THEORY-BASED EVALUATION OF INCLUSIVE BUSINESS PROGRAMMES

IDS Bulletin

Transforming Development Knowledge

institute of development studies

Volume 53 | Number 1 | February 2022

institute of development studies

IDS Bulletin

Transforming Development Knowledge

IDS Bulletin The Institute of Development Studies (IDS) delivers world-class research, learning, and teaching that transforms the knowledge, action, and leadership needed for more equitable and sustainable development globally. The *IDS Bulletin* is an open access, peer-review journal exploring emerging international development challenges. It is published in four issues per year and is the flagship publication of IDS. The *IDS Bulletin* aims to transform development knowledge through its unique thematic issues developed by global learning partnerships that bridge academic, practice, and policy discourse.

Publishing Manager/Bulletin Editorial Coordinator
Alison Norwood

Production Editor Beth Richard

Marketing and Production Officer Gary Edwards

Publisher The *IDS Bulletin* is published by the Institute of Development Studies, Library Road, Brighton BN1 9RE, UK.
Tel: +44 (0)1273 606261.

Cover photo A motorcycle used by a micro-entrepreneur to transport milk from Fulani communities to the collection centre (Oyo State, Nigeria).

Photographer © Greetje Schouten

Disclaimer The Publisher and Issue Editors cannot be held responsible for errors or any consequences arising from the use of information contained in this journal; the views and opinions expressed do not necessarily reflect those of the Publisher and Issue Editors, neither does the publication of advertisements constitute any endorsement by the Publisher and Issue Editors of the products advertised.

Online open access **This journal is available online at bulletin.ids.ac.uk. Visit the site to search for articles and authors and register for table of contents email alerts.**

Information for subscribers The *IDS Bulletin* is published in four issues per year.

Advertising enquiries Gary Edwards, Marketing and Production Officer, IDS Communications and Engagement Unit, **idsbulletin@ids.ac.uk**

www.ids.ac.uk
IDS is a charitable company limited by guarantee and registered in England (No. 877338).

IDS Bulletin
Transforming Development Knowledge
Volume **53** | Number **1** | February 2022

Theory-Based Evaluation of Inclusive Business Programmes

Issue Editors **Giel Ton and Sietze Vellema**

Acknowledgements
This *IDS Bulletin* would never have seen the light without the moral support and analytical insights of the Centre for Development Impact (CDI) and the Partnerships Resource Centre (PrC). CDI is a community of practice that reflects on appropriate, informative, and learning-oriented impact evaluation designs. It is a joint initiative between the Institute of Development Studies, the consultancy company Itad Ltd, and the University of East Anglia. PrC is an international research and knowledge centre for public–private collaboration for sustainable and inclusive development, embedded within the Rotterdam School of Management of the Erasmus University, the Netherlands. It engages with cross-sector partnerships through action research and supports strategizing and evaluative reflection processes.

Funder acknowledgements
This *IDS Bulletin* was financially supported by the 2SCALE programme, funded by the Netherlands Ministry of Foreign Affairs.

Vol. 53 No. 1 February 2022:
'Theory-Based Evaluation of Inclusive Business Programmes'
DOI: 10.19088/1968-2022.100

Notes on Contributors

Just Dengerink is an independent food systems consultant based in Amsterdam, the Netherlands. He supports development organisations to build capacity on the food systems approach and to apply its key principles in their work. Before becoming an independent consultant, Just worked for five years for Wageningen University & Research (WUR), the Netherlands, as an advisor in sustainable value chain development, carrying out food systems analyses and conducting impact evaluations of sustainable agriculture programmes in Africa, Asia, and Latin America. Just holds an MSc in Sustainable Development from Utrecht University and an MSc in International Relations from the University of Amsterdam.

Marijn Faling is a Research Fellow in Partnerships and Inclusive Development. She works with the Partnerships Resource Centre (PrC), a research centre focused on bridging science and practice, based at the Rotterdam School of Management, Erasmus University in the Netherlands. Marijn works with firms, non-governmental organisations, and governments to support interventions and policy processes through Action Research. She is interested in (systemic) change processes to address pressing societal challenges, including poverty, climate change, and food insecurity. She works with rigorous qualitative methods including process tracing and systematic literature reviewing. She has a background in political science and international development studies and holds a PhD from WUR.

Gordon Freer is a part-time lecturer in international relations at the University of the Witwatersrand, South Africa. He has over 20 years of international experience in monitoring, evaluation, and research of developmental projects, including those focused on private sector development, fragile and conflict-affected areas, and climate resilience. Gordon presents regularly at international conferences. He has a specific interest in theory-based evaluations and has published in this area.

Edward Hedley is a Principal Consultant and Evaluation Director at Itad. He has over 14 years' experience in evaluating international development programmes and he specialises in the areas of impact investing, value chain development, and market systems strengthening. In his work he deploys a range of approaches including theory-based evaluation designs and mixed methods. He worked closely with Gordon Freer to design and implement the evaluation of the Samarth-Nepal Market Development Programme (Samarth-NMDP), which drew on an adapted form of contribution analysis to capture the complexity of change at different levels of the market system.

Andrew Koleros is a Senior Researcher at Mathematica in the United States, with expertise in designing and delivering mixed-methods evaluations and programme monitoring, evaluation, and learning systems. He brings particular expertise in using theory-based approaches that integrate complexity concepts and systems thinking into programme and evaluation design processes. Andrew has also supported multiple teams in designing strategy-level theories of change to inform their monitoring, evaluation, and learning initiatives. He has published on his work in designing theory-based evaluation approaches in the *American Journal of Evaluation*, the *Canadian Journal of Program Evaluation*, and in multiple practitioner settings.

Jake Lomax is Director of 3sd Research. As well as undertaking consulting work for clients which include the World Bank, Stockholm Environment Institute, Itad, and NIRAS-LTS International, 3sd Research produces practitioner-oriented processes and tools for mapping systems, analysing conflict, designing strategies for systemic change, and evaluating systemic change programmes. Jake is also a Senior Associate at Tandem and a Senior Associate at Springfield Centre. He has a PhD in International Development and an MSc in Research Methods for International Development, both from the University of East Anglia, and a BSc in Management from the University of St Andrews.

Haki Pamuk is a development economist and impact researcher at Wageningen Economic Research, within WUR. His research focuses on technology adoption, finance, community-based organisations, and private sector development in developing countries and their food systems. He has managed field research in Tanzania, Kenya, Bangladesh, and Burundi. Haki joined Wageningen Economic Research in September 2017, and previously held a position as postdoctoral researcher in the Development Economics Group at WUR. Haki received his MA and PhD degrees in Economics from Tilburg University, the Netherlands. His research has been published in peer-reviewed journals such as *Journal of Development Economics*, *Food Policy*, *Journal of Development Studies*, and *Journal of African Economies*.

Fédes van Rijn is a Senior Researcher at WUR. Her educational background lies in business administration and development economics. She obtained her PhD on the topic of agricultural extension, social capital, and impact assessment from WUR in 2014. Fédes' research interests relate to agricultural research and development, private sector support, and sustainability standards in tropical value chains. She specialises in the design and implementation of evaluations that meet scientific standards while at the same time maximise learning. She has extensive experience in field research and stakeholder consultation in multicultural and multi-stakeholder settings.

Greetje Schouten is a Senior Research Fellow at the PrC, based at the Rotterdam School of Management, Erasmus University, the Netherlands. Through her research, she aims to contribute to a deeper understanding of the role of business – in collaboration with governments and/or civil society – in the area of sustainable and inclusive food and agriculture, especially in the context of low- and middle-income countries. Most of the research projects she is involved in have an Action Research component. Before joining the PrC, Greetje worked as a postdoctoral researcher at WUR.

Ben Taylor is chief executive officer of Agora Global and is an expert in systemic approaches to development. He has played a leading role in the conceptualisation, codification, and implementation of market systems development globally. He focuses on overcoming challenges in the implementation of the approach to new development objectives such as health or education, new contexts such as refugee camps, or new methodologies including the measurement of impact and evaluation. He has extensive experience in programme and intervention design, monitoring and evaluation, and technical advisory services. He currently holds academic positions on the boards of several journals, guest lectureships, and a Senior Research Fellowship at Durham University.

Jodie Thorpe is a Research Fellow in the Business, Markets and the State Cluster at the Institute of Development Studies (IDS), University of Sussex, UK. Her work explores the relationships between market actors, and between the market and the state, to understand what drives business investment that is more or less inclusive. She is interested in the governance processes and institutional structures that mediate these interactions, including more deliberative and democratic economic arrangements. Prior to joining IDS, she worked as a Private Sector Adviser at Oxfam GB and a Director at SustainAbility, a leading global consultancy on business risks and opportunities from sustainable development.

Giel Ton is a Research Fellow at IDS, University of Sussex, and Director of the Centre for Development Impact (CDI). He specialises in the design of mixed-methods impact evaluations in private sector development programmes. He promotes contribution analysis as an overarching approach of theory-based evaluation and a stepwise process to identify the hotspots where additional data collection and reflection are needed. He has a special interest in the effectiveness of programmes that aim to improve governance and coordination in agricultural value chains and empower smallholder farmers in collective action.

Sietze Vellema is Associate Professor in the Knowledge, Technology and Innovation group at WUR. He leads Action Research with the 2SCALE programme at the PrC, based at the

Rotterdam School of Management and is a Research Fellow at the Centre for Frugal Innovation, both at the Erasmus University, the Netherlands. Sietze is editor-in-chief of the interdisciplinary journal *NJAS: Impact in Agricultural and Life Sciences*. He is involved in integrative studies that aim to understand why and how different actors collaborate in solving organisational, managerial, and technical problems related to inclusive development and sustainable food provision, based on (Action) Research in Africa and Southeast Asia.

Introduction: Contribution, Causality, Context, and Contingency when Evaluating Inclusive Business Programmes*

Giel Ton[1] and Sietze Vellema[2]

Abstract The private sector has become an important partner in development interventions that aim to make market systems more favourable for smallholders and low-income consumers of food. How to evaluate these inclusive business programmes is the central theme of this *IDS Bulletin*. It presents real-world experiences of practitioners and academics using theory-based evaluation. This introductory article highlights the approaches and methods used to assess systemic change and provide learning for adaptive management. It acknowledges the limits to attributing outcomes to programmes alone and proposes a way to generalise about effectiveness where outcomes are highly contingent on a specific contextual embedding. The article explores the synergy of the iterative reflections on the theory of change, the analytical approach of realist evaluation, and the conceptualisation of changes in firms' practices as emerging from behaviour systems where the motivations, opportunities, and capabilities of firms are not equally distributed.

Keywords value chains, impact evaluation, market systems, realist evaluation, contribution analysis, theory-based evaluation.

1 Introduction

Theory-based evaluations (TBEs) are widely used to evaluate complex development programmes in dynamic environments. They reflect on the logic of the theory of change of programmes, and on the way that inputs are used for activities that translate into outputs, outcomes, and impact. Process evaluation focuses especially on the links between the inputs, activities and outputs. Impact evaluations help these reflections by verifying and qualifying the contribution claims at outcome and impact level. However, there are methodological limits to doing so.

The *IDS Bulletin* is published by Institute of Development Studies, Library Road, Brighton BN1 9RE, UK
This article is part of *IDS Bulletin* Vol. 53 No. 1 February 2022 **'Theory-Based Evaluation of Inclusive Business Programmes'**.

In this *IDS Bulletin*, we discuss experiences of practitioners and academics in finding doable and creative ways to conduct impact evaluations of inclusive business programmes in the domain of food and agriculture. Classic impact evaluation designs that start at baseline and proceed with a follow-up design to measure quantitative net effects in response to an intervention can be useful in a broader mix of methods in order to reflect on the theories of change (Vaessen, Lemire and Befani 2020) – but only where direct attribution of outcomes to the programme activities is possible and plausible (Ton, Vellema and Ge 2014). Attributing 'net effects' to a support programme is inappropriate for outcomes that lie beyond the sphere of influence. Other more reflexive theory-based approaches of impact evaluation are needed to assess effectiveness.

This introductory article discusses several experiences of impact evaluations that tried to develop these alternative approaches. It presents what was learnt from these evaluations about the tools and methods used to produce credible and actionable insights. It outlines an approach that combines the iterative reflection on the theory of change; the analytical approach of realist evaluation to explain the contextual embedding of change processes; and the conceptualisation of business strategies as behaviour that emerges from a complex system of incentives, and where motivations, opportunities, and capabilities are unequally distributed: some firms will benefit more than others. The article first reflects, in Section 2, on the literature that discusses inclusive business as an avenue for realising development outcomes in the domain of food and agriculture. Next, it focuses on the issues of contribution and causality (Section 3), and context and contingency (Section 4). The article ends with a reflection in Section 5 on how theory-based evaluation can become more meaningful for learning.

2 Inclusive business and development

Working with businesses to achieve development impacts has become an important strategy that is consistently featured in development policies and interventions proposed by international donor agencies, development organisations, and national governments (Heinrich-Fernandes 2016). Many of these programmes focus on the agricultural sector (Osorio-Cortes and Albu 2021). The *Rural Development Report 2021* of the International Fund for Agricultural Development (IFAD 2021) envisions a transformation of food systems driven by vast networks of mid-stream agri-food entrepreneurs, connected to small-scale farmers and consumers. The 2021 UN Food System Summit also confirmed the strong involvement of the private sector in achieving the Sustainable Development Goals.

Policies and intervention strategies labelled as value chain, market system, or food system approaches involve business partners in working towards a more inclusive development, where

smallholders are favourably integrated and low-income consumers have access to affordable and healthy food (Pouw, Bush and Mangnus 2019). We use the term 'inclusive business' programmes when we refer to these development approaches. The ideas about what inclusive business entails vary, and the reality of inclusive business is characterised by contingency and specificity (German *et al.* 2020). In general terms, inclusive business refers to including low-income communities in business chains (Likoko and Kini 2017), and it assumes that vulnerable, small-scale actors benefit through their integration into the (agri)business value chain (German *et al.* 2020; Schoneveld 2020).

There is a growing body of impact evaluations of inclusive business development interventions. The 2021 BEAM Evidence Review synthesises these in indicators aiming to compare their effectiveness and cost efficiency (Osorio-Cortes and Albu 2021). BEAM Exchange is a community of practice around the theme Building Effective and Accessible Markets. Also, the International Social and Environmental Accreditation and Labelling Alliance (ISEAL) has a repository of literature in the online Evidensia database[3] that looks at the effectiveness of market-based coordination mechanisms, especially in tropical commodity chains.

At a more general level, the Donor Committee for Enterprise Development (DCED) tries to get harmonised indicators for private sector development programmes that help to reflect on the contribution of businesses to development impacts (DCED 2017). This continuing focus on comparison for accountability is, however, not always well aligned with the objective of learning from these interventions, due to the complex nature, multiple components, and different types of support activities that are being compared. A focus on the contextualised understanding of the unfolding change processes from which outcomes emerge may reveal why some groups are benefiting more than others from the support.

There is widespread support in the literature for increasing the coordination between the public and private sectors to develop new business models that have positive developmental outcomes. Scholars use different lenses to analyse these inclusive businesses. The value chain lens looks at standard setting, coordination, sourcing arrangements, input delivery, and service provision at sector level (Stoian *et al*. 2012). The value chain literature helps to find models and modalities that are conducive to pro-poor and sustainable development along the entire value chain (Ros-Tonen *et al.* 2019), from the upstream, smallholder producers to the downstream, 'bottom-of-the-pyramid' consumers (Maestre, Poole and Henson 2017).

A recent review of this literature emphasises the contextual embeddedness of interventions and proposes a processual perspective for exploring how inclusive business 'refashions'

relations and partnering among and across market actors and public actors (Schouten and Vellema 2019). This dynamic processual understanding is also present in literature on entrepreneurship. Entrepreneurs need to find modalities to connect and coordinate behaviour of multiple actors, ranging from smallholder farmers (Kangogo, Dentoni and Bijman 2021) and youth (Barzola *et al.* 2019) to community-based enterprises (Dentoni *et al.* 2018), in order to achieve a resilient business model.

A related strand of literature calls for recognition and appreciation of the activities of economic actors that operate in the middle of agri-food chains or food provisioning (Liverpool-Tasie *et al.* 2020; Reardon 2015). The rules and practices of businesses operating in between smallholder producers and urban markets importantly shape the terms of access to markets for farmers and micro-entrepreneurs, and influence the balance between exploitation and rewarding their inclusion. This body of literature shows that value chains, businesses, and markets cannot be treated as homogenous entities, and using these concepts – as we do in this article – has the danger of simplification.

The literature also shows heterogeneous effects of inclusive business. The claims of businesses that they contribute to inclusive business and desirable development outcomes are not self-evident. Development impacts are conditional on evolving change processes that emerge in diverse actor constellations, which operate in dynamic market and natural environments. Consequently, programmes and evaluators cope with moving targets in uncertain and unpredictable market and business environments, creating conditions of constant flux. Meanwhile, they need to contribute to previously agreed impact domains, such as food and nutrition security or environmental sustainability.

To be impactful, inclusive business programmes need multiple and overlapping interventions that are implemented in collaboration with multiple partners in a highly dynamic complex business environment. Therefore, to learn about their effectiveness, evaluations will need to find a way to acknowledge the co-existence of mutually constituting practices and unravel the interdependence and interaction among mutual causal processes.

Furthermore, private sector programmes and business practices are unlikely to be the sole contributing factors to observed development processes. The behaviours, practices, and rules associated with the intervention or business generate responses in a web of interdependence. Hence, the level of control over development impacts is limited, and having impact goes beyond the programmes' sphere of direct influence. A challenge for monitoring and evaluation of these programmes is how to account for the contingency and uncertainty that are inherent in such unfolding change processes in the broader system.

Moreover, evaluations of inclusive business programmes need to be attentive to signs of changes that affect and reshape rules and practices which underlie the nature of doing business: the issue of systemic change. Likely, it is only possible to assess systemic change in markets after some years, when the changed business models or value chain coordination modalities have matured and proven sustainable at scale.

This *IDS Bulletin* presents several experiences and approaches for monitoring and evaluating development outcomes and systemic change in inclusive business programmes. For example, Taylor and Lomax (this *IDS Bulletin*) propose to capture these systemic outcomes with the Adopt-Adapt-Expand-Respond (AAER) framework, originally developed in the Springfield Centre (Nippard, Hitchins and Elliott 2014). Monitoring the reactions of stakeholders and their behaviour in reaction to pilot interventions, the AAER framework captures unintended outcomes; the Adopt and Expand quadrants are more geared to the intended outcomes of an intervention. Vellema, Schouten and Faling (this *IDS Bulletin*) describe how they developed a tool to collect comparable data from more than 60 partnerships to capture early signs of systemic change. Based on a selection of theoretical frameworks, they typify three categories of outcomes to support implementers and partnership facilitators in noticing and valuing the effects of unfolding systemic change towards more inclusivity: the refashioning of the terms of inclusion of smallholders, the access of low-income consumers to food, and innovative leadership of the private sector that is doing 'business as unusual' (Vellema *et al.*, this *IDS Bulletin*). Hedley and Freer (this *IDS Bulletin*) use the Qualitative Impact Protocol (QuIP) to scope for early signs of transformative change in markets.

The verification and critical assessment of contribution claims related to (early signs of) systemic change depend on the data that become available before, during, and after the implementation of the programme, the quality of the analysis and synthesis process, and the 'sense-making' about that evidence. Even when evaluation approaches can differ substantially in their ontological and epistemological assumptions (Stern *et al.* 2012), they all need data to generate the inputs for the sense-making process and causal inference process. The contributions by both practitioners and academics in this *IDS Bulletin* (see Table 1) respond to the call by Barbrook-Johnson *et al.* (2021: 5) to enlarge the toolbox available for complexity-appropriate evaluation.

We highlighted the methodological innovations in each of these articles that have potential to make impact evaluations more complexity-aware. At the same time, we acknowledge that the authors' accounts, including ours, of how the impact evaluations were designed and used are inevitably biased, as most of the authors operate 'in an environment where there are significant incentives to appear competent, minimise the

Table 1 Overview of the articles and programmes discussed in this introduction article

Authors	Name of programme	Nature of the interventions	Methodological innovations	
			Contribution and causality	Context and contingency
Taylor and Lomax	Generic	Market system development programmes	Nested and interlocked market systems	The Adopt-Adapt-Expand-Respond (AAER) framework captures ripple effects of pilot interventions in the wider market system
Hedley and Freer	Samarth-Nepal Market Development Programme (Samarth-NMDP)	Market system development in Nepal, especially in vegetable and dairy value chains	Contribution analysis through top-down and bottom-up research	Qualitative Impact Protocol (QuIP) captures unbiased perceptions of impacts and change processes
Ton, Taylor and Koleros	Private Enterprise Programme Ethiopia (PEPE)	Market system development, especially in the leather, vegetable, and cotton sector, including labour sourcing in industrial parks	Interlinked research design, using firm-level surveys, process tracing case studies, and macro-economic modelling	Flexible results monitoring system using actor-based theories of change
van Rijn, Pamuk, Dengerink and Ton	Pioneering Real-time Impact Monitoring and Evaluation (PRIME)	Coaching and training of small and medium-sized enterprises to improve business management and export capacities	Online survey module to ask perceptions of impact and compare contribution scores on a range of outcomes	Real-time monitoring in a setting of dynamic navigation
Vellema, Schouten and Faling	2SCALE	Facilitating more than 60 partnerships for the (scaling of) inclusive agribusinesses fostering food and nutrition security in Africa	Using structured impact pathways embedded in each partnership for spelling out the sequential change processes	Contextualised monitoring of early signs of systemic change
Faling	Community Revenue Enhancement Through Agricultural Technology Extension (CREATE)	Linking smallholders to the barley and beer value chain in Ethiopia	Assessing pieces of evidence to verify a contribution claim, using process tracing	Supporting business partners in exploring their span of influence in a sector or industry
Thorpe	Developing Effective Private Education Nigeria (DEEPEN)	Improve the quality of education provided by private schools in Lagos	Using COM-B model to shift focus to behavioural change of firm	Graphical way to distinguish outcomes at different system levels using the COM-B model
	Financial Sector Deepening Trust Kenya (FSDK)	Generate sustainable livelihood improvements through better financial sector capacity and operations		
	Gender Transformative and Responsible Agribusiness Investments in South-East Asia (GRAISEA)	Improve livelihoods of women and men small-scale producers through more responsible and inclusive value chains and private sector investments		

problems, and to make things neater than the real messy process' (Rogers and Peersman 2014: 93). We remind the reader, therefore, to be cautious, and not to see any of the presented approaches and tools as the 'silver bullet' that resolves the challenges of the impact evaluation of inclusive business programmes.

Below, we elaborate major monitoring and evaluation (M&E) challenges and the possibilities to address these under two headings: **Contribution and causality** (Section 3), pointing to the co-existence of multiple intertwined causal processes; and **Context and contingency** (Section 4), pointing to the unpredictability and uncertainty of markets and the behaviour systems of those operating in these markets.

3 Contribution and causality

Lemire, Whynot and Montague (2019) give a nice overview of the increasing complexity of systems that are present in change processes. They present the spectrum of causal complexity in programme theories, where the simplest one is described as 'A leads to C', and the most complex one (the embedded-complex version) as 'A plus B leads to C because of D, under condition E'. Inclusive business programmes are clearly to be characterised as the latter, the embedded-complex ones; inclusive business programmes intentionally try to trigger changes that depend on other contextual conditions and incentives, apart from the support provided by them. A programme is (at most) a contributory factor in the process of generating inclusive business outcomes. Mackie (1974: 63) would call them an INUS factor – an 'insufficient but non-redundant part of a condition

Box 1 How to address INUS factors?

- **Insufficient factor:** Acknowledge that other conditions need to be in place for the programme support to work – a complex change process that the programme could not create alone.
- **Non-redundant factor:** Verify whether the intervention is only 'accompanying' a causal process that would have been in place and created the outcome, without the support provided by the intervention playing any role of importance.
- **Unnecessary configuration:** Acknowledge that other configurations (where the intervention does not take part) might exist that could also have resulted in the same outcome.
- **Sufficient configuration:** Verify whether the outcome indeed happened at all and can be plausibly linked to the change process that has been supported.

which is itself unnecessary but sufficient for the result' (see Box 1). For an impact evaluation that follows a theory-based evaluation approach, we need contribution-verifying methods (did the intervention contribute to the change process?), and methods that reflect on the importance of this contribution (did it matter?). We consider these two questions in turn.

3.1 Did the intervention contribute to the change process?

Contribution-verifying methods need to show that the intervention is not redundant in a more complex change process that might not have taken place without the intervention. Firms develop new or refined service delivery models to include smallholders or to reach low-income consumers. That is the more complex change process, involving input providers, knowledge, and financial services, etc. The question in contribution-verifying research is whether the activities or resources of the support programme have played a non-redundant role in this change: would it have happened anyhow, without the support?

