Evaluation* 21 (4): 429–447.

Schut, M., C. Leeuwis, and G. Thiele. 2020. Science of Scaling: Understanding and guiding the scaling of innovation for societal outcomes. *Agricultural Systems* 184: 102908.

Schut, M., C. Leeuwis, M. Sartas, L.A.T. Andrade, J. van Etten, A. Muller, T. Tran, A. Chapuis, and G. Thiele. 2022. Scaling Readiness: Learnings from Applying a Novel Approach to Support Scaling of Food System Innovations. In *Root, Tuber and Banana Food System Innovations*, ed. G. Thiele and M. Friedmann, 71–102. Cham: Springer.

Smith, A. 2007. Translating sustainabilities between green niches and socio-technical regimes. *Technology Analysis & Strategic Management* 19 (4): 427–450.

Strohhecker, J. 2016. Factors influencing strategy implementation decisions: An evaluation of a balanced scorecard cockpit, intelligence, and knowledge. *Journal of Management Control* 27: 89–119.

Swaans, K., B. Boogaard, R. Bendapudi, H. Taye, S. Hendrickx, and L. Klerkx. 2014. Operationalizing inclusive innovation: Lessons from innovation platforms in livestock value chains in India and Mozambique. *Innovation and Development* 4 (2): 239–257.

Van Mierlo, B., M. Arkesteijn, and C. Leeuwis. 2010. Enhancing the reflexivity of system innovation projects with system analyses. *American Journal of Evaluation* 31 (2): 143–161.

Woltering, L., K. Fehlenberg, B. Gerard, J. Ubels, and L. Cooley. 2019. Scaling–from "reaching many" to sustainable systems change at scale: A critical shift in mindset. *Agricultural Systems* 176: 102652.

Wigboldus, S., L. Klerkx, C. Leeuwis, M. Schut, S. Muilerman, and H. Jochemsen. 2016. Systemic perspectives on scaling agricultural innovations. A Review. *Agronomy for Sustainable Development* 36 (3): 1–20.

Wigboldus, S.A., L.W.A. Klerkx, and C. Leeuwis. 2020. Making scale work for sustainable development: A framework for responsible scaling of agricultural innovations. In *Science, technology, and innovation for sustainable development goals* (pp. 518–543). Oxford University Press.