Faling (this *IDS Bulletin*) presents a nice example of process tracing as a way to critically verify whether a systemic change in the value chain of beer (crowding in of other malt factories) is indeed causally related to earlier support provided in the sector by an inclusive business programme. Though inspired by Bayesian reasoning (updating our confidence in the claim as more pieces of evidence become available), fortunately, Faling does not go as far as to compute the probability in a quasi-quantitative way, as recent process tracing literature suggests (Bennett, Charman and Fairfield 2021), and which Befani and Stedman-Bryce (2017) have coined as contribution tracing.

Bayesian updating starts with an estimate for the belief before a piece of evidence is considered (the prior probability), and results in an estimate that incorporates this new knowledge (the posterior probability). But, an informative evaluation of an inclusive business programme does not answer only one question – whether the programme is non-redundant in a change process – nor the probability that each arrow in the theory of change is true. Therefore, Bayesian updating might be a sophisticated but too narrow method to feed reflections about the effectiveness of inclusive business programmes.

However, what is clear is that a process tracing exercise implies a systematic process of seeking and critically assessing evidence. Faling (this *IDS Bulletin*) illustrates how process tracing can be used as a practical approach for explicating and scrutinising key assumptions in their contribution claims.

Hedley and Freer (this *IDS Bulletin*) piloted, among a wider set of top-down and bottom-up methods used, another tool to verify the contribution of the market development programme Samarth in Nepal. The tool, called Qualitative Impact Protocol (QuIP)

(Copestake, Morsink and Remnant 2019), implies that researchers talk with intended project beneficiaries about the main changes in their lives over a pre-defined recall period, and are prompted to share what they perceive as the main drivers of these changes, including to whom or what they attribute the change. The premise of QuIP is that the intended beneficiaries know a great deal about what has caused and affected changes in their lives, and what influenced their active decisions to start or stop doing certain activities (Copestake *et al.* 2019: 4). QuIP is explicitly not interested in inferring average effects but aims to explain or explore variation in the wellbeing outcomes.

Evaluators operate in an environment that creates strong incentives to look for confirming evidence (Rogers and Peersman 2014). QuIP is particularly keen to avoid this confirmation bias and is, therefore, often 'blindfolding' the researchers that do the interviews in a way that these do not know who commissioned the study or what support intervention is being evaluated. However, QuIP does not necessarily imply blindfolding. In the QuIP application in Samarth, documented by Hedley and Freer (this *IDS Bulletin*), the impact evaluation wanted to learn about the implementation modalities of service providers in value chains; they were less focused on assessing the outcomes of the changed value chain relation for the final beneficiaries (the smallholder producers). They learned in the process that evaluators needed to be sufficiently knowledgeable about the intricacies of the support and the heterogeneous effects that these modalities may have for different stakeholder groups in order to ask pertinent questions and unearth practical learning.

3.2 Did the contribution matter?

The market system development programmes discussed in this *IDS Bulletin* are all multi-year and well-resourced programmes with many activities, and are ready to respond to emerging issues and bottlenecks in the markets. Consequently, there will almost always be a clear (uncontested) non-redundant contribution to one or more intended outcomes, significant or insignificant as these outcomes might be. There will always be one or more firms that have changed their business models due to the increased market intelligence, capabilities, or opportunities that come with the support. Simply showing a contribution is, therefore, not enough. The question of the importance of this contribution needs to be addressed and is often the main reason why impact evaluations of inclusive business programmes are commissioned.

There are some principles and methods that help to reflect on the importance of contributions. Most importantly perhaps is that the commissioners of evaluations acknowledge that exact numbers are not needed but rough estimates suffice. For example, the Dutch Directorate-General for International Development[4] (DGIS-RVO 2017) accept an estimate of the number of jobs or smallholders that were directly and indirectly supported by

an intervention, asking for estimates of reach instead of net effects. This opens the way for modelling the likely impacts of a programme according to several scenarios and based on explicit normative assumptions. A computable general equilibrium model, for example, could be used to estimate how the economy reacts to a plausible range of low and high growth rates in specific subsectors. The assumptions used in these extrapolations of effects are sensitive to normative decisions about the model parameters, especially when these models are built specifically for the evaluation.

Recently, the development finance institutions agreed upon a harmonised model, the Joint Impact Model (Steward Redqueen 2021). Even when inherently speculative and inexact, the use of a joint model would yield comparable indicators of development impact to judge the importance of programmes within and across countries. Instead, Mayne (2019a) argues that, rather than measure the size of the importance quantitatively, we need to ask the question: what is the relative importance of a specific causal factor in a wider configuration of factors? He suggests several ways of data collection that may help to (normatively) judge the importance of an intervention, after it is verified as being a non-redundant causal factor (Mayne 2019a: 5–6): (1) the perceived influence of the causal factor in bringing about a change; (2) the role played by the causal factor in bringing about a change; (3) the funds expended by the causal factor; and (4) the magnitude of the constraint to change faced by the causal factor. In doing so, Mayne refrains from assessing the importance of a contribution in an objective, quantitative way.

Both van Rijn *et al.* (this *IDS Bulletin*) and Ton, Taylor and Koleros (this *IDS Bulletin*) assess the importance of a contribution in a subjective, quantitative way. They made use of a survey module that explicitly asked for the perception of the managers themselves about the importance of the contribution of the support. The survey tool was developed in response to Robert Chambers' recommendation made in the discussions about rigour in impact evaluation (Chambers 2009): 'If you want to know the impact – Ask them!'.

Van Rijn *et al.* (this *IDS Bulletin*) applied the tool in an impact evaluation of two Dutch programmes that targeted managers of small and medium-sized enterprises in developing countries with coaching and capacity building. They asked the managers about their perception of change and the contribution of the support in eight areas of business management, when comparing themselves with similar firms in the sector.

Ton *et al.* (this *IDS Bulletin*) applied a similar tool to collect perception of change in perceived constraints in the institutional environment that affect business performance. The Private Enterprise Programme Ethiopia (PEPE) worked in several sectors

of the Ethiopian economy and aimed to trigger innovation processes that would improve sector performance. The online survey module does not directly mention PEPE but asks the managers about the influence of the relevant service providers or institutional arrangements that had received support from PEPE. The contribution score tool is based on a combination of two rankings of ordinal Likert scale answer categories, one in response to a question about the level of change in a reference period, and a second in response to a question about the importance of the specific support activity to these changes.

The combination of the two answer categories results in a ranking that reflects the importance of the contribution. The overall pattern of the contribution scores per area shows where the contribution is relatively important or less important. Using the INUS wording, the perception questions help to reflect on the non-redundancy and sufficiency of the intervention in change processes. Often, the change is there, but the respondent answers that the intervention played no role at all in the change process.

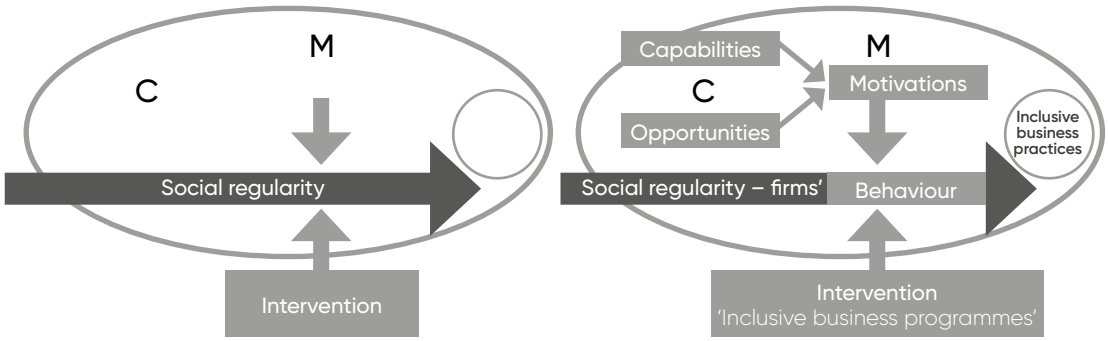
Of course, confirmation bias is still a threat to validity, but in both applications of the tool, a nuanced picture appeared with outcomes where more improvements were perceived, and with outcomes where the perception of change was less positive, which helped reflections about effectiveness of the portfolio of support activities on offer or in development.

These examples show that there are entrance points for improvement. The authors point to the need for more realism on the side of commissioners, requiring less precisely measured assessments of the ultimate outcomes and development impacts. Also, there needs to be (re)valorisation of the use of stakeholder perceptions, instead of observable outputs or outcomes to assess the importance of a programme's contributions to complex causal change processes, but acknowledging the bias that this could have. And finally, there is a need for the deployment of methods to critically assess alternative explanations, including the possibility that the intervention played no or only a marginal role in the change processes. With these 'ingredients' in the mix of methods, it proves possible to generate informative and credible impact evaluations of complex interventions like inclusive business programmes.

4 Context and contingency

The issue of contingency on context is highly relevant for inclusive business programmes, where market systems change due to the resources and reflection offered by the programme to the stakeholders involved – which subsequently triggers these stakeholders to do things differently. It is clear that the outcomes will always depend on the conditions in which these stakeholders are situated, and on the incentives and disincentives they are facing. The role of context is central to the realist evaluation approach,

Figure 1 Integrating COM-B in a realist CMO analysis of changing business practices



Note COM-B is 'capabilities, opportunities, and motivations for behaviour'; CMO is 'context, mechanisms, and outcome pattern'.
Source Authors' own, based on Pawson and Tilley (2006).

where the question 'What works, for whom, under what conditions, and why?' invites the evaluator to qualify in detail the exact role of an intervention in the change process and the conditions that are required for the intended change process to work.

In realist research, all causal mechanisms have a defined generalisation domain (Chen 1994), and need a reflection and analysis of the contextual conditions under which they are triggered (Pawson and Tilley 2006). The realist lens focuses on important components of programmes, acknowledges the heterogeneity of effects within these specific interventions, and identifies the configurations of conditions that enable these components to work well or explain why they fail.

A realist perspective translates into outlining multiple configurations of context, mechanisms, and outcomes. It is important to stress the configurational element, where 'it is not the ingredients that make the dish but how these are brought together in the cooking process' (Pawson and Manzano-Santaella 2012: 189). The object of the realist analysis will change depending on the type of support (resources and reasoning) and the scale where the activity is looking for outcomes. Each programme activity will have a specific outcome pattern, trigger particular mechanisms, and have specific contextual conditions.

However, at each scale level, the change processes can be conceptualised as changing a social regularity that results from a CMO-configuration – where CMO is context, mechanisms, and outcome pattern. The left-hand side of Figure 1 shows Pawson and Tilley's famous egg-shaped CMO-configuration (Pawson and Tilley 2006): an intervention changes the context C in a way that mechanisms M are triggered and change a social regularity resulting in an outcome pattern O.

Acknowledging that there is neither a silver bullet intervention nor a universal causal law in social systems, realists explore the generalisation domain of any conclusion about the 'what works' question. Accordingly, evaluation articulates and incorporates a selection of middle-range theories about the conditions under which an intervention might work elsewhere – and the mechanisms that would explain this. This is visible in how Vellema *et al.* (this *IDS Bulletin*) endeavour to unravel the composite nature of inclusive development and work towards categories of processual outcomes, reflecting the terms on which certain social groups are included in food provisioning. This is a step towards assessing whether the assumed achievement of 'business as unusual' materialises and, in realist terms, whether a social regularity is altered in a way that makes inclusive agribusiness rewarding for smallholder farmers and food affordable for low-income consumers. This helps to answer the 'what works, for whom' question.

To get closer to the actual mechanisms that generate change, a realist approach to impact evaluation of inclusive business programmes can opt to zoom in on choice-making and behavioural processes that generate changes in the form and substance of decision-making (Westhorp 2012, 2013) and subsequently refashion the practices and rules of individual firms or networks of firms. Following the conceptualisation of Michie, van Stralen and West (2011) and Mayne (2019b [2016]), firms or other organisational actors need to have the Capacities, the Opportunities, and the Motivations to make the Behaviour change (the COM-B model). Experiences with the new behaviour feed back into and change the capacities, opportunities, and motivations, as part of a structuration process where structures are reproduced and changed through agency (Giddens 1984). Inclusive business programmes have interventions and activities that aim to influence ongoing social processes (social regularity) by triggering the motivation of firms (mechanisms) to change their business models (behaviour) in a way that markets become better, fairer, and more inclusive (outcome pattern).

Especially for micro-level change in firms, where the outcome is a change in practices, the COM-B model of behaviour change might be useful to operationalise this realist analysis (see Figure 1). The decision-making of firms and farmers involves multiple, often competing, motivations and incentives that emerge due to certain conditions, and changes in these conditions. For example, most business-oriented development programmes discussed in this *IDS Bulletin* do not provide financial resources but knowledge, coaching, or capacity building.

These programmes address the capabilities and motivations of firms. However, in doing so, they experience that even when firms have the right motivation and capability, there must also be the opportunity to enact their behaviour. Thorpe (this *IDS Bulletin*)

presents a graphic representation using the COM-B model to illustrate the findings of three impact evaluations, with critical elements at the macro and meso levels that drive behaviour at the micro level. Ton *et al.* (this *IDS Bulletin*) apply it when they discuss experiences in the PEPE programme, where actor-based theory of change models (Koleros and Mayne 2019) are used to detail intervention strategies and guide the building of a results-based monitoring system (Yohannes 2020; Posthumus *et al.* 2020).

The combination of realist thinking and the conceptualisation of behaviour as a COM-B system offers a fruitful way to develop (nested) theories of change and middle-range theory. This may smooth the path for a learning-oriented approach to impact evaluation that recognises that achieving intended outcomes is contingent on how the interventions configure with conditions.

Some mechanisms (motivations) that lead to inclusive business outcomes are only triggered by inclusive business programmes under the right conditions. Likewise, some mechanisms (motivations) may be triggered that explain why interventions have no results. Accordingly, realist analysis generates actionable insights on the limits of interventions. The realist focus on causal mechanisms-in-context proves particularly relevant when the main role of a programme is piloting, innovating, and experimenting with intervention modalities that are expected to be scaled or implemented in other contexts.

Realist evaluators aim to develop actionable middle-range theories (Cartwright 2020; Pawson and Tilley 2006), around the question 'What works for whom, under what conditions, and why?' The message is that there are no silver bullets – there are no universal laws in social science – but all outcomes are context-dependent and often contingent on many complex, intertwined mechanisms, incentive structures, and motivations that mean that there is a high level of contingency, serendipity, and surprise involved.

5 Making evaluation meaningful

Most impact evaluations of inclusive business or market system programmes use a theory-based evaluation approach (Osorio-Cortes and Albu 2021). This raises the question, 'Whose theories are selected, combined, and refined?' A good theory-based evaluation asks for a critical engagement with a plural set of theories. Smart data collection and sharp analysis and synthesis alone are not enough. The evaluation process and outputs also need to be informative for the stakeholders involved.

Several authors in this issue (van Rijn *et al.*; Ton *et al.*; Hedley and Freer) conclude that more interaction and sense-making between implementers and evaluators is needed. Although we acknowledge that there are more stakeholders than implementing agencies involved in an impact evaluation, the

learning by the implementing agency was for most authors in this *IDS Bulletin* the main goal. But the authors also show that this is the stakeholder group for whom they struggled most to prove the usefulness and value of systematic and rigorous forms of monitoring and evaluation.

Under the right conditions, the presented approaches and tools might work and accelerate the learning loops for adaptive management. Three conditions appear as necessary components in the causal configurations that result in a high-quality theory-based evaluation: (1) interested 'listeners' as the audience of the evaluation, especially the commissioners and implementing agency; (2) rigour in anticipating and addressing validity threats to the conclusions derived from the methods used; and, last but not least, (3) sufficient resources for an appropriate mix of methods.

The inclusive business programmes in this *IDS Bulletin* provide examples of theory-based evaluation approaches that go beyond the tick-box exercises that still characterise large parts of the monitoring and evaluation field. The authors piloted processes and generated outputs that were meant to be functional for learning, and especially the comparative learning about effectiveness and relevance of intervention modalities across a portfolio of supported projects. Vellema *et al.* (this *IDS Bulletin*) discuss critical issues in the set-up of such a learning-oriented evaluation system within the 2SCALE programme that fits the navigation of partnership in dynamic contexts. The PRIME evaluation, described by van Rijn *et al.* (this *IDS Bulletin*), developed tools for real-time monitoring in an organisational setting where a monitoring, evaluation, and learning (MEL) system already existed that needed to be upgraded to meet external reporting requirements.

Other impact evaluations, such as PEPE (Ton *et al.* this *IDS Bulletin*; Koleros 2020; Yohannes 2020) and Samarth (Hedley and Freer, this *IDS Bulletin*) are embedded in the evaluation system of the donor that uses annual reviews and external impact evaluations with baseline, midterm, and endline data collection and synthesis (ICAI 2015), and where the evaluators need to ensure that learning occurs 'within these predictive management standardised tools and templates' (Koleros 2020: 63). However, all experiences presented in this *IDS Bulletin* acknowledge that it is not easy to find ways to make learning useful for commissioners and implementing agencies. The actual use of findings depends on many factors that are beyond the control of evaluators.

Notes

* We thank the contributors to the *IDS Bulletin* for their openness and reflexivity while sharing the lessons learned from designing and implementing evaluation tools in the dynamic setting of inclusive business programmes. This article benefited from the critical review by Dominic Glover and the contributing authors. Publication of the *IDS Bulletin* was supported by the 2SCALE programme.

1 Giel Ton, Research Fellow, Institute of Development Studies, University of Sussex, UK.
2 Sietze Vellema, Associate Professor, Knowledge, Technology and Innovation group, Wageningen University, the Netherlands.

3 For more information see **Evidensia**.
4 In full, Ministerie van Buitenlandse Zaken-Rijksdienst voor Ondernemend Nederland [Dutch Ministry of Foreign Affairs-Netherlands Enterprise Agency]

References

Barbrook-Johnson, P.; Castellani, B.; Hills, D.; Penn, A. and Gilbert, N. (2021) 'Policy Evaluation for a Complex World: Practical Methods and Reflections from the UK Centre for the Evaluation of Complexity across the Nexus', *Evaluation* 27.1: 4–17

Barzola, C.L. *et al.* (2019) 'Challenges of Youth Involvement in Sustainable Food Systems: Lessons Learned from the Case of Farmers' Value Network Embeddedness in Ugandan Multi-Stakeholder Platforms', in R. Valentini, J.L. Sievenpiper, M. Antonelli and K. Dembska (eds), *Achieving the Sustainable Development Goals Through Sustainable Food Systems*, Cham: Springer

Befani, B. and Stedman-Bryce, G. (2017) 'Process Tracing and Bayesian Updating for Impact Evaluation', *Evaluation* 23.1: 42–60

Bennett, A.; Charman, A.E. and Fairfield, T. (2021) '**Understanding Bayesianism: Fundamentals for Process Tracers**', *Political Analysis*, DOI: 10.1017/pan.2021.23 (advance online publication) (accessed 16 November 2021)

Cartwright, N. (2020) 'Middle-Range Theory: Without It What Could Anyone Do?', *Theoria: An International Journal for Theory, History and Foundations of Science* 35.3: 269–323

Chambers, R. (2009) 'Perspectives on Impact Evaluation: Approaches to Assessing Development Effectiveness', video, Cairo: International Initiative for Impact Evaluation (3ie)

Chen, H-T. (1994) *Theory-Driven Evaluations*, Thousand Oaks CA: SAGE

Copestake, J.; Morsink, M. and Remnant, F. (2019) *Attributing Development Impact*: *The Qualitative Impact Protocol Case Book*, Rugby: Practical Action

DCED (2017) *The DCED Standard for Measuring Achievements in Private Sector Development: Control Points and Compliance Criteria*, London: Donor Committee for Enterprise Development

Dentoni, D.; Pascucci, S.; Poldner, K. and Gartner, W.B. (2018) 'Learning "Who We Are" By Doing: Processes of

Co-Constructing Prosocial Identities in Community-Based Enterprises', *Journal of Business Venturing* 33.5: 603–22

DGIS-RVO (2017) *15 Methodological Notes – Instructions for Calculation, Validation and Reporting of Performance Indicators*, The Hague: Ministerie van Buitenlandse Zaken-Rijksdienst voor Ondernemend Nederland [Dutch Ministry of Foreign Affairs-Netherlands Enterprise Agency]

German, L.A.; Bonanno, A.M.; Foster, L.C. and Cotula, L. (2020) ' "Inclusive Business" in Agriculture: Evidence from the Evolution of Agricultural Value Chains', *World Development* 134: 105018

Giddens, A. (1984) *The Constitution of Society: Outline of the Theory of Structuration*, Berkley CA: University of California Press

Heinrich-Fernandes, M. (2016) *How to Create an Enabling Environment for* Inclusive *Business? Learning from Experience and Debates in Private Sector Development*, Business Environment Working Group Paper, London: Donor Committee for Enterprise Development

ICAI (2015) *DFID's Approach to Delivering Impact*, Report 45, London: Independent Commission for Aid Impact

IFAD (2021) *Transforming Food Systems for Rural Prosperity: Rural Development Report 2021*, Rome: International Fund for Agricultural Development

Kangogo, D.; Dentoni, D. and Bijman, J. (2021) 'Adoption of Climate-Smart Agriculture Among Smallholder Farmers: Does Farmer Entrepreneurship Matter?', *Land Use Policy* 109: 105666

Koleros, A. (2020) 'Accountable for Adaption: How Independent Evaluation Can Support Adaptive Programming within Bureaucracies', in B. Perrin and T. Tyrrell (eds), *Changing Bureaucracies: Adapting to Uncertainty, and How Evaluation Can Help*, New York NY: Routledge

Koleros, A. and Mayne, J. (2019) 'Using Actor-Based Theories of Change to Conduct Robust Contribution Analysis in Complex Settings', *Canadian Journal of Program Evaluation* 33.3: 292–315

Lemire, S.; Whynot, J. and Montague, S. (2019) 'How We Model Matters: A Manifesto for the Next Generation of Program Theorizing', *Canadian Journal of Program Evaluation* 33.3: 414–33

Likoko, E. and Kini, J. (2017) 'Inclusive Business – A Business Approach to Development', *Current Opinion in Environmental Sustainability* 24: 84–88

Liverpool-Tasie, L.S.O. *et al.* (2020) 'A Scoping Review of Market Links Between Value Chain Actors and Small-Scale Producers in Developing Regions', *Nature Sustainability* 3: 799–808

Mackie, J.L. (1974) *The Cement of the Universe: A Study of Causation*, Oxford: Oxford University Press

Maestre, M.; Poole, N. and Henson, S. (2017) 'Assessing Food Value Chain Pathways, Linkages and Impacts for Better Nutrition of Vulnerable Groups', *Food Policy* 68: 31–39

Mayne, J. (2019a) ***Assessing the Relative Importance of Causal Factors***, Centre for Development Impact (CDI) Practice Paper 21, Brighton: Intitute of Development Studies (accessed 16 November 2021)

Mayne, J. (2019b [2016]) 'The COM-B Theory of Change Model', Working Paper V5, unpublished

Michie, S.; van Stralen, M.M. and West, R. (2011) 'The Behaviour Change Wheel: A New Method for Characterising and Designing Behaviour Change Interventions', *Implementation Science* 6: 42

Nippard, D.; Hitchins, R. and Elliott, D. (2014) ***Adopt-Adapt-Expand-Respond: A Framework for Managing and Measuring Systemic Change Processes***, Springfield Centre Briefing Paper, Durham: Springfield Centre (accessed 9 September 2021)

Osorio-Cortes, L.E. and Albu, M. (2021) *The Results Achieved by Programmes that Use the Market Systems Development (MSD) Approach: A Narrative Synthesis of Current Evidence*, BEAM Evidence Review 2021, London: BEAM Exchange

Pawson, R. and Manzano-Santaella, A. (2012) 'A Realist Diagnostic Workshop', *Evaluation* 18.2: 176–91

Pawson, R. and Tilley, N. (2006) *Realist Evaluation, Realistic Explanation: Linking Explanatory and Interpretive Accounts in Applied Social Research,* Utrecht and Wageningen: CERES

Posthumus, H.; Shah, R.; Miehlbrandt, A. and Kessler, A. (2020) *A Pragmatic Approach to Assessing System Change*, Durham: Springfield Centre

Pouw, N.; Bush, S.R. and Mangnus, E.P.M. (2019) 'Editorial Overview: Inclusive Business for Sustainability', *Current Opinion in Environmental Sustainability* 41: A1–A4

Reardon, T. (2015) 'The Hidden Middle: The Quiet Revolution in the Midstream of Agrifood Value Chains in Developing Countries', *Oxford Review of Economic Policy* 31.1: 45–63

Rogers, P.J. and Peersman, G. (2014) '**Developing a Research Agenda for Impact Evaluation in Development**', *IDS Bulletin* 45.6: 85–99, DOI: **10.1111/1759-5436.12115** (accessed 16 November 2021)

Ros-Tonen, M.A.F. *et al.* (2019) 'Conceptualizing Inclusiveness of Smallholder Value Chain Integration', *Current Opinion in Environmental Sustainability* 41: 10–17

Schoneveld, G.C. (2020) 'Sustainable Business Models for Inclusive Growth: Towards a Conceptual Foundation of Inclusive Business', *Journal of Cleaner Production* 277: 124062

Schouten, G. and Vellema, S. (2019) 'Partnering for Inclusive Business in Food Provisioning', *Current Opinion in Environmental Sustainability* 41: 38–42

Stern, E. *et al.* (2012) *Broadening the Range of Designs and Methods for Impact Evaluation*, London: Department for International Development

Steward Redqueen (2021) ***Joint Impact Model – Version 1.6*** (accessed 11 November 2021)

Stoian, D.; Donovan, J.; Fisk, J. and Muldoon, M. (2012) 'Value Chain Development for Rural Poverty Reduction: A Reality Check and a Warning', *Enterprise Development and Microfinance* 23.1: 54–69

Ton, G.; Vellema, S. and Ge, L. (2014) '**The Triviality of Measuring Ultimate Outcomes: Acknowledging the Span of Direct Influence**', *IDS Bulletin* 45.6: 37–48

Vaessen, J.; Lemire, S. and Befani, B. (2020) *Evaluation of International Development Interventions: An Overview of Approaches and Methods,* Washington DC: World Bank

Westhorp, G. (2013) 'Developing Complexity-Consistent Theory in a Realist Investigation', *Evaluation* 19.4: 364–82

Westhorp, G. (2012) 'Using Complexity-Consistent Theory for Evaluating Complex Systems', *Evaluation* 18.4: 405–20

Yohannes, L. (2020) *Adaptive Management: From the Inside Looking Out*, Addis Ababa: Enterprise Partners

This page is intentionally left blank

Systems, Sapiens, and Systemic Change in Markets: The Adopt-Adapt-Expand-Respond Framework*

Ben Taylor[1] and Jake Lomax[2]

Abstract Systemic change is universally desirable and poorly defined. This article seeks to refine a practitioner-developed framework – Adopt-Adapt-Expand-Respond (AAER) – for conceptualising systemic change, and offers case studies to demonstrate its utility in planning for and measuring such change. To do so, the article firstly seeks to define the nature of a system and the components of change within that system. It also discusses the relevance of behaviour change among both actors and institutions in conceptualising systemic change. Finally, in exploring the utility of AAER throughout the implementation of development interventions, it examines the role of the framework in adaptive management: utilising data on observed changes to alter programme intervention tactics.

Keywords systemic change, market systems development, sustainability, adaptive management, monitoring, M4P, feedback loops, institutional economics.

1 Introduction

Considering something as 'systemic' sometimes seems to be shorthand for politicians, academics, or practitioners to refer to something 'important', without necessarily understanding what it is or how it came about. Talk of systemic change abounds in the area of market system development, but clarity in understanding is needed. The opposite of systemic change seems to be consistently considered as undesirable – change that is temporary, superficial, or tokenistic. It seems logical, then, that systemic change should be something we seek – and to do so, we need to be able to plan for it and measure whether or not it has happened.

This article further develops the Adopt-Adapt-Expand-Respond (AAER) framework, introduced by Nippard, Hitchins and Elliott (2014)

The *IDS Bulletin* is published by Institute of Development Studies, Library Road, Brighton BN1 9RE, UK
This article is part of *IDS Bulletin* Vol. 53 No. 1 February 2022 **'Theory-Based Evaluation of Inclusive Business Programmes'**; the Introduction is also recommended reading.

in the world of implementers of market systems development (MSD) programmes. It highlights the utility of the framework to a research and evaluation audience in planning for and measuring systemic change. To do so, the article utilises and clarifies some key concepts for understanding the market system as an interconnected set of transactions within an institutional environment, where interventions look to address market failure. It highlights how behaviour change is a key outcome of MSD interventions, both in changing the role of direct value chain actors and in supporting functions and institutions. Ultimately, this article wants to present AAER as a tool to plan for and measure systemic change in markets.

The article builds on practitioner-led literature on Making Markets Work for the Poor (M4P) (stemming from Springfield Centre 2008), which in turn builds on academic work in new institutional economics (North 1990; Williamson 2000). In establishing the desire for systemic change in development, a general awareness of the critiques of aid (Easterly 2002; Moyo 2010) is useful. To comprehend the difficulties of delivering on and measuring systemic change, an understanding of complexity in development is also useful (Ramalingham and Jones with Reba and Young 2008; Taylor 2014).

The AAER framework itself emerged from the practice of the Katalyst programme in Bangladesh in 2014 (Nippard *et al*. 2014) but the confusion about its concepts has led to continued attempts to clarify its usage and prevent inconsistent application in programmes (Taylor 2016; Lomax 2020). These various attempts may even have increased the confusion for practitioners, with their emphasis on unmeasurable notions of change using the lens of complex adaptive systems (Cunningham and Jenal 2016). Others have conflated aid-led private sector partnerships or interventions with the more complex innovation and system-level changes in markets (FSD Kenya 2016). This article represents an attempt to capture learning to date, refine the framework, and showcase its use by employing stylised case studies with two intervention arms that illustrate the breadth of MSD interventions. Further, the article discusses the framework's dual purpose in both planning for and measuring systemic change that results from MSD programmes.

The article is structured as follows: Section 2 presents the boundaries of the market system as commonly used in MSD programmes, and clarifies the system components and interactions that MSD interventions seek to change and monitor. Section 3 then serves to define the AAER framework and its components. Section 4 uses illustrative examples to demonstrate how AAER can be used to develop a vision of where activities need to focus, and how it helps to identify indicators that can capture (early signs of) systemic change. Section 5 then concludes with an agenda for further application of the framework.

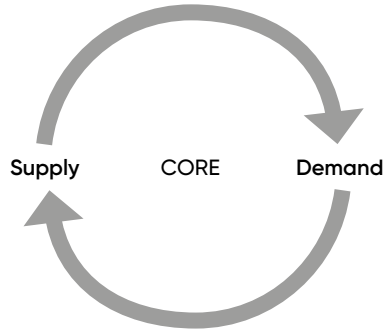
2 Market systems

MSD is a practitioner-derived approach to development interventions based on new institutional and transaction cost economics. It aims to provide an analytical framework and some guidance on interventions geared towards large-scale, sustainable social change. Over the years, discourse within new institutional economics has engaged in debate about how different levels and scales can be combined to analyse the institutional arrangements in markets, including the categorisation of organisations and institutions within this ambiguity. Add to this the considerations of varying economic rationality among individuals and you get a confusing picture of individuals, organisations, and institutions, all with relative autonomy and agency but all constrained by social structures (Agora Global 2019).

These structures and institutions provide the mental orientation or resources for individuals to act, but at the same time these individuals may act to change, tweak, or evade it (Coase 1992; Cheung 1983; Ménard 2004; North 1990). For example, a law might traditionally be considered as a 'solid' institution from the point of view of an individual, but unpacked, that institution would reveal several organisations comprised of several individuals, each with a role to play in developing, changing, or enforcing that rule (Ramström 2018). Further, many firms look for ways to avoid the costs and capitalise on the benefits of that law, not only when paying taxes but in other forms of patronage which capitalise on related individual, organisational, or social characteristics.

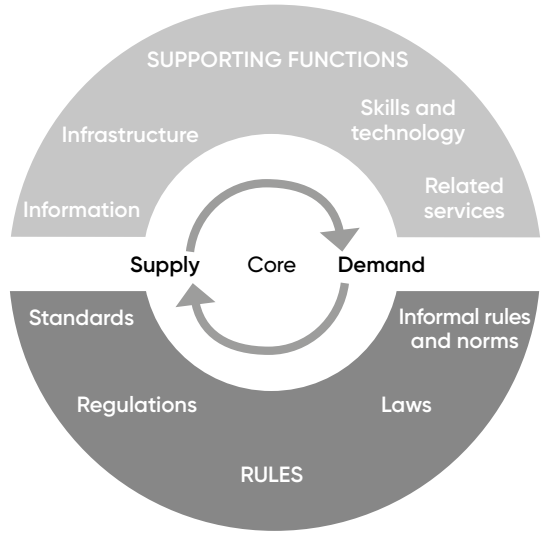
Markets and institutions do not change their behaviour; people change their behaviour. Without behaviour change among at least some of the actors who make up the institutions in the market system, no systemic change takes place. Here, most analytical frameworks of MSD have been lacking – simplistic and focused only on those actors who have a direct engagement with an intervention and ignoring (or taking for granted) the actors who have an indirect influence on the transaction in the market.

Figure 1 Core transactions



Source Springfield Centre (2008), reproduced with permission.

Figure 2 **Market system diagram**



Source Springfield Centre (2008), reproduced with permission.

MSD has a relatively straightforward way in which systems are conceptualised and depicted. MSD, as an approach that is designed to facilitate intervention for social change, applies specific boundaries to the system that is being analysed. In this market system, the 'actors' at the core (Figure 1) of the system diagram (Figure 2) are the target group – those whom the development actor wishes to benefit from an intervention. Every other component of the institutional web of the market is incorporated only when it has a function in the delivery of the core transaction.

2.1 Core transactions

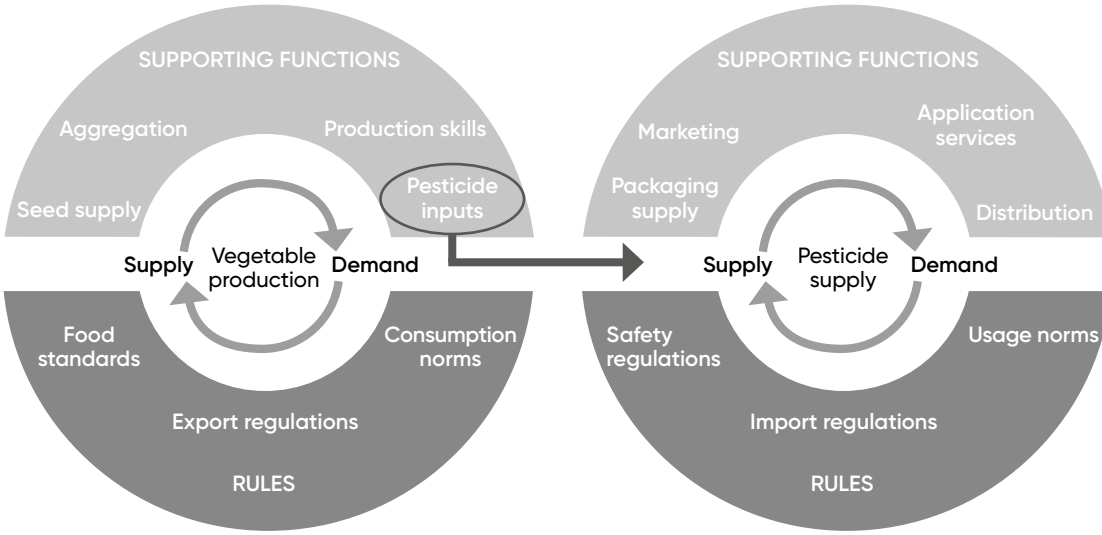
Market systems are composed of transactions. One party supplies – goods, services, labour, employment, or rights – and another party receives. This does not necessarily imply direct communication between both parties. Linking supply with demand is the core function of markets.

Who plays each of these roles varies according to the nature of the transaction. However, in market system development, the target group – the poor or disadvantaged – should always play the role of either supply or demand in this central transaction.

2.2 Supporting functions and rules

This core transaction is enabled or inhibited by a range of functions and formal and informal rules. These can be divided into the supporting functions and rules that affect supply, the supporting functions and rules that affect demand, and the supporting functions and rules that affect exchange (Figure 2).

Figure 3 Principal and supporting markets



Source Adapted from Springfield Centre (2008).

Take, for example, agricultural production. The supporting functions and rules affecting supply include input supply, labour availability, and agricultural information. For demand, relevant supporting functions and rules might include marketing and informal norms around consumption behaviours; while for exchange, important supporting functions and rules might include infrastructure provision, market information, and export regulations.

2.3 Nested and interlinked systems

Every one of the supporting functions or rules – whether related to supply, demand, or exchange – can be placed at the core of a nested market system diagram and each forms part of its own supply, demand, and exchange transaction. This is, in turn, enabled or inhibited by its own supporting functions and rules. As such, it can be conceptualised as a network of market system 'doughnuts', or even as a 'galaxy' of interlinked and nested system objects (Figure 3).

In the example here, where the core transaction relates to agricultural producers selling to consumers, the conditions of the input supply will influence the price, quality, and quantity of the agricultural produce in the core transaction. Examples of supporting functions of the input supply support system might include, on the supply side, finance or skills required to develop new products, or on the demand side, the distribution to get inputs to rural areas or the marketing required to increase the quantity of demand. Rules might include licences that increase

the price of exchange or norms around use of chemicals which affect demand.

This represents a stylised version of what an interlinked system looks like. The diagram imposes a notion of order on the complexity and messiness of the markets in the real world. In order to stimulate a change in a system, or at least observe how change is happening, one must understand the core transaction and how the outcomes of it are influenced by the supporting functions and rules which surround it. Understanding these functions and rules, the interactions between them, and how they affect outcomes in a core transaction allows for a vision of how they might work differently to improve these outcomes.

This conceptual framework for understanding market systems means that the boundaries of the system of interlinked market systems are, potentially, indefinite. Even in a simple system, one might conceive of a market system of a processor, a producer, and an input supplier. For example, there might be a relevant market system of research that might produce the basic seeds for propagators that delivers to a market of input suppliers. The boundaries of the system, therefore, are a key part of market system analyses; they are often pragmatically defined (and redrawn) based on expert judgement and influenced by factors such as the exact definition of the development challenge that is addressed, the potential to influence certain actors or components, and the likelihood and speed or scale of the envisaged impact.

Importantly, the conceptual framework is flexible, pragmatic, and intuitive, and is designed to help understand a market in order to reach an outcome for a target group. It is **not** designed to reflect how the system **should** work – governance and institutional change programmes are most often normative about what the institution should look like (e.g. the nature of a rule, which actor should perform which function, which functions **need** to be performed). In MSD, the object of study (and eventually intervention) is the observed realities of the market as it works for the target group. Instead of being normative about market structure or the type of actor who **should** perform certain functions, the main normative aspects of MSD are the definition of the target group and the outcome of interest that a programme might seek to achieve. A desirable outcome is placed at its core and the analysis is about enablers and barriers to its realisation.

3 Capturing the systemic change with the AAER framework

As documented above, intervening to affect systemic change is about altering functions and rules, or 'structures'. It is not aimed simply at the technological uptake of a new product or service if that does not alter the way the system operates for the benefit of the target group. It can be referred to as an innovation in the way the system operates, with an innovation being defined as:

> A change in the way that one or more supporting functions and rules of a principal or support system operate(s) that confers a benefit to the target group in the principal system. This will consist of one or more actors changing their behaviour in one or more ways.

Based on the goals of sustainability (Mosley and Taylor 2014) and scale of impact at this system level, the changes in the performance of supporting functions and rules identified above must demonstrate:

- uptake, ownership, and investment by relevant actors within the system, in the absence of external involvement;
- increased impact over time, creating more benefits for more people in the target group;
- changes in other supporting functions and rules to stabilise or augment the impact of the innovation (Taylor 2016).

Market systems development needs to be evidence-based rather than normative in the choice of support interventions. This is in contrast to many approaches to private sector development that utilise predefined instruments such as matching grants, credit guarantees, or capacity building. This means that there is significant discretion for implementers as to which instruments to deploy. They need to continually monitor and adapt their intervention depending on the degree to which it is achieving progress towards systemic change. However, in order not to lose sight of the various intended pathways of change, it requires a way of systematically monitoring what they hope to achieve.

A useful heuristic for achieving these objectives is the Adopt-Adapt-Expand-Respond (AAER) framework, also called the **Systemic Change Framework**. As defined in the Merriam-Webster dictionary, a heuristic is an 'aid to learning, discovery, or problem-solving by experimental and especially trial-and-error methods… [they] utilise self-educating techniques… to improve performance' (Merriam-Webster 2021). While uptake has been broad in the professional field by implementers of MSD programmes (Jenal and Gray 2019; DCED 2017; Samarth-NMDP 2015; Kosoris 2018), like many heuristics that have evolved from practice, the framework's theoretical foundations are unclear and its use pluralistic. The four key components of the AAER framework are explained below.

3.1 Adopt

In the first instance, the role of an agent external to a system, such as an MSD programme, is to identify what change is needed – which of the supporting functions and rules within a system are underperforming, how might they perform better, and what actions should be taken to bring that change about. This assumes that the system is not generating this solution of its own accord and so programme interventions to instigate an innovation are necessary.

Adopt is a process where an innovation in the operation of one or more supporting functions or rules of the market system is introduced, and ownership over it is gradually institutionalised or adopted by the relevant actors in the system. This will involve action by different actors that perform different roles in the systems. In Adopt, for example, a programme could be testing a technology or refining a product or service in partnership with one or more firms whose incentives are expected to be similarly aligned should the innovation be successful. It may be the case that multiple models of innovation fail at Adopt – constraints may be too strong and intractable, or the barriers to seize the opportunities are too high (e.g. costs), to warrant further programme investment.

There are two main criteria against which innovations generally fail in Adopt. Firstly, even with programme support, various actors who are needed for the innovation may not see the benefit of the change in their practice and may stop that behaviour. Potential reasons for the failure here are numerous. Incentives may not be sufficient to sustain the behaviour change, personal circumstances or the wider social or economic environment may shift, or it may be as simple as personalities being unconducive to continuing the new relationships.

The second criterion is whether the model actually leads to the intended outcomes. Development programmes have a pro-poor objective in mind. Because the programme aims to improve the scope for the target group to realise improved outcomes, when an innovation fails to impact upon them, it cannot be considered as being a systemic change within that system.

In the example above, a programme might want to change the way in which farmers access inputs, changing the performance of the input supply function. In the support market, the innovation may be a new way for input supply companies to use rural distribution networks. Considering both the core market and the linked-support market systems, it will be clear that the adoption of any innovation requires several behaviour changes by various actors. The input suppliers may need to seek access to information on new inputs, source a supplier, decide to spend the necessary money to acquire the inputs, instruct staff to promote them and so on.

Meanwhile, farmers need either to seek or otherwise receive information on the availability and use of these inputs, decide to spend and then proceed to spend the money on them, invest in ground preparation, allocate time to tend to the crops and so on.

In most cases, there will also be several further changes needed in the practices of other actors for the innovation to have an impact. The department of trade might need to proactively permit the import of new products or introduce quality control

on the imported seeds or veterinary products that farmers cannot easily observe themselves, in order to close the door to harmful behaviour of opportunistic input sellers. And, when the goal is about income change rather than simply yield, there are several actors on the demand side whose behaviour will need to change to ensure that they buy the right quality of produce at the right prices to sustain the farmer's new practice and to ensure that the theory of change holds. Mapping – and continually adapting this mapping – of these market actors and the behaviour changes necessary helps programmes to determine where an intervention is and is not working and why, as part of adaptive management.

3.2 **Adapt**

As a component of the systemic change framework, **Adapt** refers to sustained behaviour change by relevant actors. The actors involved in the innovation – both those who were supported by the programme and those who weren't – must have adopted new behaviour for the model to work. They need to incorporate this into their 'normal' operations, without the need of programme resources, with their own investment of time, money, and other resources. Evidence of institutionalisation is often seen not so much in the continuation of the initial adopting practices (particularly if that behaviour was the result of external stimulus), but is evidenced by continued refinement and alteration of these practices to the dynamic realities of each actor (Glover *et al.* 2019).

In response to different conditions, this adaptation process can vary a lot between different actors, such as groups of farmers, and can result in different pathways that might even create the need to identify and map a different subsystem, such as organic producers for export versus smallholder production to local markets. Moreover, the process of institutionalising the changed behaviour – moving from Adopt to Adapt – will happen at the system level only when there is an appropriate configuration of supporting functions. Many of the supporting actors involved in the innovation will start to experience the results of the innovation and adjust their behaviour for their own purposes. For example, after a successful adoption of a seedling provisioning programme with programme support, the same provisioning system might, with other crops, become reliant on bank loans instead of project funding. In the Adapt phase, the initial arrangements between the market actors are fine-tuned through experiential learning and in response to changing market dynamics.

3.3 **Expand**

Expand is about pushing the boundaries of the innovation – scaling the process in order to have more benefits for more people. Expand is not the repetition of a support intervention or partnership but an expansion of the change process in the market system, including both scaling and deepening, with various mechanisms that underpin these processes (Table 1).

Table 1 Components and mechanisms of Expand

Change		Mechanisms
Scaling	**Deepening**	
● New geographies ● New segments of target group ● Income groups ● Inclusion of marginalised segments – women, minorities, etc.	● Lower costs ● Higher incomes from products ● Greater health or wellbeing benefits ● Better resilience and protection of future incomes through disease resistance or genetic diversity	● Imitation of practices of the original beneficiaries ● Strengthening of existing actors ● Roll-out at scale ● Involvement of new actors ● Expansion to new geographies ● Increased competition ● Lower transaction costs ● Further innovation

Source Authors' own.

Imitation of functions or products and increased competition between firms to serve the changed market are often early signs of sustainability. If others are not imitating or emulating innovations, this can be indicative of more fundamental problems in the way the system operates, including information transmission blockages or high entrance costs. As a result, a programme might want to re-engage in order to include new actors or new areas.

Once an innovation has proven its worth, the risk for other 'imitating' actors will be lower. It may also be that the programme initially targeted easier-to-reach areas and so heavier programme involvement is required in order to push impact into more marginal areas. Different partners also have different needs determined by their capacities, and so the type of programme support might also differ from that in the initial innovation.

The obvious group to look at for a change in behaviour that enables the innovation process to expand and scale are the actors who play similar roles in the system – often competing companies or, for changes in rules or sector policies, different government departments. If these new actors change their behaviour, adopting relevant aspects of the innovation – such as ways of packaging, seedling distribution models, certification systems – the systemic change triggered by the pilot innovation will expand and affect more people. Using the previous example on distribution and marketing of agricultural inputs, the new model may have been successful in that the marketing firms, input suppliers, agricultural extension agents, and farmers may all have changed their behaviour in the required way.

However, the impact of this change in model may still be limited, for example, to only certain geographical areas or to certain segments of the population such as male farmers. In such cases, it may be necessary to partner with other actors. For example, women's cooperatives, which have a closer engagement with segments of the target group, could be supported to provide the supporting function instead of a more distant private company. These modifications and extensions of the model may result in an expansion of the impact of the innovation – more benefits to more people.

3.4 Respond

Respond, as a component of the systemic change matrix, looks at whether supporting functions and rules other than those explicitly targeted within the innovation's core objective are changing in response to the innovation. It assesses the other changes that are happening in supporting functions or institutions, and the degree to which they are supportive of or obstructive to the desired impact. In order for an innovation to reach sustainability and scale of impact, it is essential to monitor the role of those not directly involved in the original innovation and outside the boundary of the targeted core market system – actors whose role is in delivering support functions and rules in support markets. These behaviours are often difficult to anticipate at the start of an innovation process. For example, it may be that service providers involved in the innovation start marketing their services in totally different markets, or that government institutions use the new organisational model of stakeholder consultation in other sectors.

Where the components Adopt, Adapt, and Expand target changes in the operation of one or more initial supporting functions or rules which are part of a programme's vision for how a sector might work better to improve outcomes for the target group, **Respond** addresses the changes in other supporting functions or rules that may reinforce and enhance the changes from the initial innovation.

4 Example: AAER as a planning and monitoring tool

The AAER framework is applied following a market systems diagnostic process, mapping supporting functions, rules, and transaction modalities in a system. This is necessary to understand the causes of underperformance within a system and to find an opportunity to address a constraint that may yield a better outcome for the target group. However, it is always the actors' behaviour in performing these functions and implementing these rules which are generally the entry point and levers for systemic change and, therefore, the main units of analysis for intervention development in MSD.

As a tool, AAER helps to plan for and reflect on changes in the system. When considering all four components of the AAER framework, it is obvious that multiple actors will have to change

several of their behaviours simultaneously for the objective of systemic change to be reached. The four components are not different phases but focus on processes that may happen simultaneously. As discussed throughout, the AAER framework helps to identify several points at which changes in supporting functions or rules are needed, and it helps for learning and adaptive management of the MSD programme. Building on the work of Lomax (2020), we use feedback loops to show how AAER informs further programme decisions about where to intervene and when to withdraw (see Figure 4). We illustrate these feedback loops with experience from several MSD programmes in horticulture.

4.1 AAER in planning for change

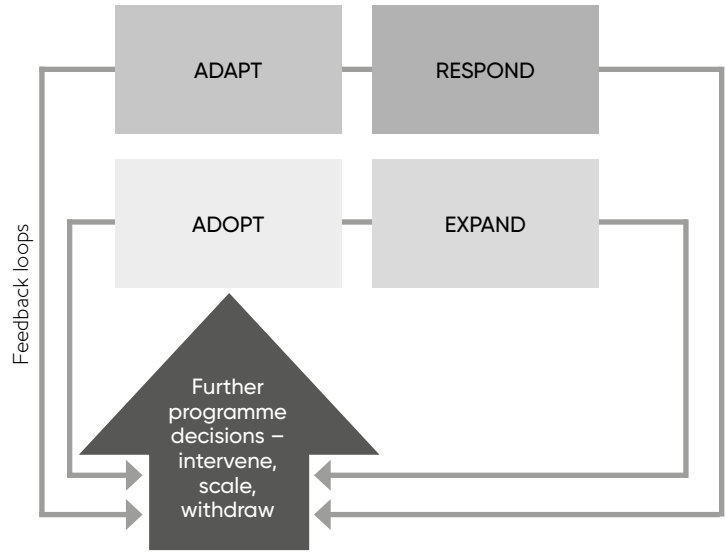
Many MSD programmes in the vegetable sector seek to increase the incomes of smallholder vegetable producers as their central aim. We combine real-world insights from horticulture MSD programmes in Bangladesh, Ethiopia, and Fiji in this somewhat stylised example. One of the support markets that was seen as a cause of this underperformance was seed supply. Within this support market system, several constraints were framed as root causes.

- Farmers were not aware of the potential commercial returns to planting improved seeds. Some farmers simply were not aware these seeds existed as they were not available through their usual supply channels.
- Some farmers who had attempted planting improved seed had not employed good agricultural practices and so yield was low.
- Some farmers had experienced adulteration or counterfeit seed and no longer trusted the products sufficiently to invest.
- Further, many farmers could not afford to invest in these new technologies – or rather viewed the risk as too high – in the way they were currently marketed.

The MSD programme identified the need for a pilot to target a change in the way the **marketing** and **distribution** functions around seed provisioning worked. The vision at this stage was relatively simple; seed companies would trial a range of marketing techniques that would build (and rebuild) consumer confidence in their products, in tandem with a strategy to get these products closer to potential consumers with innovative distribution techniques.

At this stage, in terms of **Adapt**, **Expand**, and **Respond**, the rationale was also straightforward. Market analysis had demonstrated that sufficient market potential existed, and the private companies who were piloting the new marketing techniques had sufficient incentive to embrace the model.

Figure 4 Adopt-Adapt-Expand-Respond with feedback loops



Source Authors' own, adapted from Lomax (2020).

The clear yield and market potential for new varieties would give farmers sufficient incentives to institutionalise the necessary behaviour change. Considering the market potential, there were also very clear pathways to the refinement of the products themselves, such as offering other seeds suited to the local conditions in these markets.

If the innovation and pilot experimentation were successful, the mechanisms for expansion were clear. The mechanisms of expansion were located both within partner firms and through more competitive seed markets that would attract new firms and farmers. This competition would also likely drive down prices and, in doing so, the benefits to each farmer would increase over time. Finally, the programme anticipated potential for positive change in other supporting functions and rules as a result of the innovation. If the distribution system began to work, then these could become 'hubs' or channels for agricultural extension and the provisioning of other services to farmers.

The programme then had to refine, with and through partners, exactly what these new marketing and distribution strategies would be. In MSD, the critical part of the analysis here is around the capabilities and incentives of those actors in the system who could play an ongoing role in the delivery of the intervention.

Adopt: Initially, the ideas for behaviour change by the seedling producers were straightforward. In seed marketing, it was envisaged that demonstration plots would overcome the trust barriers to uptake, while in distribution, the programme saw the

potential of a low-cost mobile seed vendor model incorporating bicycle distribution which was more trusted and formalised than other distribution systems.

Adapt: At this stage, the key to institutionalising change was ensuring that the model became aligned with the incentives of relevant actors. From the outset, therefore, the programme partnered simultaneously with different types of organisation. Firstly, the programme reached an agreement with a market leader who could achieve scale quickly when the model proved successful, providing this firm with market research and information. Simultaneously, the programme partnered with a start-up, providing it with an innovation grant and technical assistance.

Feedback loop: having identified the market actors who would need to change their behaviour in order to institutionalise change, the interventions recognised that their incentives were not sufficient to incorporate the activities as part of their normal business. The reason was that the government needed to allow this new model of seed distribution and it became clear that a combination of power dynamics meant that this did not happen. Therefore, the interventions needed to be refined by revisiting the Adopt strategy, including support interventions to trigger behaviour change in the local government and the agricultural department in the geographies where the new model was to be trialled.

Expand: There was an inherent consideration of scaling in the design of the interventions. These were not demonstration plots run by an NGO with replication dependent on this NGO's continued existence; instead, these were demonstration plots within the seed companies' established practices so that they would take the model to scale within normal market conditions. The demonstration plots were not farmer-led, considering that attending a demonstration has been shown to be as effective as running the demonstration plot itself for adoption of practices (Duflo, Kremer and Robinson 2007; Khan *et al*. 2009). Moreover, the fact that the demonstration plots were managed internally by seed producers also assured that the quality of the seeds could be controlled more easily.

Feedback loop: while achieving impact at scale had been considered in the intervention design, at this planning stage, it became clear that certain key assumptions were unlikely to hold. The demonstration plots needed to be complemented with better distribution systems; no scale can be achieved if farmers cannot buy the seeds that are being demonstrated. Market analysis revealed that informal mobile seed vendors (MSVs) bridged this gap for seeds of other crops. However, these seed vendors lacked quality control and had limited agronomic knowledge. The programme tried to overcome these challenges by revisiting the intervention design, formalising the model of MSVs and incorporating them into the business model of seed companies.

Respond: In developing this innovation, the programme acknowledged that there was potential for other actors in the wider market systems to respond, which could have positive (and negative) consequences for the programme objectives.

Feedback loop: at the actor level, then, the programme began to plan for, observe, and respond to these broader market system changes with deliberate tactics. It monitored, for example, whether government extension policies could respond to the new modalities. As such, regular meetings were held from the outset with the relevant ministry to assess their capacities and incentives for change. Exogenous shifts in some support markets also had an impact on the intervention tactics. For example, new products in crop protection and crop nutrition became available and were incorporated in the adoption pilot – this greatly influenced the success of this seed system intervention.

4.2 AAER in monitoring change

The programme used a three-step process for monitoring actor-level change in relation to broader system-level change and adaptive programming:

1 Define the behaviour change and from whom you expect to see that change.

2 Establish whether that behaviour change has taken place, and to what extent.

3 Decide on course corrections/additional intervention tactics depending on the outcome.

These steps are represented for the seed case in Tables 2–5, illustrating how the behaviour changes are monitored and measured. It should be noted that more behaviour changes would be covered here when fully implementing the framework, as this example is for illustration purposes only. The examples are drawn from routine programme monitoring data unless otherwise indicated.

An important side note is warranted here on the methodologies that might be included in these programme-monitoring systems in order to operationalise this framework for adaptive management. Monitoring of adoption is well established, utilising conventional methodologies – field observations of the number of people who attended training or the sales of partner firms, for example. Given the intention of MSD interventions, programmes need to ensure that this evaluation of impact goes as far as capturing the intended development outcomes – that is, not only measuring sales of seeds, but tracking this through to the sales of the produce of those seeds and incomes for the farmers. This need not all be primary data, and aspects can be based on explicit assumptions. In addition, having mapped the relevant actors and behaviours, it is important to include data capture of

Table 2 **Adopt**

Behaviour change process	Evidence	Course correction
Partner seed companies establish demo plots in the target location.	All partner seed companies successfully established demo plots.	
Partner seed companies distribute seeds through MSVs.	All partner seed companies piloted MSV distribution models.	
Farmers attend demo plots.	Variation was perceived between attendance at the different sites.	Share learning from successful sites to modify other pilots in awareness raising.
Farmers buy the improved seed.	Purchase was closely linked to attendance. Repeat purchases were dependent on the proximity of sale i.e. people needed both marketing and distribution exposure to benefit.	Shift focus from knowledge to access.
Farmers plant improved seed.	Attending the demo plot was more successful than purchasing the seed from MSV alone, in terms of yields.	Analyse constraints to MSV sales (customer research).
Farmers sell vegetables and increase income.	The market is robust. Farmers demonstrate increased sales and income.	

Source Authors' own.

perceptions and behaviours beyond programme partners. For these observations, it may be necessary to incorporate formalised recording of qualitative observations from the field or key informant interviews with relevant actors.

For assessing adaptation, it is important to include an element of *ex post* data collection in a monitoring system. Methods will likely be similar to those deployed in the measurement of adoption but will assess the degree to which relevant actors have institutionalised a behaviour. This means that data collection must occur substantively after the end of programme support to partners – 12 months later, for example – and also that it must attempt to observe any changes in the nature of an innovation, such as being applied to different elements of the organisation.

Measurement of expansion needs to include observations within partner organisations (have sales continued to increase?) as well as measurement of the broader market, such as competing companies. For the latter, it is useful to maintain ongoing relationships with selected experts – for example, in business membership organisations, relevant government departments, or research organisations – who keep track of behaviour changes across the sector. Similar insights may be available through interviews with suppliers. Interviews with competing firms are often

Table 3 **Adapt**

Behaviour change process	Evidence	Course correction
Partner seed companies invest in new distribution and MSV model.	Three of the five partner companies continued the innovation, with the other two dropping out due to the financial and time investment required.	When examining scale-up and the introduction of new partnerships, the programme sought modifications to the model, which required less intensive up-front engagement by seed companies.
Farmers continue to **buy** and **plant** seed and **sell** produce.	Sales for all partner seed companies had increased, showing that farmers were buying. Farmer surveys showed increased incomes from sales by the vast majority of adopters.	
Seed companies continue to **adapt** model including price changes, product variation, and further development of marketing strategy.	Further investment is evidenced by the inclusion of additional complementary marketing techniques, including signboards and flyers.	
Some **farmers** refuse to buy seed owing to a lack of trust and consistency in seed supply.	Even farmers receiving promotional material and with access to seed purchased lower than expected amounts of seed, and surveys showed that the distrust related to the intrinsic quality of the seeds.	The programme began to examine the potential for national and independent seed certification processes. This resulted in supplementary intervention providing technical assistance to the Ministry of Agriculture [Respond].

Source Authors' own.

useful although access can be a problem when the programme does not have a relationship with these firms. Findings may be triangulated with demand-side studies to assess **who** is receiving the intended benefit to assess whether any progress is being made towards the deepening of impact.

Finally, measurement of market response will likely involve a less formalised process of monitoring changes in a market. In some cases, this can come in the form of repeated market system assessments to understand whether changes in supporting functions and rules are happening. (These principally take the form of secondary data analysis and key informant interviews with actors involved on the supply and demand sides as well as those involved in the performance of supporting functions and rules.) In other cases, it might involve more targeted data collection around specific functions where the programme implementers suspect a change might be happening, such as a survey of business development service providers to assess whether they have started to offer services to seed companies to improve their marketing strategies.

Table 4 **Expand**

Behaviour change process	Evidence	Course correction
Non-partner seed companies adopt new marketing tactics and formal MSV model.	Market surveys revealed that MSVs have become the norm across seed companies. Uptake of new marketing techniques has been lower with only three non-partner seed companies adopting the model.	Low uptake of marketing techniques indicates issues with the nature of the tactics used. While MSVs alone likely mean improved access to seeds, pilots showed that uptake, especially among low-income farmers, remained low. As such, this feedback allowed the programmes to design new interventions attempting alternative marketing strategies.
Farmers who are not customers of partners buy and plant seeds, and sell produce.	Performance naturally varied but on average farmers buying the improved seeds from MSVs and having attended demo plots continued to report higher incomes than before.	
Partner seed companies expand offering to new geographies and products.	The geographical spread was not as anticipated as seed companies continued to focus on other more remunerative markets.	Supplementary programme interventions were necessary in order to de-risk investment into areas with a smaller and more unproven market.
Seed companies reduce prices of seeds.	While there was an initial and slight increase in prices after the pilot period, when more farmers wanted to procure seeds, later, in areas where several companies operated the same model, prices did fall.	Ensure mechanisms for competition are in place. Supporting additional pilots with new partners even where the model existed was considered valid, although not a priority.
Seed companies offer increased variety of products available to farmers.	While the pilot started with a handful of high-value vegetables, across seed companies, after some years, more than 100 varieties were on offer through formalised MSVs.	

Source Authors' own.

The use of AAER as a planning tool should help to construct a measurement system for its use as a measurement tool. Using the seed example, Tables 2–5 demonstrate this link by looking at the expected behaviour changes from different actors underpinning an innovation, the evidence the programme looked for in determining whether that change had taken place, and the course correction as a result of what they found.

5 Conclusion

There are few who would argue that sustainability and scale of impact are desirable outcomes of development intervention and, in recent years, these qualities have become intertwined in the discourse around 'systemic change' (Taylor and Donovan 2016). That discourse is, however, messy. While everyone seems to

Table 5 **Respond**

Behaviour change and group	Evidence	Course correction
Policymakers seek information on results of new distribution mechanisms.	Little engagement was seen from policymakers, who continued to favour distribution through public channels or distributors with links to them.	The programme diversified the government engagement strategy, working both with different departments but also different individuals with a view to achieving buy-in.
Seed companies invest in research and development on this specific market segment or geography.	The larger seed companies had a greater capacity to invest in R&D and so began to invest in product development. They found a way to reduce the cost by reducing packaging size and embedding alternative financing models.	The programme both supported these innovating firms to scale the innovation more quickly and also introduced new related innovations concerning the financing aspect of the innovation, partnering with financial institutions to underwrite new financing products.
Ministry of agriculture implements new seed certification law.	Law is passed by parliament.	
Ministry of agriculture establishes new seed certification body.	Staff are recruited and organisational structures are put in place.	
Farmers trust new seed certification system and **begin to purchase** additional certified seed.	Sales of seeds to farmers increase in similar areas and using similar marketing and distribution techniques as in the pilot. Farmers report greater trust in the seeds they buy.	Supplementary intervention developed at earlier stage in these new markets to improve trust in the seed system.

Source Authors' own.

agree that a nebulous concept of systemic change is important, there is an absence of a clear conceptual framework to help plan for it and measure whether it has happened. This article has attempted to contribute to this discussion by restating and clarifying a commonly used conceptual framework in market systems development, the AAER model. In this framework, **the market system** is conceptualised as a transaction embedded in supporting functions and rules, and as part of a network of nested and interlinking systems. The most important objective of the article is to show how the AAER framework can be used to dissect and design support components that increase the likelihood of achieving systemic change.

There are two roles of the AAER framework. Firstly, it is a grammar to articulate the programme's vision. If a programme aims to bring about systemic change, and the AAER framework helps to articulate what this could look like, then a programme should be better able to design support activities that could leverage systemic change. However, systems are dynamic and

complex, and initial plans for MSD support are rarely borne out in reality – and rightly so. Therefore, a second and perhaps even more important function of the AAER matrix is as a heuristic for monitoring, reflection, and adaptive management during the implementation of an MSD programme. The article presented a way that this was operationalised in a horticultural programme.

However, as is always the case with approaches and recipes, 'the proof is in the pudding'. And there is clearly a need for more empirical contributions, with experiences in other sectors, employing this framework and evaluating its usefulness for adaptive programming. These may yield other practical examples of how and when the different components of the framework can be operationalised according to the ambitions, the scale, and the length of programmes, and, of course, the budgets available for monitoring, evaluation, and learning. While this article has had to dedicate considerable attention to clarifying a number of interrelated concepts, further refinement is needed to allow learning from cross-case comparisons.

Notes

* This article draws extracts from unpublished work by the authors (Taylor 2016; Lomax 2020). The article also benefited significantly from inputs from Giel Ton, to whom the authors are incredibly grateful.

1 Ben Taylor, CEO, Agora Global, UK.

2 Jake Lomax, Director, 3sd Research, UK.

References

Agora Global (2019) ***Firmo Economicus: How Late Capitalism Loses the Motivations of the People that Comprise Capital***, Agora Global blog, 24 April (accessed 27 August 2021)

Cheung, S.N.S. (1983) 'The Contractual Nature of the Firm', *Journal of Law and Economics* 26.1: 1–21

Coase, R.H. (1992) 'The Institutional Structure of Production', *American Economic Review* 82.4: 713–19

Cunningham, S. and Jenal, M. (2016) '**Rethinking Systemic Change: Economic Evolution and Institution**', *Building Effective and Accessible Markets (BEAM) Exchange Policy Brief*, London: BEAM Exchange (accessed 27 August 2021)

DCED (2017) ***The DCED Standard for Measuring Results in Private Sector Development: Control Points and Compliance Criteria***, Donor Committee for Enterprise Development (accessed 8 September 2021)

Duflo, E.; Kremer, M. and Robinson, J. (2007) ***Understanding Technology Adoption: Fertilizer in Western Kenya***, Innovations for Poverty Action (accessed 8 September 2021)

Easterly, W. (2002) '**The Cartel of Good Intentions: The Problem of Bureaucracy in Foreign Aid**', *Journal of Economic Policy Reform* 5.4: 223–50, DOI: 10.1080/1384128032000096823 (accessed 8 September 2021)

FSD Kenya (2016) ***Formalising Informality: Savings Groups, Community Finance, and the Role of FSD Kenya,*** Nairobi: FSD Africa (accessed 8 September 2021)

Glover, D.; Sumberg, J.; Ton, G.; Andersson, J. and Badstue, L. (2019) '**Rethinking Technological Change in Smallholder Agriculture**', *Outlook on Agriculture* 48.3: 169–80, DOI: 10.1177/0030727019864978 (accessed 8 September 2021)

Jenal, M. and Gray, J. (2019) ***Measuring Systemic Change in Market Systems Development – A Stock Taking***, Tegucigalpa: United States Agency for International Development/Honduras (accessed 27 August 2021)

Khan, A.; Pervaiz, U.; Khan, N.M.; Ahmad, S. and Nigar, S. (2009) '**Effectiveness of Demonstration Plots as Extension Method Adopted by AKRSP for Agricultural Technology Dissemination in District Chitral**', *Sarhad Journal of Agriculture* 25.2: 313–19 (accessed 5 October 2021)

Kosoris, J. (2018) ***The Benefit of Hindsight: How Market Systems Theory Could Have Predicted Aflasafe Uptake in Nigeria***, BEAM Exchange blog, 17 December (accessed 8 September 2021)

Lomax, J. (2020) ***AAER Revisited: From Systemic Change Narrative to Systemic Change Analysis***, 3sd Research Briefing Paper 6, Falmouth: 3sd Research (accessed 8 September 2021)

Ménard, C. (2004) 'The Economics of Hybrid Organizations', *Journal of Institutional and Theoretical Economics* 160.3: 345–76

Merriam-Webster (2021) ***Heuristic*** (accessed 16 September 2021)

Mosley, P. and Taylor, B. (2014) 'Who Wants to Give Forever? Giving Meaning to Sustainability in Development', *Journal of International Development* 26.8: 1181–96

Moyo, D. (2010) *Dead Aid: Why Aid is Not Working and How There is Another Way for Africa*, London: Penguin

Nippard, D.; Hitchins, R. and Elliott, D. (2014) ***Adopt-Adapt-Expand-Respond: A Framework for Managing and Measuring Systemic Change Processes***, Springfield Centre Briefing Paper, Durham: Springfield Centre (accessed 9 September 2021)

North, D.C. (1990) *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press

Ramalingham, B. and Jones, H. with Reba, T. and Young, J. (2008) ***Exploring the Science of Complexity: Ideas and Implications for Development and Humanitarian Efforts***, ODI Working Paper 285, London: Overseas Development Institute (accessed 8 September 2021)

Ramström, G. (2018) 'The Analytical Micro-Macro Relationship in Social Science and Its Implications for the Individualism-Holism Debate', *Philosophy of the Social Sciences* 48.5: 474–500

Samarth-NMDP (2015) ***Results Measurement System: User Manual Version 2.0***, Lalitpur: Samarth-Nepal Market Development Programme

Springfield Centre (2008) *The Operational Guide for the Making Markets Work for the Poor (M4P) Approach*, report commissioned by Swiss Agency for Development and Cooperation and Department for International Development

Taylor, B. (2016) ***Systems and Systemic Change: Clarity in Concept***, Springfield Centre Working Paper, Durham: Springfield Centre (accessed 27 August 2021)

Taylor, B. (2014) 'In Vogue and at Odds: Systemic Change and New Public Management in Development', *Enterprise Development and Microfinance* 25.4: 271–87

Taylor, B. and Donovan, J. (2016) 'New Approaches to Old Problems: Systemic Change as a Unifying Objective', *Enterprise Development and Microfinance* 27.1: 1–4

Williamson, O.E. (2000) 'The New Institutional Economics: Taking Stock, Looking Ahead', *Journal of Economic Literature* 38.3: 595–613

Using Theory-Based Evaluation to Evaluate Systemic Change in a Market Systems Programme in Nepal

Edward Hedley[1] and Gordon Freer[2]

Abstract The complexities of markets and market environments are felt in the design and the evaluation of market systems development (MSD) programmes. The authors reflect on a recent evaluation of an MSD programme in Nepal in which they used contribution analysis as a means of navigating these complexities. The planned niceties of the proposal soon departed ways from the reality on the ground, forcing the authors to adopt a more iterative evaluation approach, while ensuring evaluative robustness. This article outlines the iterative process and what the authors have learned regarding the applicability of contribution analysis within a theory-based evaluation, in a dynamic, changing environment.

Keywords theory-based evaluation, market systems development, contribution analysis, dairy value chain, inclusive business, Qualitative Impact Protocol (QuIP), Nepal.

1 Introduction

Market systems development (MSD) programmes work with a variety of public and private sector actors to improve the way that markets function for the poor as consumers, producers, or employees. These programmes often work in complex environments where the level of complexity may grow more intricate given circumstantial factors such as a rich donor environment or thin markets or fragile socioeconomic contexts. Evaluating MSD programmes necessitates a methodology that is 'complexity-aware' and enables the evaluators to unpick the role of multiple overlapping drivers of change to uncover and understand the inner workings of programme processes (Chen and Rossi 1980, 1983, 1992) – unpacking the notorious 'black box' of evaluation (Stame 2004). In a recent evaluation of an MSD programme, we used contribution analysis as a means of categorising market forces and contributions within the

The *IDS Bulletin* is published by Institute of Development Studies, Library Road, Brighton BN1 9RE, UK
This article is part of *IDS Bulletin* Vol. 53 No. 1 February 2022 **'Theory-Based Evaluation of Inclusive Business Programmes'**; the Introduction is also recommended reading.

programme 'black box'. In this article, we explain the design and application of our process, as one possible tool to use within theory-based evaluation.

The programme we evaluated worked in a variety of agricultural sectors and we illustrate the approach we took by drawing on examples from our evaluation of the dairy sector. Initially, the evaluation was planned as a two-step process to determine the level of contribution of the programme. However, early in the process, we realised that we needed to adapt this approach. In the article, we explain the original design, our adaptation, and our rationale for taking these steps. We then reflect on the application, noting the insights gained, and using examples that illustrate how the methodology helped us to answer the question 'What interventions worked and why in generating systemic change?'

The article is structured as follows. Sections 2, 3, and 4 provide some context to the evaluation and focus on why we selected a theory-based evaluation design to meet our objectives. Section 5 highlights how we applied it – drawing on a worked example from the dairy sector. It discusses what we learned at each step in the process and how we adjusted our approach along the way. We conclude in Section 6 with our reflections on the process as a whole.

2 Background

In 2020, Itad Ltd[3] completed an evaluation of the Samarth-Nepal Market Development Programme (Samarth-NMDP) for the Foreign, Commonwealth & Development Office-Nepal (FCDO-Nepal) (Itad 2019). The Samarth-NMDP programme ran over a six-year period from April 2012 to March 2018 and was the first programme to apply the MSD[4] approach in Nepal. It worked across several agriculture sectors and tourism, with the aim of making these markets more inclusive of poor people. Itad was commissioned to undertake an endline impact evaluation of the Samarth-NMDP programme, with several major objectives. A significant objective of the evaluation was to determine the extent to which programme interventions had initiated change that might result in systemic change.[5]

3 The programme context and the evaluation challenge

By their nature, MSD programmes present evaluators with challenges. Their *raison d'être* is to include the excluded poor in fair, functioning markets; their ultimate objective is systemic change to overcome this challenge; and their *modus operandi* is to operate at arm's length, encouraging sustainable solutions while avoiding creating distortions within emerging market dynamics. In this way, MSD programmes seek to change the way markets work to benefit the poor by facilitating systemic change. Fundamentally, promoting systemic change derives from a wish to make the benefits of development intervention as inclusive and long-lasting as possible. Programmes achieve that aim by empowering market players to understand and overcome the challenges in market

rules and functions, rather than by offering more traditional direct delivery development assistance. However, the lack of direct involvement, the longer-term goals, and the complex, multifaceted market environment present the evaluator with evidence that is more circumstantial than direct or physical in nature.

The Samarth-NMDP programme was designed from the outset to be an MSD programme. In the first three years of implementation, however, in an experimental attempt to achieve scale more rapidly and to gain buy-in for the unfamiliar MSD approach from key government stakeholders, the programme partnered with established non-governmental organisations (NGOs) that were active in the programme's sectors of focus and unfamiliar with the MSD approach. Unintentionally, this resulted in a more direct delivery model being practised and limited traction for the MSD approach. After a 'midcourse correction', the programme opted for a more facilitative market systems approach. Market development programmes generally require time to catalyse systemic change; with the programme implementing a fully MSD approach for only two or three years, we anticipated that any observed changes in the market system would likely be embryonic. This made our evaluation that much more difficult. Systemic change is generally difficult to pinpoint, and now we were obliged to look instead for early indicators of this change.

Events external to the programme also created a challenging environment. The programme worked in a congested donor environment in which multiple government and other donor programmes occupied similar geographic locations. These were working in the same sectors, often with overlapping aims and objectives and sometimes in direct tension with the programme's market systems approach. The programme also worked against a backdrop of rapid socioeconomic change (such as rural-to-urban migration), a changing and fluid political landscape, and encountered a series of serious shocks, including the 2015 earthquake and the 2015 Indian Economic Blockade. Finally, while some of the interventions focused on sectors with significant market activity, in other areas, the programme operated in a classic 'thin market' context characterised by sparse and underperforming market-supporting functions – especially market failures in agricultural input markets and post-production services, poor physical infrastructure, and weaknesses in the policy and regulatory environment.

In short, given this context, we acknowledged at the outset that it would be difficult to collect evidence of systemic change and that we would have to seek indirect evidence that would give a plausible indication that systemic change would manifest itself in the future. Even where we found proof, the programme's contribution to change might not be obvious, given the multiple overlapping drivers of change, and the long and indirect pathways to impact.

In response to these challenges, we developed an overarching theory-based approach to the evaluation framed by contribution analysis. We selected contribution analysis precisely for its ability to deal with complexity, for its capacity to weigh up the relative importance of different factors, and its iterative and exploratory nature. Within this contribution analysis framework, we developed different evaluation modules based on a mix of different methods, including household surveys and key informant interviews.

4 What were we looking for?

The holy grail of MSD programmes is a systemic transformation that makes a market beneficial for poor producers and consumers. However, despite decades of implementation, a clear definition of systemic, transformative change remains elusive.

Given that Samarth-NMDP (hereafter known as 'Samarth') had already used the Adopt-Adapt-Expand-Respond (AAER) framework (see Taylor and Lomax, this *IDS Bulletin*) in its reporting on results achieved by the programme in the area of systems change, we opted to use this framework as a primary lens to identify evidence of systemic change. We purposively selected five programme sectors (dairy, vegetables, pigs, ginger, and tourism)[6] for study, based primarily on the programme's reporting of achievement against the AAER framework and the sectors' reach across the programme portfolio, considering their importance in terms of numbers in the target population (smallholder producers). This selection was made to balance the twin evaluation aims of accountability and learning. The dairy sector was selected, for example, because of its importance within the Samarth portfolio and its importance to the Nepali economy (contributing an estimated 8 per cent to Nepal's gross domestic product, with more than 3.5 million households engaged in the sector, of which 500,000 are producers and sellers of milk). Within each sector, we then identified three or four interventions for in-depth study based on similar criteria.

Figure 1 Examples of systems change within the AAER framework

Source Authors' own, adapted from Nippard, Hitchins and Elliott (2014).

Box 1 Top-down and bottom-up lenses

The **top-down phase** was designed to develop our understanding of the market contexts in which the programme had intervened, to deepen our understanding of the mechanisms through which the programme sought to influence these market systems, and to assemble existing evidence of change in the market system. This phase was designed to cover the first four steps of the typical contribution analysis cycle, drawing largely on programme data and secondary sources, but with additional primary research among key informants. This helped us to refine the research agenda for further data collection.

The **bottom-up phase** was designed to collect evidence of change in the market system introduced by individual interventions 'on the ground'. It harnessed a range of research methods appropriate to the different types of programme participants: among poor producers we used household-level quantitative surveys and participatory Qualitative Impact Protocol (QuIP) interviews (see Box 2); among market-level participants (business, associations, and government agencies) we used semi-structured qualitative interviews following a snowball sample).

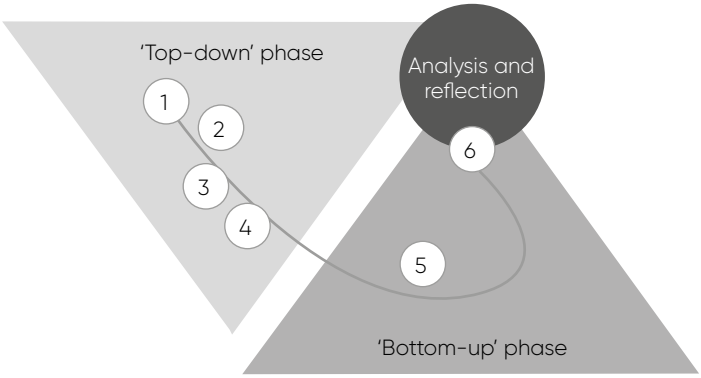
Source Authors' own.

In terms of what would constitute evidence of systemic change against the AAER framework, and taking our constraints into consideration, we were looking for evidence within the AAER quadrants which would suggest a subtle change of role or 'function shift' of some of the market actors (Fowler and Lomax 2021). We were also looking for evidence of replication of the interventions by other market actors ('crowding in') and for similar shifts in thinking and action by market actors in other areas of the same value chain. Figure 1 highlights the type of change that would constitute elements of systemic change against the AAER quadrants, with examples drawn from the dairy sector for illustration.

5 How did we identify early signs of systemic change?

Our approach built on one that Itad first developed for the Growth and Employment in States (GEMS) project in Nigeria to capture change at different levels of a system and then to attribute this change where possible to MSD interventions (Ruffer 2012). During the inception phase of the Samarth evaluation, we structured our approach into two broad phases, '**top-down**' and '**bottom-up**'. We mapped these research phases onto the standard phases of contribution analysis as illustrated in Figure 2.

Figure 2 Steps in contribution analysis illustrating our top-down and bottom-up phases



Step 1 Set out the attribution problem.

Step 2 Develop a theory of change and identify the risks to it.

Step 3 Gather existing evidence on the theory of change.

Step 4 Assemble and assess the contribution claim and the challenges to it.

Step 5 Seek out additional evidence.

Step 6 Revise and strengthen the contribution story.

Source Authors' own, adapted from programme documentation.

Structuring our research in this way served a similar purpose to the 'helicopter' and 'intervention' lenses recently described by others in their *Pragmatic Approach to Assessing System Change* (Posthumus *et al*. 2020).

At the end of the process, we put the two lenses together through a process of reflection and analysis to understand not only **if** interventions (or combinations of interventions) had influenced change in the system (and the role played by the programme), but also **why** this change had occurred, and if this change was likely to be sustained and scaled, even if evidence for this was nascent.

By structuring our research into these top-down and bottom-up phases we expected that, collectively, our varied methods would build a good picture of how the market system(s) were operating and why certain interventions within these systems appeared more likely to generate lasting change than others.

5.1 Top-down phase

5.1.1 Defining and understanding the system under study

During the top-down phase of research, we first developed results chains for our selected sectors and interventions (in consultation with programme staff and drawing on programme documentation and reporting). These built on the programme's existing results chains developed as part of the Donor Committee on Enterprise Development (DCED) audited results measurement system. However, it added further detail in order to capture not only the programme's vision for system change in each sector, the role played by different interventions in supporting this vision and key assumptions and risks, but also external factors, including the perceived role and influence of other actors. These complexity-aware results chains (Britt 2013) therefore took into account what were thought to be relevant factors in the broader environment, including other market players and drivers of change, and

captured our initial understanding of the parameters and dynamics of the market systems that the programme sought to influence.

Our dairy results chain covered Samarth's work to improve incomes for poor households through improved market access and improvements in productivity and milk quality by adopting good manufacturing practices (GMP) and improving animal husbandry. We captured the roles and responsibilities of the most significant actors who were either engaged directly in these interventions or had a relevant role in the wider milk production and marketing system. As such, it covered smallholder producers and milk aggregators (including milk cooperatives and private regional-level milk processors), national-level milk processors of milk and cheese products, national-level dairy associations, and the various government agencies responsible for setting milk prices and defining and enforcing milk quality standards. While recognising and appreciating the 'soft' boundaries of market systems, this range of actors acted as a *de facto* delineation of the dairy market system, for the purposes of our evaluation.

At the end of this stage, we developed an initial set of sector-specific research questions, which translated our evaluation questions into a set of more tailored research questions for each sector based on our sector and intervention results chains.[7] These guided the next stages of our research.

5.1.2 Top-down research

Our top-down research phase consisted of a review of secondary documents (in particular, Government of Nepal and donor reports) and a series of high-level interviews and workshops with sector key informants to get a 'birds-eye view' of the sector. These key informants were drawn from the public, private, and non-governmental sectors.[8]

The purpose of this phase was to deepen our understanding of system dynamics in our selected sectors and to focus on topics such as recent socioeconomic history and key development trends, as well as contemporary political economy and incentive structures for key actors. As part of this research, we interviewed our key informants about changes and developments in the sectors and the relationship between these changes and the programme's activities.

In the dairy sector, the top-down phase of research helped us to identify the key opportunities and challenges faced by the sector as a whole and in doing so to understand the aims, objectives, and reported achievements of programme interventions in this context. It also enabled us to develop a more informed and nuanced research agenda for the bottom-up phase of research to follow.

Our initial top-down research focused on validating programme information, filling evidence gaps, and providing us with a foundation from which to assess programme contribution to change. This confirmed that the dairy sector had experienced rapid growth in annual demand for milk and milk products (estimated at 8 per cent p.a.), which had outstripped supply from domestic sources and offered opportunities for producers to increase production. It also confirmed some of the key constraints in the sector related to production, value chain coordination, and the enabling environment. These constraints ranged from supply-side constraints such as inconsistent knowledge and application of GMP to regulatory constraints such as poorly enforced national quality standards. These constraints combined to tighten profit margins for milk producers despite the increasing national demand for the product.[9]

5.1.3 Learning from the top-down assessment of systemic change

A strong and evaluable sector-level results chain or theory of change (ToC) that is 'complexity-aware' is critical to the research process. We initially aimed to use the programme's overarching ToC and its sector results chains. However, these did not systematically and explicitly capture the expected system-level changes in all sectors. Nor did the overarching ToC articulate the relevant system constraints, assumptions, and risks which might have a bearing on these changes. The ToC diagram and results chains needed to be much more granular. Therefore, the research team spent more time than anticipated in creating 'nested results chains' in each sector in which groups of results chains for selected interventions fed into one 'complexity-aware' results chain for each evaluated sector. This proved to be a useful and necessary investment. It enabled us to define the systems we sought to study, understand their wider dynamics and strengthen our understanding of the intended role of the programme in driving change. These results chains provided the backbone for our subsequent research agenda and the platform against which we later combined our top-down and bottom-up lenses during our reflection and analysis phase.

A top-down preparation and research phase is particularly useful to gain a broader perspective on the system(s) under study and to help the evaluators get 'up to speed'. Our top-down research helped us to quickly understand the dynamics of the systems under study. We found the combination of programme reporting, secondary document review, and interviews and workshops with sector key informants to be a particularly effective and efficient way of building this broad but detailed picture. The analysis of systems change is highly context-specific and requires in-depth knowledge of the sectors under study and the roles, responsibilities, and incentive structures of key actors. The top-down phase helped us to build a more detailed picture of the key assumptions and risks present in programme delivery

across the sectors and alternative explanations for change, given the wider dynamics present in the sectors.

A careful selection of programme areas for evaluation is vital. Evaluators seeking signs of systemic change may need to delve more deeply into fewer programme areas, rather than adopting a wider but shallower approach. Systemic change itself takes time, and a deeper study is more likely to uncover these early signs. The inherent risk of a shallower approach is that the evaluated programme areas do not have sufficient depth to be able to reveal any signs of systemic change. This makes the initial selection process vital for evaluation success. To maximise the likelihood for success and opportunities for learning from this evaluation, we prioritised sectors in which the programme had been working for longer periods of time and had reported stronger early signals of systemic change. As a result, we decided not to include a number of more recent programme sectors (for example, the fish sector). Only where programme sectors were important from an economic and portfolio perspective – but where we determined that these had not been designed according to MSD principles, such as tourism – did we adopt a lighter-touch approach.

Evaluators need to be careful and deliberate in defining and delimiting the boundaries of the system(s) under study. Beyond selection of programme areas for study, a key finding for the team was that evaluators also need to be careful in how they define the systems selected for study and, within a finite resource envelope, not to overcommit in the research phase. Many potential avenues of research may open up during the top-down phase (and later during the bottom-up phase); the evaluator needs to make explicit choices as to where to focus their energy and resources. For example, in each sector the team made numerous decisions concerning which other potential drivers of change they should investigate and to what depth. In the dairy sector, for example, the team determined that Danida (Danish International Development Agency) support to develop guidelines for improved milk production was directly related to the aims and objectives of Samarth's interventions and was worthy of further investigation. By contrast, the team determined that the new relationships that Samarth had facilitated among market players in the sector had not been significantly impacted by another programme, the United States Agency for International Development's (USAID) large-scale Agricultural Growth Programme for livestock.

5.2 Bottom-up phase

5.2.1 Research strategy

Our bottom-up research phase involved a mix of different research methods designed to meet the varied aims of our evaluation.

Box 2 The Qualitative Impact Protocol (QuIP)

The Qualitative Impact Protocol (QuIP) draws on contribution analysis. QuIP's approach places project beneficiaries' voices at the centre of the evaluation, enabling them to share and feed back their experiences in an open, credible, and respectful way. QuIP gathers evidence of a project's impact through narrative causal statements collected directly from intended project beneficiaries. Respondents are asked to talk about the main changes in their lives over a pre-defined recall period. They are prompted to share what they perceive to be the main drivers of these changes, and to whom or what they attribute any change – which may well be from multiple sources. In some applications of QuiP, the researchers do not know for which project the analysis is being done, limiting bias in deciding on contribution claims.

Source Authors' own based on the method developed by Bath Social and Development Research (2021).

We conducted **quantitative surveys** of programme participants (treatment groups) and compared the results with groups of non-participants with similar characteristics (comparison groups). Our aim was to determine with confidence whether programme interventions had resulted in producers adopting the new practices, whether these changes had become embedded in the market and were sustained, and whether they were continuing to produce benefits for poor households in terms of increased productivity and income (all core aims of the MSD approach). Data from these surveys also supported our understanding of the resilience and gender equity of the 'adapt' phase of market system change.

In the dairy sector, we undertook two rounds of data collection with 500 producer households, which were consistent with, and incorporated baseline and endline data collected through Samarth's own monitoring system. The evaluation team aimed to add further rigour by providing additional resources to increase sample sizes beyond those used normally by the programme; we added comparison groups and a further round of data collection after the programme had ended to better assess the sustainability of programme interventions.

Paired with these quantitative surveys, we commissioned **Qualitative Impact Protocol (QuIP) studies** (see Box 2 for further detail on the method) to understand the issues faced by dairy farmers in greater depth and to identify the most important challenges, obstacles, and drivers of change. We proposed to triangulate this information with other evidence on the

contribution played by programme interventions that we had collected in both the top-down and bottom-up research phases. The design of the QuIP studies also enabled us to explore these issues in a participatory way from the perspective of smallholder dairy farmers on the ground, especially women and marginalised groups. In the dairy sector, our QuIP study comprised 24 interviews with individual respondents and four focus group discussions in two locations – one in a programme district and one in a matched control location.

Finally, since we aimed to understand the broader impact of interventions beyond the direct sphere of programme influence, we deployed a **semi-structured survey that asked qualitative information of market actors** following a 'snowballing sample' technique. These interviews started with market actors who were directly engaged in our selected interventions and had been identified with the support of programme staff, and then expanded outwards from there. We used insights gained through the initial interviews and the team's own market intelligence to identify additional actors who were thought to have reacted to, and potentially replicated, new practices as a result of the programme interventions. In the dairy sector, this included interviews with local milk producers, aggregators and traders, regional and national-level processors, producers' cooperatives, and local government agencies.

We opted for this semi-structured survey of market actors based on the expectation that evidence of systemic change would be limited and would need to be carefully identified and 'unpicked'. We were also faced with resource constraints and out-of-date and/or incomplete lists of market actors operating in programme districts and neighbouring locations. Unfortunately, there had been very limited record-keeping on the part of Samarth in terms of 'adjacent' actors that they had engaged with tangentially or indirectly during programme implementation. This made identifying credible sample frames difficult. In this context, we determined that quantitative surveys of a sample of market actors would not be feasible or cost-effective. This was one of the reasons why a snowballing sample technique was more feasible than a sampling strategy, which attempted to draw a representative sample of market actors from a pre-determined sample frame.

During our qualitative surveys, our researchers played the role of 'detective'; they tracked down systemic change 'leads', often based on partial or contradictory information, until they were satisfied that they had developed a fairly accurate picture of the change that had occurred and what factors were driving it. In situations with potentially conflicting data, asking probing questions of all the respondents and carefully corroborating this evidence was important, especially where the evidence was incomplete and potentially based on interviewees' incomplete recall over a time horizon of several years.

This combination of methods helped us to build up a picture of change in the market system 'on the ground' and to understand the programme's role in sparking this change, both within its sphere of direct influence and more broadly. While the qualitative interviews of market actors were undoubtedly the most important method in identifying and following up specific evidence of changes in the system beyond the direct influence of the programme, each method had a role to play in developing our understanding of the market and the programme's place within it. This is discussed in Section 5.3 on analysis and synthesis.

5.2.2 **Learning from the bottom-up perspective of systemic change**

The qualitative interviews with market actors placed particular demands on the team. We learned that researchers need to be knowledgeable about the MSD approach and be highly versed in the aims and objectives of the interventions and sectors they are studying. Experience in conducting qualitative research was particularly important so that researchers could effectively step into the role of 'detective' to identify, follow up on, and substantiate 'clues' as to the role of the programme, while at the same time knowing when to stop. In practical terms, we found that there needed to be consistency in the research process, with researchers being brought in as core members of the team for both data collection and analysis. Where we hired in researchers for a few interviews, the quality of the evidence suffered.

The reality of our research was 'messier' than anticipated and our bottom-up research did not always follow on neatly from the top-down phase as initially envisaged. From a practical perspective, a number of external factors complicated our research timings. These included agricultural cycles and intervention close-out activities which in some instances dictated that our bottom-up research needed to commence more rapidly than anticipated and overlapped with the conclusion of our top-down research phase. In some cases, our understanding of system dynamics was not as strong as we would have preferred and we missed opportunities to ask salient questions in our surveys.

We also found that our understanding of the system and its boundaries continued to evolve throughout our research. In practice, we continually updated our intervention results chains and research questions as subsequent research activities deepened our understanding of the role of key actors and important aspects of context; in some cases, we identified additional strategic top-down activities (additional research and interviews with sector key informants, for example) as a result. In sum, our research was much more iterative and irregular than we first envisaged.

Nepal's fluid political situation arising from the ongoing process of decentralisation proved to be challenging in a practical sense (although it did hold broader opportunities). It proved

Figure 3 The evaluation team plot evidence for dairy sector interventions on a white board

Source Authors' own.

challenging during our bottom-up research phase to identify and track down key respondents in local government who may have moved position following their central role with the programme. This resulted in numerous dead ends, with little value. However, it also presented an opportunity to contact a range of respondents who we may not have ordinarily identified and who were willing to provide their own thoughts and commentary on the sector situation and the role of different interventions in their geographic area of responsibility. This allowed us a far broader view of the practical realities on the ground, as well as a more comprehensive, unfiltered view of the regulatory environment governing the sector from the bottom up.

5.3 Analysis and synthesis: putting the 'top' and 'bottom' together

Given the challenging context for this evaluation, the final phase of analysing and synthesising the data was of central importance to building the programme's contribution story. Throughout the data-gathering process we had identified strands or threads of evidence that pointed to early indications of systemic change, which included hints and nuances in interviews and glimpses into the manner in which 'business was being done differently'. The phase of synthesis and analysis aimed to gather these hints, nuances, and threads of evidence, triangulate them, and weave them into a coherent evidence-based contribution story.

To do this, we reverted to the results chains we had developed during the top-down phase. We engaged the whole evaluation team in a participatory exercise to visually map the evidence we had collected through our research phases onto the result chains (see Figure 3). This mapping process was an iterative and inclusive

process, held over a few days, allowing all team members to record their thoughts and interpretations of data and evidence. All team members were encouraged to challenge the data, the strength of evidence, and even the positioning of the data within the visual map. All of this added value in testing the strength of evidence for early signs of systemic change, where this change seemed to be happening, and the contribution played by the programme.

Throughout the mapping process, we colour-coded the evidence by source and strength to aid the triangulation process and paid particular attention to evidencing the key assumptions made by the programme. Merging the wider, more inclusive top-down evidence with the more fine-grained evidence collected bottom-up allowed us an opportunity to critically examine our own 'data picture'. We could then take a step back to scrutinise this picture and to understand the selected interventions' influence within the system as a whole. This allowed us to consider other explanations of change in light of what we knew about the context in which they were operating.

Synthesising the data from both the top-down and bottom-up approaches in a visual manner assisted us in understanding the dynamics and external factors within the market system. The process was particularly useful in helping us to contextualise the changes introduced by the programme and to make an assessment as to the extent to which changes within the market system were likely to be sustained and/or scaled. This process sometimes led to a satisfactory understanding of the programme's contribution to change; on other occasions it raised further questions which led to additional data collection, either through further top-down document review and interviews, or additional interviews from the bottom up.

In the dairy sector, evidence from our various research methods helped us to identify a clear difference in the depth and sustainability of market system change brought on by first- and second-phase interventions. For example, although interventions in the first phase had successfully encouraged smallholder producers and the cooperatives and processors that purchase their milk to adopt changed practices, leading to increased productivity and milk quality, our quantitative household surveys revealed that adherence to these practice changes had begun to erode over time and the expected impact on household income had not emerged.

During the synthesis we had hoped that our QuIP studies would provide a source of data triangulation to understand the programme's contribution in bringing about these changes, but this proved difficult given the challenges encountered in isolating the precise role of programme interventions within the QuIP data. This resulted from a number of factors including the nature of

Box 3 A new business model in the dairy sector

In a second-phase intervention in the dairy sector, Samarth provided support to national-level processors (in the form of brokering new linkages with suppliers, providing technical advice and some financial resources) to enable them to form new working relationships and adopt new business models with regional suppliers of fresh milk and semi-processed dairy products. The national-level processors were encouraged to provide technical support to these regional processors to introduce new production practices, with the expected benefit of improved quality and consistency in the supply of raw milk to these national-level processors. In turn, the regional suppliers were encouraged to work with their smallholder producers of milk to support them to adopt new practices to improve raw milk quality, underpinned by the incentive of increased prices.

Source Authors' own.

the QuIP design, especially the 'blindfolding' of researchers, the indirect nature of the MSD approach (working at arm's length through the market), and the complexity of the research setting (in which multiple interventions were operating, often with similar objectives).[10] Nevertheless, the QuIP data (alongside evidence from our qualitative survey of market actors and data from the household surveys) helped us to better understand the contextual factors that undermined the durability of the changes introduced into the market system by first-phase interventions and that limited their scope.[11]

A key factor to emerge through the synthesis from both top-down and bottom-up evidence was the impact of increasing production costs. Programme participants identified increasing costs as a primary factor in eroding incomes from milk over time. The costs had not been offset by the expected 'price premium' from the production of higher-quality milk. This resulted in reduced willingness on the part of producers to sustain new practices. This, in turn, highlighted the importance and impact of a key contextual factor identified during the top-down research: out-of-date and poorly enforced national quality standards for milk which the first-phase interventions in the dairy sector had been unsuccessful in addressing. This undermined incentives for milk aggregators and processors to offer higher prices on a sustained basis for higher-quality milk.

The picture to emerge through the evidence synthesis for second-phase interventions in the dairy sector was quite different. There were early signs that a number of these interventions were gaining

traction in the market system and were contributing to changes that held the prospect of sustainability and scale. In particular, evidence from our qualitative survey of market actors revealed that a second-phase intervention that had facilitated the introduction of new linkages and working practices between large national-level dairy processors and regional-level suppliers of processed milk and cheese products (see Box 3) had resulted in both of these actors adopting changes to their practices to the benefit of smallholder producers. Those changes were sustained beyond the end of the financial support offered by the programme.

In this intervention, these new working relationships seemed to be proving to be durable because of the way that incentives had been aligned along the value chain: our qualitative survey of market actors revealed that the regional processors were continuing to implement improved practices. They were also continuing to work with smallholder producers after programme support had been withdrawn in response to the prospect of continued access to new, large, and higher-value urban markets. For their part, national-level processors were also continuing to support these suppliers in order to secure access to additional supplies of higher-quality and more consistent milk products.[12] In doing so, they had overcome the twin constraints of increasing production costs and weak official quality frameworks.

There was also evidence from the qualitative surveys of market actors, albeit nascent, that this intervention had gained traction in the market and held the potential for scale. The national-level processors engaged by Samarth were seeking to adapt the model (for instance, by offering finance to support smallholder producers to adopt practice changes to improve the quality of milk produced) and to expand their relationships to other regional processors. Regional processors meanwhile were taking further steps to adapt their production processes to the demands of urban markets and offer new processed and semi-processed milk products. In addition, other national-level dairy factories were seeking to crowd in with similar versions of the model described in Box 3.

It should be noted that Samarth was not the only programme to support some of these market actors – other programmes had offered equipment to at least one of the regional milk processors, for example. One of the key questions debated during the synthesis, therefore, was the degree to which Samarth could claim to have contributed to these changes. Based on evidence from multiple interviews with market actors and set against evidence collected through the top-down phase, the team concluded that Samarth had indeed made a very significant contribution, given that it could claim credit for initiating and supporting the development of the new working relationships between national and regional milk processors – which was the critical factor in explaining the intervention's success.

6 Concluding reflections on the process as a whole

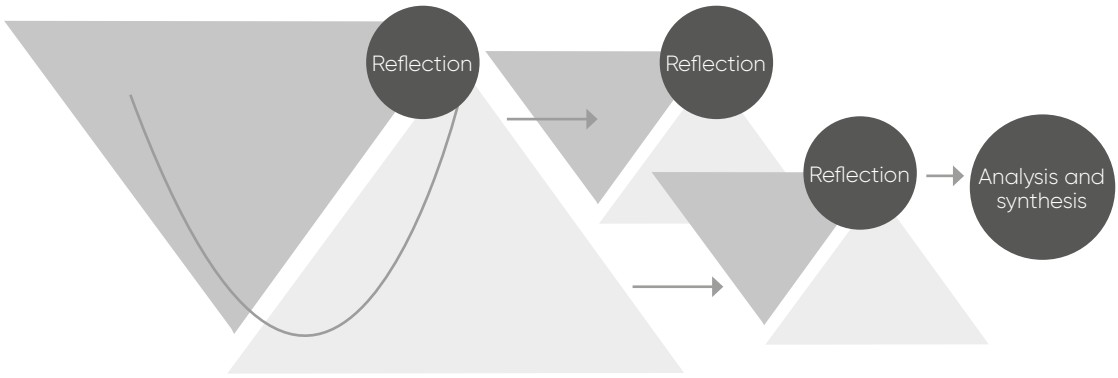
Overall, we found that our contribution analysis-based approach, framed by top-down and bottom-up research lenses, assisted us in identifying early signs of systemic change. Contribution analysis provided an appropriate analytical framework to navigate a complex environment (Mayne 2011), characterised by long and indirect results chains and multiple overlapping interventions and actors. The 'thinness' of the market and the impact of two huge, unanticipated external events (the trade embargo and an earthquake) would have significantly impacted our ability to identify a plausible 'business as usual' counterfactual through the results of the quantitative survey only, given its focus on inferring impact through with–without measurements in treatment and comparison groups only (Stern *et al*. 2012).

Our top-down and bottom-up research phases provided complementary inputs to help us determine if and how the systems we were studying had changed, the likelihood that these changes would be sustained and scaled in the market, and the programme's role in driving this. Faced with the initial programme ToC and results chains which were not 'systems-aware' and only providing a very loose definition of the market system, the top-down phase enabled us to develop more detailed impact logics for our chosen interventions; to define more clearly the boundaries of the systems in which they were working; and to deepen our understanding of these systems' characteristics (in essence, the contextual backdrop against which the programme was seeking to effect change). This work helped us to develop a detailed research agenda for our bottom-up phase.

Through our bottom-up phase, we progressively collected evidence of change in the market system against our interventions' results chains, drawing on a range of methods. We found that an exploratory, flexible qualitative survey of market actors, based on a snowball sampling approach, was an appropriate technique to identify and follow up on emergent evidence of systemic change, especially where it lay beyond the immediate scope of programme influence. However, this approach required skill on the part of researchers, who needed to know enough about the context and the intervention to develop relevant questions and have the ability to know where to probe further and where not.

We had initially planned that our contribution analysis framework would follow what might be referred to as a 'linear spiral' of ongoing evaluative activity, with our top-down and bottom-up lenses forming discrete phases of research, followed by a period of reflection and analysis. However, one of our key findings was that the reality of our evaluation implementation was much more iterative and 'messier' in reality, as illustrated in Figure 4. For example, we found that our understanding of the market systems in which our interventions operated continued to evolve during

Figure 4 **'Messy' reality of applying our top-down and bottom-up research lenses**



Source Authors' own.

the bottom-up data collection phase – as we learned more about context and the roles and responsibilities of key actors, for instance – and this necessitated further top-down research activities and periods of reflection.

Nevertheless, we found that our top-down and bottom-up framing added considerable value, especially during the analysis and synthesis phase. By combining a wide-angle view of the system as a whole with a narrower, more detailed perspective on change from the bottom up, we were able to take a step back to understand the selected interventions' influence within the system as a whole. This took into account context and other influences and explanations of change, which ultimately enabled us to identify those interventions that were demonstrating early signs of sustaining and scaling change in the market system.

In this case, we applied contribution analysis to identify embryonic traces of systemic change and to weigh up evidence from multiple perspectives as to whether the programme had contributed to the existence of these changes. Trying to define and identify systemic change in any MSD programme is challenging. However, we may take refuge in Justice Stewart's[13] wisdom that often we may 'know what it is when we see it'.

Notes

1 Edward Hedley, Principal Consultant and Evaluation Director, Itad, UK.
2 Gordon Freer, Evaluation Team Leader, Department of International Relations, University of the Witwatersrand, South Africa.
3 Edward Hedley is a Principal Consultant in Itad's Inclusive Growth and Climate Change Practice and was Project Director for this evaluation. Gordon Freer is an independent evaluator and was Team Leader for this evaluation.

4 At the time of programme inception, the MSD approach was widely known as Making Markets Work for the Poor (M4P).
5 While there is no clear definition on what constituted systemic change, four perspectives of different types of systemic change are presented in Jenal (2019). See also Nippard *et al.* (2014).
6 Out of a total of 11 programme sectors which included pigs, dairy, fish, feed, vegetables, ginger, mechanisation, crop protection inputs, agriculture reconstruction, tourism, and media.
7 These questions covered the relevance, effectiveness, impact, and sustainability of the selected interventions, including questions around the extent to which they had produced systemic change, either singularly or collectively.
8 These key informants were selected based on their anticipated knowledge of the sector, including the work of Samarth and other programmes. These key informants included individuals who had a direct relationship with Samarth and those who did not in order to get an external perspective on the work of the programme. This latter group included representatives of other donor organisations, other government departments, and other 'apex' firms not engaged directly by Samarth.
9 A comprehensive review of dairy sector constraints is provided in Pant *et al.* (2017).
10 The role of the QuIP studies in our evaluation and some of these challenges encountered are discussed further in an upcoming Centre for Development Impact Practice Paper co-authored with Bath SDR.
11 Factors included increasing production costs and poorly enforced milk quality standards which served to undermine incentives to improve the quality of production.
12 Depending on the timing of intervention close and our research, evidence indicated that these actors were still implementing improved practices between 12 and 18 months after the end of programme support.
13 *Jacobellis v Ohio* 378 US 184 (1964).

References

Bath Social and Development Research (2021) (accessed 8 October 2021)

Britt, H. (2013) *Complexity-Aware Monitoring*, Discussion Note, Washington DC: United States Agency for International Development

Chen, H-T. and Rossi, P.H. (eds) (1992) *Using Theory to Improve Program and Policy Evaluations*, New York NY: Greenwood

Chen, H-T. and Rossi, P.H. (1983) 'Evaluating with Sense: The Theory-Driven Approach', *Evaluation Review* 7.3: 283–302

Chen, H-T. and Rossi, P.H. (1980) '**The Multi-Goal, Theory-Driven Approach to Evaluation: A Model Linking Basic and Applied Social Science**', *Social Forces* 59.1: 106–22 (accessed 23 August 2021)

Fowler, B. and Lomax, J. (2021) 'Measuring what Matters: Monitoring and Results Measurement', in J. Ledgerwood (ed.),

Making Market Systems Work for the Poor: Experience Inspired by Alan Gibson, Rugby: Practical Action Publishing

Itad (2019) ***Samarth-NMDP Programme Final Evaluation Report***, Oxford: Itad and e-Pact (accessed 17 January 2022)

Jenal, M. (2019) *Measuring Systemic Change in Market Systems Development – A Stock Taking*, Washington DC: United States Agency for International Development

Mayne, J. (2011) 'Contribution Analysis: Addressing Cause and Effect', in K. Forss, M. Marra and R. Schwartz (eds), *Evaluating the Complex: Attribution, Contribution, and Beyond*, Brunswick NJ: Transaction Publishers

Nippard, D.; Hitchins, R. and Elliott, D. (2014) *Adopt-Adapt-Expand-Respond: A Framework for Managing and Measuring Systemic Change Processes*, Briefing Paper, Durham: Springfield Centre

Pant, S.R.; Acharya, M.; Sah, A.K. and Pandet, G. (2017) 'Present Status, Problems and Strategies for the Development of Cattle Farming in Nepal', Proceedings of the 10th National Workshop on Livestock and Fisheries Research in Nepal, Lumle, Kaski, Nepal, 5–7 March

Posthumus, H.; Shah, R.; Miehlbradt, A. and Kessler, A. (2020) ***A Pragmatic Approach to Assessing System Change***, Boekel: Hans Posthumus Consultancy, Springfield Centre, Miehlbradt Consulting, and Donor Committee for Enterprise Development (accessed 23 August 2021)

Ruffer, T. (2012) ***GEMS Results Measurement Handbook***, Hove: Itad (accessed 20 September 2021)

Stame, N. (2004) 'Theory-Based Evaluation and Types of Complexity', *Evaluation* 10.1: 58–76

Stern, E. *et al.* (2012) ***Broadening the Range of Designs and Methods for Impact Evaluations: Report of a Study Commissioned by the Department for International Development***, Working Paper 38, London: Department for International Development (accessed 23 August 2021)

Assessing the Contribution to Market System Change of the Private Enterprise Programme Ethiopia*

Giel Ton,[1] Ben Taylor[2] and Andrew Koleros[3]

Abstract An impact evaluation of a pro-poor market system development programme, 'Making Markets Work for the Poor' (M4P), poses several methodological issues for evaluators. M4P interventions intend to change the contextual conditions in which stakeholders take business decisions so that it triggers change processes in the wider social system and ultimately benefits poor people. An impact evaluation design for such a programme thus needs to be robust enough to adequately capture these systemic outcomes, acknowledging dynamic changes in intervention delivery as well as in market conditions over time. Theory-based evaluation can provide learning and accountability when it incorporates methods that allow a critical reflection on the key causal steps in an intervention's theory of change. We present our learnings about indicators and methods to reflect on the importance of the contributions to market system change of a large M4P programme in Ethiopia.

Keywords market systems, value chain development, mixed methods, job creation, theory-based evaluation, monitoring and evaluation, logframes.

1 Background

In interventions aimed at catalysing change in a complex system – that is, systems where multiple actors act and interact with each other and the wider environment to bring about change, such as economic sectors in a country, transformation of a political system, or mitigation of climate change – it becomes unreasonable and methodologically challenging to assess the relative effectiveness of the support of one of these actors to wider systems changes, let alone credit this to a single contributor (Earl, Carden and Smutylo 2001). The systemic effects that are measured at the system-wide level are well outside the sphere of direct influence of any one actor group, and hence any direct

The *IDS Bulletin* is published by Institute of Development Studies, Library Road, Brighton BN1 9RE, UK
This article is part of *IDS Bulletin* Vol. 53 No. 1 February 2022 **'Theory-Based Evaluation of Inclusive Business Programmes'**; the Introduction is also recommended reading.

intervention with that actor group. For example, if wider systemic outcomes were to decline over time, it would be unreasonable to blame an intervention for this negative change.

Moreover, an implementer would come up with countless reasons as to why this negative change could not be attributed to this intervention. The reverse logic is also true; if wider systemic outcomes are positively changing, it is unreasonable for the intervention to claim the credit for an improvement. Nevertheless, commissioners of evaluations often want to have an idea about the size or importance of a contribution to change at this systems level for multiple reasons, including the need to present this information at an aggregate level; for example, to account for their commitments to the Sustainable Development Goals. This creates the paradox that commissioners of impact evaluations pose legitimate but unanswerable questions about the precise size of their contribution to wider systems changes. As such, impact evaluators are often faced with identifying ways to reconcile the impossible with the possible (Ton *et al*. 2019).

In this article, we present learning from an attempt to do so in the area of market system development. The Private Enterprise Programme Ethiopia (PEPE) was an ambitious, £69m programme, funded by the UK's Department for International Development (DFID, now Foreign, Commonwealth & Development Office – FCDO), designed following the Making Markets Work for the Poor (M4P) approach. M4P is an approach to developing market systems so that they function more effectively, sustainably, and beneficially for poor people, building their capacities and offering them the opportunity to enhance their lives (Elliott, Gibson and Hitchins 2008; Nippard, Hitchins and Elliott 2014). By addressing underlying causes (rather than symptoms) of weak performance of the market system, M4P aims to unleash large-scale change. Interventions may be small in themselves but are expected to leverage the actions of key market players to bring about extensive and deep-seated systemic change (Tschumi and Hagan 2008).

The M4P activities of PEPE were implemented from 2013 to 2020 by a consortium led by the global consultancy firm DAI through a team based in Addis Ababa, named Enterprise Partners (EP). EP supported M4P innovations in three priority sectors (leather, textiles, and horticulture) and provided technical assistance to the financial sector to improve access to finance by micro, small and medium enterprises (MSMEs). Additionally, DFID contracted an external evaluation team, led by the consultancy firm Palladium. DFID had ambitious objectives both with PEPE and its external evaluation. It was intended to be the first evaluation to include an *ex post* analysis five years after programme completion, to capture the scale and sustainability of the innovations developed through the M4P approach within the wider market systems over a longer time horizon.

The total budget for the external evaluation was £2m and covered the costs of ten process evaluations to assess the outputs and progress of EP on an annual basis (referred to as annual reviews by DFID), and four impact evaluation milestone reporting events, at baseline (2016), mid-term (2018), endline (2020), and *ex post* (2024), to assess the outcomes and impact of EP's support to wider market systems changes in the three priority sectors: leather, textiles, and horticulture. In addition to the data-gathering exercises conducted by the evaluation team, EP had its own monitoring and results measurement (MRM) system following best practice in MRM system design (DCED 2017), which consisted of intervention logics for each support component along with progress indicators that were reviewed quarterly by programme teams to drive intervention adaptation and pivots (Enterprise Partners 2020; Yohannes 2020).

In this article, we present lessons learnt from this ambitious monitoring and impact evaluation effort. It starts with a brief introduction to the main features of M4P programmes. This includes a discussion of the adaptive management that is required for these programmes to work effectively, and the challenges this presents to rigorous impact evaluation, particularly when the desired impact is a wider sector-level change, such as job creation. Section 2 then discusses two major challenges faced by evaluators: defining and capturing early signs of systemic change; and designing an appropriate mix of methods to reflect on the importance of the contribution of the programme to these changes. Next, Section 3 illustrates how our impact evaluation addressed these challenges. Finally, Section 4 reflects on the results and Section 5 draws conclusions with advice for commissioners of impact evaluations in these types of programmes.

2 Challenges in the evaluation of M4P programmes

The ambition of M4P programmes is to find leverage points in market systems that change the dynamics in these systems in a way that more poor people benefit. That is, an M4P programme wants to trigger the motivation of firms and other stakeholders to innovate existing production, service delivery, or transaction practices in order to improve the functioning of the market system while including more poor people within markets. Consistent with wider principles around how change happens in complex adaptive systems such as economic systems (Beinhocker 2006), these changes rarely follow a dose-response relationship: the amount of effort or investment in an activity is not proportional to the size of the outcome; a small change in one actor or institution can cause a dramatic shift in the overall systems performance. These innovation processes often involve many stakeholder groups, and each will have a different perception of the related risks and rewards.

M4P programmes try to find leverage points by multiple activities, such as organising brokering events, elaborating proposals for

a policy change, or peer-learning activities around experiments and innovation. However, market systems change continuously and M4P programmes might find that the relevance of these activities shift or fade over time; what promised to be a leverage point may cease to be one once there is a shift in the market constellation. Due to this uncertainty inherent to complex systems (Snowden and Boone 2007), M4P programmes need to adapt and improvise, trying multiple activities while making sense of the ripple effects – experimentation and innovation is inherent to M4P programmes.

Moreover, the effects of support activities can manifest themselves much later. For example, an event where various sector stakeholders meet for the first time, such as producers and processing companies that discuss the strategies to improve the quality of raw material inputs, may appear to be fruitless in the short term, when they do not reach any common agreement on the ways to tackle the issue. However, this 'fruitless' activity may have resulted in personal networks between persons and organisations that lead to important systemic effects several years later, when the same persons contact each other for a rapid response to a policy proposal in the sector. What first could appear as being an insignificant event may prove a key event in the explanation of significant outcomes some years later. Evaluators of M4P programmes need to find ways to capture these unpredictable outcomes as a result of multiple, adaptively managed activities.

Implementers of M4P programmes, of course, have strong economic and reputational incentives to attribute results to themselves. They are often international consultancy companies that rely in their business model on successful projects or, at least, satisfied commissioners. They will have a tendency to overestimate their contribution. This requires evaluators thus to critically scrutinise both the rationale of the support activities and the evidence that supports the contribution claims (Stern *et al*. 2012; Mayne 2019). In M4P programmes – due to the multitude of activities – there is almost always a contribution to changes at the direct beneficiary level, often through business service providers or beneficiaries of innovation grants. The more interesting, but also more contestable, claims are usually related with its contribution to the performance of firms in the sector, such as increased trade, employment, or value addition in the sector, among firms that are not directly supported by the intervention and lie outside the sphere of direct influence.

Evaluators must thus find ways to verify whether the support can indeed be considered as a contributing factor in the wider configuration of changes among actors and other external factors over time that produced the observed outcome in performance. This implies a structured process of critical counterfactual thinking about alternative explanations of the

change process (Spellman and Mandel 1999; Stern *et al*. 2012; Yin 2013), and answering the question whether it is plausible – as assumed – that the support has been a non-redundant component in the configuration of causes that resulted in the outcome (Mackie 1974; Shadish, Cook and Campbell 2002; Mahoney 2008).

Even when the evidence suggests that an M4P programme has contributed to a systemic change in the market, this does not answer the question about the importance and effectiveness of this public investment in for-profit private actors. Donors need to aggregate and compare programmes at a higher, portfolio level, for example when deciding on new programme priorities in the region (ICAI 2015). John Mayne (2019: 4) indicates various ways for reflecting on the relative importance of the intervention's efforts in bringing about change in comparison to other factors. However, understanding the relative contribution of a support programme in one particular complex change process (= a causal configuration), is not enough for this portfolio analysis; there is still a need for some sort of ranking of various, alternative programmes (= multiple configurations) according to the size or importance of the outcome that resulted. Commissioners legitimately ask evaluators to give them an idea of the size of the impact to make priorities in future programming and budget allocation decisions.

This outlines two big challenges in the impact evaluation of M4P programmes. First, it is difficult to pinpoint what a systemic change in markets is, and how to 'capture' and monitor the early signs of it with sensible indicators. Second, impact evaluators need a research design that not only verifies whether a programme contributed to this change, but also helps to reflect on the importance of this contribution to judge the relevance for future funding of similar programmes.

3 Impact evaluation of PEPE

In this section, we describe how we attempted to address these two main challenges through the design of the impact evaluation of PEPE. As mentioned above, PEPE intervened in three priority sectors of the Ethiopian economy: leather, textiles, and horticulture; as well as interventions in the cross-cutting financial sector. DFID had selected these sectors in 2013 because of their potential for sector-wide transformation and poverty reduction, and followed the priorities defined in Ethiopia's Growth Transformation Plans (Diriba and Man 2019). For example, the focus on horticulture linked to the ambition to support the large number of newly commercialising smallholder farmers who could increase their incomes. With the focus on labour sourcing for industrial parks and the development of the leather manufacturing value chain, DFID expected to provide employment to part of the growing youth population.

3.1 Theory of change and logframe indicators

The overarching theory of change of PEPE and its accompanying logical framework (DFID 2018) were developed over time through a participatory process between the programme commissioners, the implementers, and evaluators, and provided the framework for the logframe indicators; i.e. how the programme was intended to report on progress and outcomes over time to DFID. Figure 1 is a simplified version of PEPE's theory of change; the full logframe differentiates between agro-industrial sectors and financial services, and included two non-M4P output areas, related to work with the International Labour Office and the Ethiopian Competitiveness Facility (a grant fund), and the external evaluation, led by Palladium. The backbone theory of change depicted in Figure 1 has a linear dimension that shows the intention to create higher level outcomes related with poverty alleviation through sector-level outcomes by multiple activities and outputs that are intended to find the leverage point in the market system.

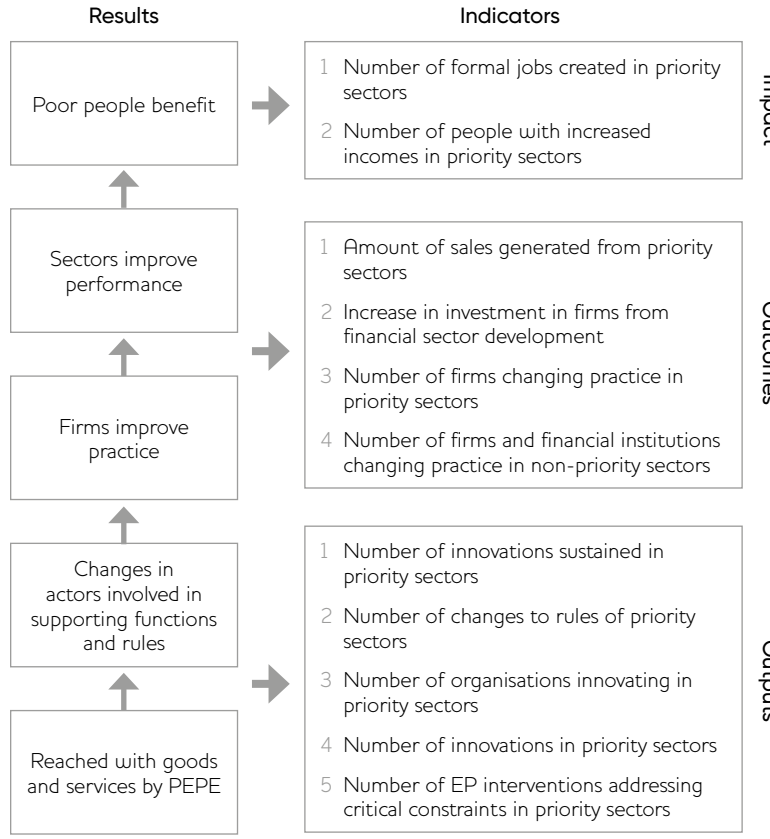
The logframe indicators align with the principles of M4P and allow outputs even when these do not directly lead to outcomes and impact. Changes in market systems result from non-linear processes and multiple activities need to be trialled to find leverage points that shift the system to a higher level of performance.

Following the M4P logic, all of these interventions were designed to trigger an innovation, but it was acknowledged that only part of these innovations would be successful in doing so and result in a significant and sustained change in practices of core market firms. EP describes its M4P approach as:

> a process of 'facilitated muddling through' [that] can only take place if a programme is set up as a learning organisation, able to adapt and respond to the context it engages with. Core to adaptive management in an MSD [market system development] programme is an ability to design innovative partnerships, map out how they are expected to work by means of a theory of change, test whether this theory holds true by means of continuous monitoring, and use the insight generated to engage stakeholders.
> (Bekkers 2020: 47)

The PEPE theory of change served as the backbone of EP's monitoring and result measurement system (Enterprise Partners 2020) but was much more granular in practice. Before starting with an activity, EP mapped the market system and developed more granular, actor-based theories of change (Koleros *et al*. 2020) with indicators and targets related to the specific subsector. Across the lifespan of EP, there were a total of around 100 interventions, each with their own intervention logic and activity plan.

Figure 1 PEPE intervention logic and logframe indicators



Source Authors' own, simplified from the PEPE logframe (DFID 2018).

EP developed a programme-level monitoring and results measurement (MRM) system that followed best practice guidance in the field (DCED 2017; Posthumus *et al*. 2020). The MRM team in EP generated annual, bi-annual, and quarterly data on the results of each of the programme's interventions. Each sector team had a person responsible for monitoring the results: the MRM person generated real-time data and analysis to support intervention managers to make decisions, and managers in turn provided qualitative input to data-gathering activities. Quarterly half-day workshops for each sector team created an opportunity for staff to provide input into each other's decision-making (Enterprise Partners 2020; Yohannes 2020). The MRM system was third-party audited by the Donor Committee for Enterprise Development (DCED).

Below we provide a more detailed description of the logframe indicators designed to capture the systemic nature of M4P (see Figure 1), and how these were operationalised by EP and the external evaluation team. Progress on the indicators was

self-reported by EP, using its MRM system, and reviewed by the independent impact evaluators during the annual reviews before being submitted to DFID.

3.1.1 Output indicators

As described above, output indicators were designed to report progress around how EP's interventions supported the development and sustainability of an innovation or rule change within the priority sectors. In the logframe (DFID 2018: Output 1), 'innovation' was defined as 'a change in the way a supporting function works in response to a critical constraint identified in the sector strategy'. Changes in rules included policies enacted, standards revised, regulations released, strategies validated, and directives or other rules enabled that address critical constraints in the relevant sectors. The term 'sustained' meant that the innovation continued for a minimum of 12 months after the end of direct support to the intervention. It was assumed that 50–70 per cent of innovations would be sustained, recognising that it would take two years after the start of the innovation before they could be reported as sustained.

For all output-level changes, EP developed a results chain showing how it contributed to the new innovation or rule change. The evaluation of the outputs sometimes implied expert judgements about what was considered as being an innovation – where to draw the boundary? Generally, the differences in judgement between EP and the evaluation team were small at output level, reflecting an unwritten rule that allowed flexibility in outputs as long as at least some of these delivered outcomes.

3.1.2 Outcome indicators

Much more discrepancy between the assessments of EP and the evaluators was present when the evaluation reflected on outcomes that resulted from these outputs. The logframe acknowledges that the indicators around investment posed particular problems for attribution. This is because investment is often a significant decision for a company, which is made based on many factors – not just EP. The guidance provided in the logframe, therefore, suggested that EP would seek to assess attributable investment where possible. Where it reports contribution, it isolates the specific investment that it has contributed to, and explains how this contribution was made, rather than reporting the whole investment. This obviously opened up a discussion between EP and the independent evaluation team, and indicated the methodological rigour required from the latter when verifying the reported outcomes.

3.1.3 Impact indicators

In spite of a consensus between implementers and evaluators that net effects of M4P programmes are only meaningful when applied on outcomes that are still in the sphere of influence of the intervention – firms that change their business practices in

response to the support – DFID maintained an interpretation of impact as being 'additional jobs created' and the 'number of smallholders that increased their incomes by a minimum of 20 per cent' (DFID 2018: Impact 1) as this was the basic metric on which the programme was awarded to DAI through a competitive procurement process. The logframe clearly asked for an estimate of the plausible size of the impact that resulted from their £69m investment in PEPE. The evaluation team, therefore, had to come up with a research design to do the impossible with the possible, and decided to give well-reasoned plausible range for this impact, instead of point estimates.

3.2 Impact evaluation design

Throughout the programme period, but particularly in the first half, the external evaluation team functioned as technical back-stoppers to the programme, with the annual reviews as the key moments of interaction. The accountability question became more dominant in the second half of the programme, from 2017 onwards, not least because the logframe targets became partly linked to a 'payment for results' element (DFID 2015). The methodological design for the impact evaluation was a learning process that can be divided into three phases, each associated with a different core method to assess and quantify the outcomes and impacts along the theory of change. Each phase had a methodological design that was revised and approved by DFID's Evaluation Quality Assurance and Learning Service (EQuALS).

3.2.1 Phase 1 – baseline

At baseline (2015/16), the core method proposed was to measure net effect in sector performance through changes in constraints in a firm's business environment. This followed the logic used to assess the value for money in the DFID Business Case by estimating the induced growth of value added at sector level. The impact evaluation team planned to use different data set analytical methods (econometric methods, social network analysis, and qualitative comparative analysis) that could show that the sector performance was associated with EP-induced changes in the perceived severity of constraints. The data needed included questions that asked for business performance (sales, employment, exports), and modules to identify and rank a long list of constraints and incentives that affected a firm's decision-making to invest, and asked respondents to rank their importance, similar to the World Bank Enterprise Surveys.

The results of the baseline survey, however, proved somewhat unsatisfactory. While the response rate of the survey was good and covered most formally registered small and medium-sized enterprises (SMEs) in each sector, not all the surveyed firms answered all questions related to the business constraints which made it difficult to aggregate results. Moreover, the team concluded that the baseline constraints prioritisation exercise was too imprecise to be used at mid-term as the core method

for inferences about the impact of EP support, for three main reasons. First, the persons that would respond at mid-term and endline about the firm's status and constraints could well have changed. Second, the prioritisation of constraints was unlikely to capture longer-term, more structural changes in the market system over time. Third, the data collected about the firm's economic performance were highly incomplete, as firm owners did not always want to give the exact figures – which often resulted in lacking or unreliable data (e.g. often, the enumerator was told to come back to interview other staff, even when this later proved to be unfeasible).

3.2.2 Phase 2 – mid-term

In 2017, based on this baseline survey experience, the methodology was revised. Compared with phase 1, the qualitative and quantitative research were much more interlinked in a mixed-method design that allowed triangulation of findings. Additional to the survey, the evaluation introduced process tracing case studies (see Figure 2), in the subsectors where EP claimed that the most significant outcomes had occurred. These process tracing case studies assessed the strength of the evidence for the claim that EP's outputs were a necessary, non-redundant component in the configuration of factors and actors that caused the change at outcome and impact level – the (early signs of) systemic change. The evidence for the claim was provided by EP and complemented by additional interviews and analysis by the evaluation team.

The case studies verified the logic of the result chain and, for example, probed whether the service providers had indeed improved their services due to EP support or would have provided these services to the firms in any case. To assess the strength of the contribution claim, we used four leading questions in a process of logical reasoning about the counterfactual situation (Ton and Glover 2019). These questions were adapted to the specific case at hand:

1 Did the change occur?

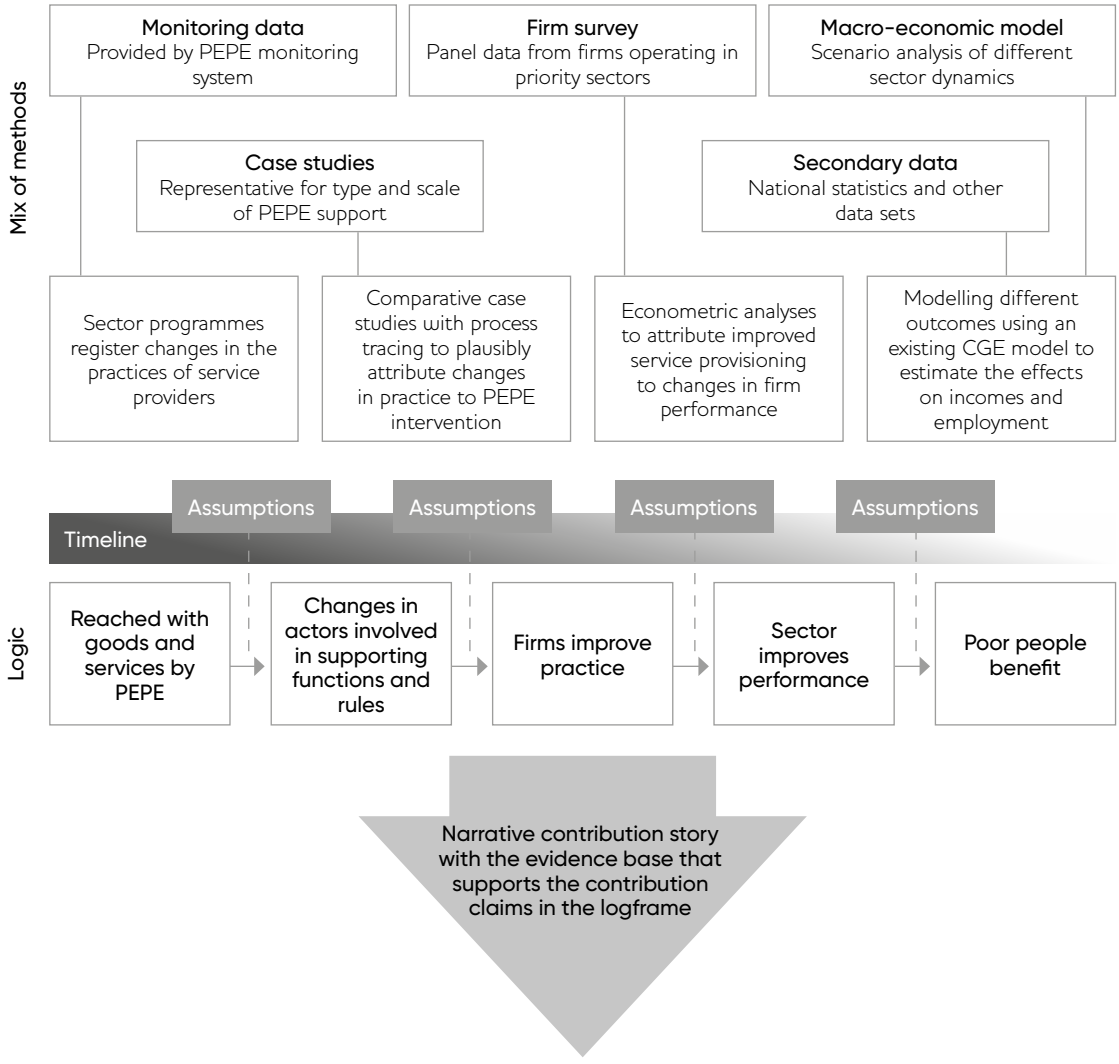
2 Did it result from a process in which PEPE-supported services were used?

3 Can this support be considered as a necessary (non-redundant) causal factor for that process to have taken place? And, if not, was it a necessary causal factor in accelerating or scaling of the outcomes?

4 Are there any other institutions or programmes that may have provided similar support to the change process, if the PEPE-supported services had not been present?

Each case study took around 25 days of research. At mid-term, this included a week of interviews in Ethiopia, and at endline, it included a series of online interviews. Most time, however, was

Figure 2 Methodological design at mid-term



Source Authors' own, based on Koleros, Taylor and Ton (2018).

spent on reviewing the documentation provided by EP in the 'evidence pack' distilled from their MRM system, and additional information provided by EP at the request of the evaluators. The case studies, especially at mid-term, explored the sustainability and importance of the effects of EP's activities at output and outcome level. At endline, the case studies focused primarily on the claims related to investment being mobilised and the impact level, the jobs created, and the number of smallholders with at least 20 per cent improvement in income.

Another adaptation to the methodology between baseline and mid-term in order to more accurately measure firm-level

changes was a complementary way to assess the firms' business performance. In order to address the real-world problem of incomplete survey data on the financial performance of firms that often need to operate in the grey area between the formal and informal in relation to the payment of taxes, the team decided to include a second, complementary way to ask for the performance change, using a less threatening way of asking about business performance.

Instead of relying only on the formal reported figures for before-after estimates of impact, the survey introduced questions that asked for perceptions of change in these performance indicators over the last three years. The scale had four intervals to indicate an increase or decrease in sales, exports, and employment: 0–25 per cent; 25–50 per cent; 50–100 per cent; more than 100 per cent. The information was based on the firm manager's perception of the change over the last three years, without requiring the exact figures of this change. This resulted in complete data on these percentual estimates and, again, a high number of missing values for the formal, absolute numbers from the financial statements of the firms. The survey resulted in data on 335 firms.

Moreover, the baseline modules for the prioritisation of constraints were substituted by modules that asked, for each of the 23 constraints, two perception questions that could be used to compute 'contribution scores' (Waarts *et al*. 2017; van Rijn *et al*. 2018). The perceived change in the severity of the constraint (using a five-point Likert scale) was combined ('multiplied') with the information about the perceived influence of the EP-supported service providers on this improvement (also a five-point Likert scale) into a contribution score. This was a sector- and constraint-specific list of business service providers and government institutions provided by EP's MRM team. These contribution scores can be interpreted as the 'perceived impact of EP-supported service providers on the constraint/outcome'. The average contribution score, considering all relevant constraints for which the perception questions were asked, was converted in percentage points and could fluctuate between 0 (no change or no influence) and 100 per cent (large change with a large perceived influence). The contribution scores allowed comparative analysis and subgroup analysis to detect meaningful differences in outcome pattern between types of firms and between sectors in statistical analyses.

3.2.3 Phase 3 – endline

The Covid-19 pandemic and related budget and logistical constraints forced us to make a change in the survey method. In 2020, when the endline survey was held, it was decided to limit the sample to only firms that were likely to have been within the influence of the EP-supported service providers. The 2020 endline survey covered 74 firms that had been in contact with one or

more EP-supported service providers. The Covid-19 pandemic and lockdowns affected the firms. The perceived change in performance, therefore, used the question 'Imagine the situation that the Covid-19 pandemic had not affected your firm. Can you give an estimate of the percentage change [in sales/exports/profits] that you would have had, without Covid-19, compared with three years ago?'

While the mid-term analysis of the contribution scores was largely restricted to map the differential impact of PEPE between different types of firms and between different sectors, for the endline evaluation, we went a step further and used them to assess the outcomes in sales, exports, and profits. To estimate a plausible treatment effect, we converted the 23 contribution scores for each of the 74 firms into seven support components using principal-component analysis. These components were used in regressions to test their association with the outcomes. For those sectors where a component proved significant (and only when the case studies confirmed the contribution claim), the coefficient in the regression was used as a scenario in the macro-economic CGE model of the Ethiopian economy (Tebekew *et al*. 2015) to estimate the lower and upper bounds of the EP-induced employment effects in the economy.

During the annual reviews in 2019 and 2020, it became evident that EP's M4P interventions managed to meet the output targets in the logframe but that these outputs did not (yet) result in the outcomes and impacts that were expected at the start. Most of the job creation was due to the support to a labour-sourcing innovation in Hawassa Industrial Park and the financing of women entrepreneurs and SMEs in two programmes funded by the World Bank that had received the direct technical assistance of EP. The challenge for the PEPE evaluation team was to come up with a reasonable way to estimate the EP-attributable effects within the total effects of these large multi-donor programmes. Therefore, two of the three process tracing case studies at endline verified the contribution claims related to the work in the financial sector.

3.3 Results

The impact evaluation resulted in an endline report that synthesised and combined the information from the methods depicted in Figure 2 (Koleros *et al.* 2018; Ton *et al*. 2021). The report shows that PEPE managed to reach their output targets but that this did not result in the expected level of outcomes and impact. Six case studies estimate, for each case, the causal effect of the activities and outputs on these outcomes and explain how this is backed up by the evidence and data generated in the MRM and the critical verification by the external evaluation team. For reasons of space, we illustrate the results by zooming in on only three of the six markets where PEPE claimed to have made a contribution to outcomes and impact.

These three cases best exemplify how the evaluation methods helped to critically verify the contribution claim of the implementer. The cases concern the support to labour sourcing in industrial parks, the development of agent-based seedling propagation models for smallholders, and the activities to develop a private capital investment advisory market. The texts in these three cases are taken from a much more comprehensive analysis of the cases in the endline report (Ton *et al*. 2021) and illustrate the way inferences were made.

3.3.1 Seed and seedling market

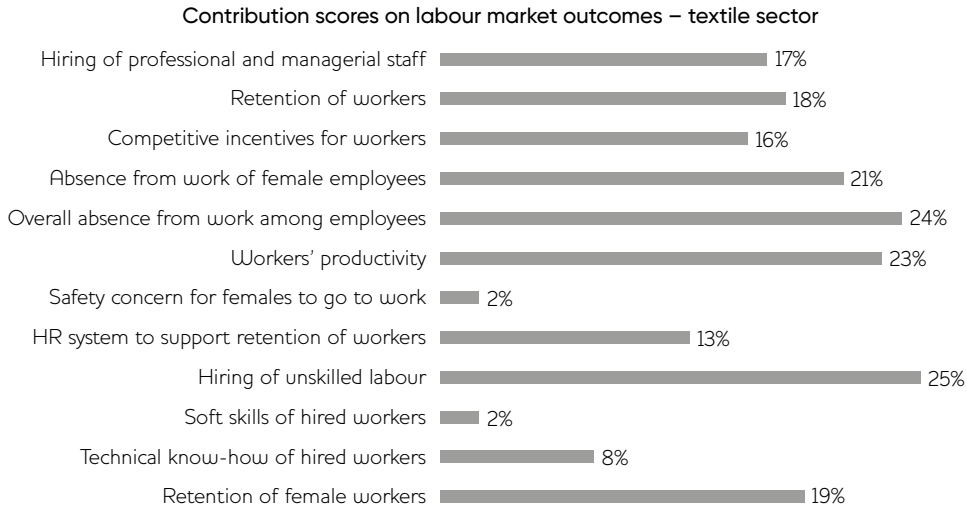
In the vegetable and fruit sector, PEPE identified that smallholder farmers struggled to consistently procure good-quality seeds and seedlings. With PEPE support, 14 propagators set up satellite nurseries in remote locations, involving around 400 agents who provided on-farm extension support to farmers that bought the seedlings. The propagators reached the vegetable farmers mainly through model farmers. Each of these was assumed to reach around five fellow farmers, who learn through demonstration effects. For the seed distribution model, the MRM system included a comparison between participating and non-participating farmers that suggest income increases in vegetable producers of more than 30 per cent.

However, the sample included mainly model farmers who are more likely to receive better training and technical support, and to have established stronger market linkages. In the seedling distribution model, farmers adopted improved fruit tree seedlings. However, it is not yet certain whether these farmers will have an income increase of 20 per cent in the future, because the income rise is still uncertain and contingent upon the continuation of care of these trees and future harvests. Taking both considerations into account, the endline report estimates that the vegetable seedling programme improved the income of a minimum of 3,416 farmers and a maximum of 17,082. The difference is due to this uncertainty in spread of the innovation beyond the model farmer.

3.3.2 Labour sourcing in industrial parks

An important theme of discussion in several annual reviews related to the way that jobs were created by the innovative labour-sourcing system in Hawassa Industrial Park (HIPSTER[4]), where EP had helped to establish a system of sourcing and grading of labourers to meet labour demand by the textile manufacturers that started operating there. Hawassa is the first and largest industrial park in Ethiopia located in a region where the potential workers are primarily located in rural villages. Consequently, Hawassa Industrial Park had unanticipated problems in attracting sufficient workers for the (textile) factories. The case study concluded that EP had effectively become part of the problem-solving task force to address issues with labour in Hawassa.

Figure 3 Perceived contribution of EP support to labour market outcomes in the textile sector



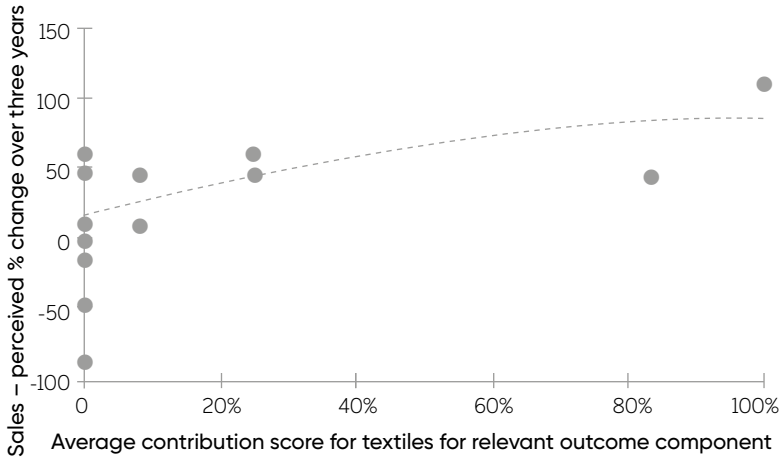
Source Authors' own, using endline data (Ton *et al*. 2021).

Since its start, more than 70,000 workers in Hawassa were screened and 60 per cent were also graded based on their skills. The problem for the impact evaluation was that the labour-sourcing system was mandatory for all firms in the industrial parks that required workers. The screening and grading component in HIPSTER is where PEPE has put most efforts. The screening and grading component was, however, also the component that did not appear to have worked particularly well in Hawassa and has not been replicated in other, more recently established industrial parks, which carry out combined sourcing and screening through local government departments.

The contribution scores (Figure 3) suggest that there is almost no perceived effect of EP support on soft skill but a small to fair effect on absentee reduction and workers' productivity. To assess the impact, the evaluators used the definition of job creation in the intervention logic that envisioned job creation through increased productivity. This means that it is inappropriate to count the number of workers that went through the sourcing system as jobs created by EP. However, HIPSTER has undoubtedly increased the effectiveness of the system. Without having a point estimate, the evaluation estimated that this efficiency is most likely not less than 4 per cent and not more than 10 per cent of the jobs that were created in Hawassa Industrial Park.

The contribution scores also helped to reflect on the performance-enhancing effect of EP, and an econometric regression showed that, for textile firms, the support was associated with an increase in sales of 2 per cent per year (Figure 4). Both elements were used to estimate the total

Figure 4 Association between contribution scores and performance in the textile sector



Note The curved line, instead of a linear one, being the most approximate to the regression results.
Source Authors' own, using endline data (Ton *et al*. 2021).

additional induced job creation with the CGE model in the range of 1,184 and 2,963 jobs in the priority sectors and an induced job creation in the national economy in the interval of 5,672 and 13,413 jobs.

3.3.3 Private capital advisory market

The case of the private investment advisory market was important for the outcome target related to investments in firms mobilised by new financial instruments, and illustrates the forensic approach used to assess the non-redundancy of the support in the complex process of systemic change. With a Private Capital Advisory Fund (PCAF), PEPE created a fund (initially intended as a revolving fund) for companies to hire investment advisors, who would help facilitate transactions by ensuring they meet the requirements of private equity investors in areas such as business plans, international financial reporting requirements, and valuations of assets. More than 30 grants were made available in 2018–19. In 2020, four companies were able to create improved business propositions and also attract a business partner, and generated a total investment of US$25m.

However, interviews by the evaluation team with the firms involved showed that in three of these cases, the advisors already had existing working relationships with companies prior to PEPE's intervention. While PCAF certainly helped to cover some of the costs associated with these deals, the deals themselves did not rely on PCAF. Moreover, PCAF saw uptake from companies only when it offered funds in the form of a grant rather than a loan, implying that there was no appetite for taking on the risk of hiring

an investment advisor. It was too early to ascertain any significant change in the market system. The case also highlighted the issue that the outcome indicators were not defined as additional (net effects) unlike the impact indicators. Therefore, even the small contribution of PCAF could be registered as investment mobilised in the logframe agreed upon with DFID.

4 Reflection

We found that the indicators used at output level in the PEPE logframe reflected the programme's theory of change and implementation approach reasonably well. However, they did not provide a particularly good accountability framework. The flexibility afforded at the output level – designed to allow evidence-based and adaptive programming – meant that it became possible to achieve the output targets by developing interventions that would never have impact on job creation or smallholder incomes at scale. And in some activities for which outcomes were reported, as in the case of PCAF, the importance of the contribution could be called into question. The disaggregate analysis of the logframe indicators helped the reflection on the importance of the different M4P components but the aggregates will hide nuances and lead to opportunistic, direct, programme-funded support activities with farmers and firms instead of the indirect support that characterises the M4P approach.

Monitoring of the ultimate outcomes and impact indicators is important for reflection on the relevance of the £69m investments of UK public development funds. However, estimating the net effects of changes in market systems that are well beyond the sphere of influence of an intervention is trivial and methodologically problematic (Ton, Vellema and Ge 2014). DFID insisted in its requirement for the evaluation team to quantify the net effects of PEPE on job creation. The sophisticated method developed to do so showed that PEPE's impact was far below targets, even when considering the higher bounds of the confidence interval.

Other donors, such as the Netherlands, decided to shift the focus in evaluating the importance of private sector development programmes away from this net-effect perspective and to ask for monitoring data about the aggregate sales and employment of all firms that were reached by a private sector intervention with a 'significant contribution' (DGIS-RVO 2017: 4). Monitoring the reach of a programme in relation to the number of firms, farmers, or jobs supported is far easier than computing net effects and still results in rough, indicative numbers that help to compare between programmes and interventions. Instead of requiring precise baseline–endline data with counterfactual designs, this requires research methods that evaluate the significance or importance of a contribution made by an intervention but without the need to quantify it, which appears, similar to what is argued by other

scholars (Goertz 2006; Mayne 2019), a better and more workable approach for evaluating the development impact of private sector development support, such as M4P programmes.

Especially in M4P programmes, there is a need to redefine what a rigorous impact evaluation design implies. Our experience showed that when the 'treatment' is highly variable, as inherent to M4P programmes, a treatment-comparison design to assess changes in outcomes or impact indicators is extremely vulnerable to changes in intervention modalities. We show that a careful analysis of change trajectories, and within the group of supported firms only, can yield a plausible estimate of impact, without the use of a comparison group. Estimates of the relative change over the last two years in sales, profits, and exports, combined with the perception questions used to compute contribution scores, proved sufficient to roughly estimate the impact of the support provided. We argue that asking directly for the perceptions of contributions or impact is a useful add-on to any survey that wants to capture M4P effects. Perception questions allow cross-sectional analyses and real-time reporting, can capture a wide range of outcomes, and help to build the resiliency of an impact evaluation design to changes in interventions, sample attrition, and evaluation conditions.

We learnt that rough measures of performance with high response rates are preferable over precise measures but with many missing data points. The competitive nature of firms makes it difficult to collect precise performance data. Therefore, even with relations of trust between the respondent and the enumerator and with well-crafted confidentiality agreements, missing data on sales, profits, and investments is notorious in firm surveys. Less precise but easier-to-collect data, for example asking for rough percentual changes in business performance indicators, as we did, helped to get a full data set that allows statistical pattern detection.

We maximised the potential to capture evidence/responses that could support the contribution claim but at the same time made it possible to critically assess the effects. The two core methods used, the firm survey, and the process tracing case studies, had features that allowed falsification of the claims. The survey did so quite straightforwardly, by asking the firm managers directly whether they used the improved services or regulations that addressed each constraint (see Figure 2) and, if so, how they rated the influence of these services in their business development. The contribution scores showed that only a few firms perceived a positive effect on these outcomes that they attributed to some degree to EP support. The average contribution scores in each sector rarely exceeded 16 per cent, which reads as 'a slight improvement and a slight influence' (Ton *et al*. 2021: 52).

The major drawback in the evaluation of PEPE related to the division of labour between the leading implementing company (DAI) and the leading impact evaluation company (Palladium). The contract stated that the evaluators were not allowed to influence the field activities and detailed interventions of the implementer, except by reflecting on the theory of change and the M4P approach. The evaluation's main task was to help DFID reflect on the effectiveness of the M4P approach and assist in fact-checking the reported progress according to the logframe targets. As inherent to *ex post* impact evaluations, the learning from the impact evaluation often comes too late to have a short-term follow-up. Also, in this case of PEPE, the decision to follow up was taken long before the results of the endline evaluation findings about the importance and size of the contribution to employment and smallholder incomes were available.

This delink between the endline evaluation outputs echoes the warning of the ICAI who warned that 'the more that evaluations are seen as prompts to evaluative and strategic thinking by programme teams, rather than products in their own right, the more useful they are likely to be' (ICAI 2015: 23). We think that as external evaluators, we could have done better in creating and feeding this strategic learning, continuing the more developmental evaluation process that characterised the annual reviews in the early stages of the PEPE programme. The logframe targets and external accountability became more important in the last years and, logically, created more sensitivities around the way the contributions to outcomes and impact were assessed and quantified. Together with personnel changes in EP, DFID, and Palladium, the decision to design a non-M4P programme as a follow-up, and the logistical challenges due to the Covid-19 pandemic, this translated into a more distant relationship between the stakeholders involved in the evaluation.

5 Conclusion

In private sector development programmes, where government funds are used to support private profiteering, the impact on development and public goods needs critical scrutiny; the risks of market distortion and corruption are simply too big to ignore. Therefore, we argue that it is legitimate to ask for an assessment of the size and importance of a contribution claim. However, computing a precise quantitative estimate of the size of a contribution is not possible. Nevertheless, as shown in the PEPE example, it might be possible to give a rough idea of the plausible range of effects that result from the support.

In the end, the quality of any evaluation design depends on the room for and quality of critical scrutiny (Patton 2012; Pawson 2013; Yin 2013). We argue that the critical scrutiny of contribution claims, articulated by the implementing stakeholder, and based on a reflection on the theory of change or intervention logic

provides a good starting and endpoint for an impact evaluation in M4P programmes. The theory of change provides the grammar for the contribution claims, and proper logframe indicators help to pinpoint the expected size and importance of the impact that is being pursued. However, due to contractual obligations and donor dependency, the targets specified in a logframe often take on a life of their own and activities are geared towards meeting the logframe's targets without concern for quality of the outputs and outcomes, and the nature of the impact.

We found that *ex post* process tracing of the most significant outcomes reported by the implementers is a method of critical inquiry and counterfactual reasoning that helps to balance the overreporting bias. *Ex post* process tracing is inherently resilient to changes in interventions, and economic and policy dynamics, including changes in the expectations and evaluation questions of the commissioners. It is especially useful when the contribution claim includes an outcome at the boundary of the sphere of influence (e.g. the ultimate outcomes) where the causal arrow is important but contested (or uncertain).

We argue that the commissioners could do better in prioritising methods and sense-making events that can inform the discussion around impact at mid-term in the terms of reference, instead of the current emphasis on rigorous impact evaluation designs that only produce evaluative insights at endline. Our advice for future impact evaluation in M4P programmes is threefold: verify the logic of the contribution claims with critical, forensic research methods; take perceptions of firm managers seriously; and refrain from point estimates of outcomes and impact but use minimum and maximum bounds of plausible effects. In sum, combine 'good-enough methods' with critical, evaluative, and counterfactual reasoning, to feed iterative learning cycles, involving the implementers, commissioners, and evaluators together in reflecting on the importance and logic of the evolving intervention logics of a programme.

Notes

* The authors were contracted by Palladium. Andrew Koleros was team leader between 2013 and 2017 before he moved to Mathematica. Ben Taylor of Agora Global was a member of the evaluation team between 2013 and 2021. Giel Ton of the Institute of Development Studies (IDS) joined the evaluation team in 2016. The authors acknowledge the support of Seife Ayele, Marinella Leone, Dirk Willenbockel, Ayako Ebata, and Keir Macdonald at IDS and Deepti Sastry and Ciana-Marie Pegus at Palladium, and Adam Kessler and Shumete Belete for designing and implementing EP's monitoring and results management system. We also acknowledge the constructive peer review of Sietze Vellema at Wageningen University that helped us to shape and sharpen the arguments.

1 Giel Ton, Research Fellow, Institute of Development Studies, UK. Corresponding author: **g.ton@ids.ac.uk**.
2 Ben Taylor, CEO, Agora Global, UK.
3 Andrew Koleros, Senior Researcher, Mathematica Policy Research, USA.
4 Hawassa Industrial Park Sourcing and Training Employees in the Region (HIPSTER).

References

Beinhocker, E.D. (2006) *The Origin of Wealth: Evolution, Complexity, and the Radical Remaking of Economics*, Boston MA: Harvard Business School Press

Bekkers, H. (2020) *Enterprise Partners in Support of Industrial Transformation: Building an Industrial Labour Services Market in Ethiopia*, Enterprise Partners Case Study Series 2, Addis Ababa: Enterprise Partners

DCED (2017) *The DCED Standard for Measuring Achievements in Private Sector Development: Control Points and Compliance Criteria*, London: Donor Committee for Enterprise Development

DFID (2018) *Private Enterprise Programme Ethiopia Logframe 2018*, London: Department for International Development

DFID (2015) *Private Enterprise Programme Ethiopia Annual Review 2015*, London: Department for International Development

DGIS-RVO (2017) *15 Methodological Notes – Instructions for Calculation, Validation and Reporting of Performance Indicators*, The Hague: Dutch Ministry of Foreign Affairs

Diriba, G. and Man, C. (2019) *Building a Big Tent for Agricultural Transformation in Ethiopia*, CSIS Global Food Security Project, Washington DC: Center for Strategic and International Studies

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

Elliott, D.; Gibson, A. and Hitchins, R. (2008) 'Making Markets Work for the Poor: Rationale and Practice', *Enterprise Development and Microfinance* 19.2: 101–19

Enterprise Partners (2020) *Enterprise Partners' Monitoring and Results Measurement System and DCED Experience*, Addis Ababa: Enterprise Partners

Goertz, G. (2006) 'Assessing the Trivialness, Relevance, and Relative Importance of Necessary or Sufficient Conditions in Social Science', *Studies in Comparative International Development* 41.2: 88–109

ICAI (2015) *DFID's Approach to Delivering Impact*, London: Independent Commission for Aid Impact

Koleros, A.; Taylor, B. and Ton, G. (2018) *Independent Evaluation of the Private Enterprise Programme Ethiopia (PEPE): Midterm Evaluation Final Report*, London: Palladium

Koleros, A.; Mulkerne, S.; Oldenbeuving, M. and Stein, D. (2020) 'The Actor-Based Change Framework: A Pragmatic Approach to Developing Program Theory for Interventions in Complex Systems', *American Journal of Evaluation* 41.1: 34–53

Mackie, J.L. (1974) *The Cement of the Universe: A Study of Causation*, Oxford: Oxford University Press

Mahoney, J. (2008) 'Toward a Unified Theory of Causality', *Comparative Political Studies* 41.4–5: 412–36

Mayne, J. (2019) ***Assessing the Relative Importance of Causal Factors***, CDI Practice Paper 21, Brighton: Institute of Development Studies (accessed 13 July 2021)

Nippard, D.; Hitchins, R. and Elliott, D. (2014) 'Adopt-Adapt-Expand-Respond: A Framework for Managing and Measuring Systemic Change Processes', *Briefing Paper*, Durham: Springfield Centre

Patton, M.Q. (2012) 'A Utilization-Focused Approach to Contribution Analysis', *Evaluation* 18.3: 364–77

Pawson, R. (2013) *The Science of Evaluation: A Realist Manifesto*, London: SAGE

Posthumus, H.; Shah, R.; Miehlbradt, A. and Kessler, A. (2020) *A Pragmatic Approach to Assessing System Change*, Boekel: Hans Posthumus Consultancy, Springfield Centre, Miehlbradt Consulting, and Donor Committee for Enterprise Development

Shadish, W.R.; Cook, T.D. and Campbell, D.T. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Boston MA: Houghton Mifflin

Snowden, D. and Boone, M. (2007) 'A Leader's Framework for Decision Making', *Harvard Business Review* 85.11: 68

Spellman, B.A. and Mandel, D.R. (1999) 'When Possibility Informs Reality: Counterfactual Thinking as a Cue to Causality', *Current Directions in Psychological Science* 8.4: 120–23

Stern, E. *et al.* (2012) *Broadening the Range of Designs and Methods for Impact Evaluations: Report of a Study Commissioned by the Department for International Development*, Working Paper 38, London: Department for International Development

Tebekew, T. *et al.* (2015) *Ethiopia: Input Output Table and Social Accounting Matrix*, Addis Ababa and Washington DC: Ethiopian Development Research Institute and International Food Policy Research Institute

Ton, G. and Glover, D. (2019) ***Improving Knowledge, Inputs and Markets for Legume Expansion: A Contribution Analysis of N2Africa in Ghana and Ethiopia***, IDS Practice Paper 10, Brighton: Institute of Development Studies (accessed 13 July 2021)

Ton, G.; Vellema, S. and Ge, L. (2014) '**The Triviality of Measuring Ultimate Outcomes: Acknowledging the Span of Direct Influence**', *IDS Bulletin* 45.6: 37–48 (accessed 13 July 2021)

Ton, G. *et al.* (2021) *PEPE Evaluation: Endline Impact Evaluation*, London: Palladium

Ton, G. *et al.* (2019) ***Contribution Analysis and Estimating the Size of Effects: Can We Reconcile the Possible with the Impossible?***, CDI Practice Paper 20, Brighton: Institute of Development Studies (accessed 13 July 2021)

Tschumi, P. and Hagan, H. (2008) *A Synthesis of the Making Markets Work for the Poor (M4P) Approach*, Bern: Swiss Agency for Development and Cooperation, Federal Department of Foreign Affairs

van Rijn, F. *et al.* (2018) *Verification of PUM's Intervention Logic: Insights from the PRIME Toolbox*, The Hague: Wageningen Economic Research

Waarts, Y. *et al.* (2017) *Assessing IDH's Contribution to Public Good Impacts at Scale*, Wageningen: Wageningen University & Research and KPMG

Yin, R.K. (2013) 'Validity and Generalization in Future Case Study Evaluations', *Evaluation* 19.3: 321–32

Yohannes, L. (2020) *Adaptive Management: From the Inside Looking Out. Managing the Enterprise Partners Market Systems Development Programme in Ethiopia*, Addis Ababa: Enterprise Partners

This page is intentionally left blank

The Search for Real-Time Impact Monitoring for Private Sector Support Programmes*

Fédes van Rijn,[1] Haki Pamuk,[2] Just Dengerink[3] and Giel Ton[4]

Abstract An increasing consensus exists in the impact evaluation literature on using detailed theory-based evaluations to evaluate complex programmes such as private sector development (PSD) programmes. At the same time, PSD managers expect periodic and timely (so-called 'real-time') input from evaluators to improve programmes throughout their implementation. This article presents insights from real-time theory-based monitoring and evaluation shaped by the needs of policymakers in two Dutch PSD programmes. To learn about their experiences, we held in-depth interviews with researchers and policymakers involved in the evaluation. The interviews indicated that theory-based evaluation improved reporting on the programmes' contribution to higher-level impact areas and credibly quantified the importance of that contribution. The insights showed too that real-time monitoring and evaluation of PSD programmes requires more flexibility in data collection and increased interaction with mid-management.

Keywords private sector support, business coaching, economic development, theory-based evaluation, impact analysis.

1 Introduction

Private sector development (PSD) programmes aim to contribute to overall economic development through providing business support services (such as technical assistance, management provision, export training) or financial aid (Schulpen and Gibbon 2002). Over the past two decades, pressures on development budgets have increased the demand to show results at the impact level in order to legitimise public funding. While many PSD programmes have monitoring systems in place that collect information on outputs (for example, number of companies trained) and to a lesser degree on uptake (such as use of

The *IDS Bulletin* is published by Institute of Development Studies, Library Road, Brighton BN1 9RE, UK
This article is part of *IDS Bulletin* Vol. 53 No. 1 February 2022 **'Theory-Based Evaluation of Inclusive Business Programmes'**; the Introduction is also recommended reading.

training), they face difficulties when asked to report on the impact on business performance. PSD programmes, therefore, need to upgrade their monitoring system in such a way that they can respond to donors' expectations with regard to evaluation. The Donor Committee for Enterprise Development (DCED) recommends building a results management system that tracks, among other things, the effects of support activities as attributable changes to business performance, job creation, and export performance (DCED 2017).

To get quantitative estimates of impact that can be attributed to programme activities, most impact evaluation approaches (Khandker, Koolwal and Samad 2010) ask for the comparison of programme outcomes before and after the programme, and ideally between supported and unsupported firms. In the case of PSD programmes, impact evaluations usually require information on business performance indicators such as employment, sales, and exports before and after participation in the programme, and information to correct for contextual influences for a regression-based estimate. The programmes contribute to changes in high-level indicators such as employment, sales, and exports of firms, but these net-effect indicators of impact are not very actionable and useful for adaptive management during programme implementation, because at best, they are only available after some time (Apgar, Hernandez and Ton 2020). Often the programme management learns about the impact only when the decisions about continuation, adaptation, or finalisation of the PSD support have already been made.

Therefore, not many PSD programmes choose to rely on quantitative research designs for computing attributable net effects. Instead, most PSD programmes provide illustrative examples of their relevance at impact level in a more qualitative way, through case studies. However, case studies rarely provide a representative picture of the quality and impact of the portfolio of activities. The challenge for evaluators is, thus, to find practical ways to report the size and importance of the support that are lean enough to be incorporated into a programme's monitoring and evaluation (M&E) system for portfolio-level monitoring but are rigorous enough to result in credible estimates of the overall impact of the PSD programme to allow a reflection on its relevance, effectiveness, and efficiency. There is an increasing consensus among the quantitative and qualitative-oriented impact evaluators that for complex programmes – such as PSD programmes – programme theories need to be the backbone of an impact evaluation design (Chen 1994; Blattman 2008; White 2009; Bates and Glennerster 2017; Davey *et al*. 2018). In these theory-based evaluations, the data collection is designed in response to key assumptions in the programme theory.

The policy relevance of impact evaluations depends on the extent to which programme management has access to

these findings to refine and adapt their programmes. Ideally, information for monitoring, evaluation, and learning is shared throughout the implementation in 'real time', using methods and processes that enable adaptive management (Giordano 2017).

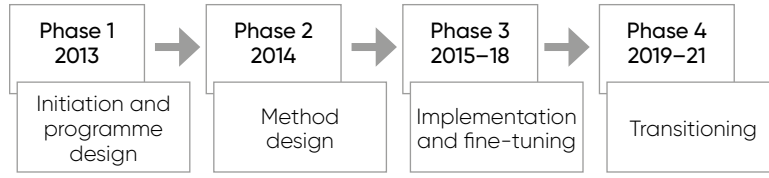
This article reflects on how theory-based mixed-methods impact evaluation can assess the importance and impact of PSD support in terms of offering accountability to the funders, while serving the information needs of programme managers. For this purpose, we distil lessons learned from the implementation of the Pioneering Real-time Impact Monitoring and Evaluation (PRIME) programme between 2013 and 2021. In PRIME, two large Dutch PSD organisations, the Centre for the Promotion of Imports from developing countries (CBI) and the Netherlands Senior Experts (Programma Uitzending Managers, or PUM), used similar tools to report on the impact of their PSD support. The programmes differ in aims but both have business coaching as a common approach. CBI promotes exports from developing countries through sectoral programmes that provide advice, counselling, and export market entry support to small and medium-sized enterprises (SMEs) and business support organisations. PUM organises business-level and sectoral missions that help SMEs to improve their business practices. Both CBI and PUM are funded by Dutch official development aid because they aim to generate additional (export) sales and employment in those countries, and therefore contribute to sustainable and inclusive economic growth.

To review the lessons learned in the implementation of our theory-based mixed-methods impact evaluation approach to assess the impact of PSD support, Section 2 explains the evolution of PRIME between 2013 and 2021. Section 3 reflects on the PUM and CBI programme managers' experiences with the approach and their assessments about its policy relevance. The section builds on information from in-depth interviews held in 2017 and 2021, and a workshop with a wider group of programme stakeholders conducted in 2017. We discuss user feedback and the main trade-offs and tensions that researchers and programme managers encountered in implementing PRIME. Finally, Section 4 provides recommendations for a better integration of theory-based impact evaluation and M&E systems of PSD programmes.

2 PRIME approach

The PRIME partnership was established in 2013 by CBI, PUM, the Erasmus School of Economics (ESE), and Wageningen University and Research (WUR), to develop and implement a methodology to monitor and evaluate the real-time impact of private sector development support by PUM and CBI. We distinguish four phases in PRIME (see Figure 1). In this section we describe each phase.

Figure 1 **The four phases of PRIME**



Source Authors' own.

2.1 Phase 1: initiation and programme design

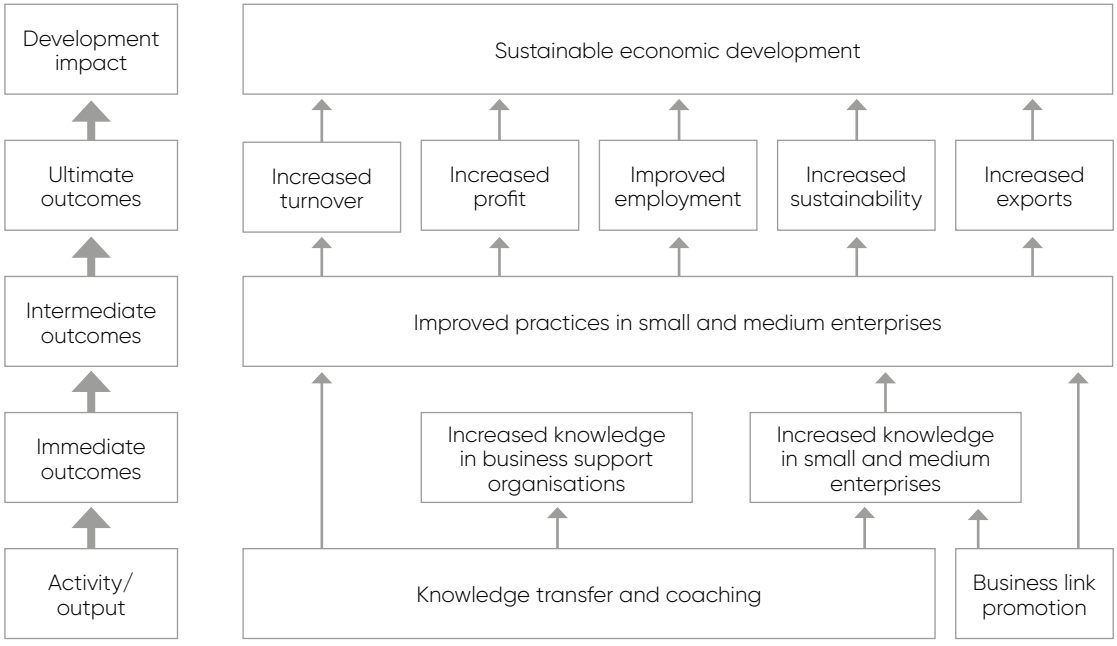
Following an instruction from the Dutch Ministry of Foreign Affairs in 2011 (DGIS 2011), all Dutch PSD organisations with a budget above €10m were made responsible for evaluating the impact of their work on sustainable development outcomes. The guidance also emphasised the need to show net effects of impact and the use of counterfactual research designs to do so. Many of these organisations struggled with this need and started to experiment with methods to generate credible evidence.

The idea for the PRIME partnership emerged in 2012. It was the fruit of informal discussions during a series of seminars hosted by WUR for the 'PSD Platform', where most Dutch PSD support organisations are represented. There were three reasons for establishing the PRIME partnership. First, the necessity of reporting the impact of private-sector support on the harmonised impact indicators defined by the DCED (jobs, revenues, and scale) – in other words, **accountability needs**. Second, the difficulty of going beyond 'before/after' measurements and the use of comparison groups – **a methodological need**. Third, a desire for meaningful impact evidence which can be used during the implementation of programmes – **a learning need**.

CBI and PUM, both prominent members of the Dutch PSD Platform, decided to address the challenges together as their organisations had complementary objectives. They also aimed to work more closely together and were accountable to the same governmental body and civil servants. The assumption was that a better understanding of each other's strengths, using a similar method for benchmarking effectiveness, would create synergies between both programmes. They approached Wageningen University and Research (WUR) and ESE to help them. WUR had a track record in developing evaluation methods for value chains in agriculture, forestry, fisheries, natural resources, and consumer markets. ESE had a track record in performance measurement and development of corporate social responsibility programmes of companies.

The organisational structure of the PRIME programme was designed to ensure the involvement and ownership of CBI and PUM, while at the same time maintaining sufficient independence

Figure 2 Simplified intervention logic of SME support provided by CBI and PUM



Source Authors' own.

to meet the quality criteria for external evaluations defined by the Dutch Evaluation Office (IOB). PRIME had a Programme Board, consisting of the managing directors from CBI, PUM, and higher management in WUR and ESE, plus an Advisory Committee, consisting of six external representatives, including the Ministry of Foreign Affairs, the IOB, the International Trade Centre (ITC), the International Initiative for Impact Evaluation (3ie), and two other knowledge institutes (Panteia and The Hague University of Applied Sciences).

During this inception phase, overarching theory of change (ToC) charts (using the term 'intervention logics') were developed to allow theory-based evaluation and to sketch the preliminary mix of core method. The research partners facilitated the process, as the concept of a ToC chart was not yet used by either organisation at the time. Several underlying assumptions for the change process were added to the chart, including identification of risks and plausible unintended effects, for the main causal links in the chart (Ruyter de Wildt *et al*. 2013).

The overarching intervention logic (Figure 2) was used to identify indicators at several outcome levels that could capture the effects of CBI's export promotion and PUM's business coaching. The ToC chart used the disaggregated outcome categories as suggested by Mayne (2001): immediate outcomes (knowledge);

intermediate outcomes (business practices); and ultimate outcomes (firm performance).

The programme design document (Ruyter de Wildt *et al*. 2013) proposed a mix of core methods plus additional methods and routines to anticipate the main validity threats to these (Ton 2012). The core methods were: (1) a literature review; (2) a cohort design to collect panel data; and (3) case studies in six countries to explore whether the support led to systemic change. The programme design was approved by the end of 2013, leading to the next phase of PRIME.

2.2 Phase 2: method design

In phase 2, the method of PRIME was operationalised. Most attention was on the development of indicators for the intermediate outcomes, as these were deemed to be less context-specific than the immediate outcomes and, therefore, have more generic characteristics that enable benchmarking. Ultimate outcome indicators (firm performance in terms of profit, employment, exports, and so on) are even more standardised, but, as anticipated by the researchers in PRIME, could be outside the span of influence in terms of time or programme effect (Ton, Vellema and Ge 2014).

Literature review. A key activity in this phase was a literature review on the current evidence regarding SME support (Harms, Ton and Maas 2014), which benefited from several extensive systematic reviews that became available at the time (Grimm and Paffhausen 2014; Piza *et al*. 2016). The literature review confirmed the plausibility of the ToC but indicated a lack of evidence about the assumed employment effects of PSD support at the sector and national level. The study also showed a list of indicators used by other scholars to track changes of intermediate and ultimate outcomes.

Survey design. PUM had already collected information on immediate outcomes, while CBI collected data mostly on the ultimate outcomes. Data on intermediate outcomes (business practices) were hardly collected. The indicators had to capture meaningful change and be general enough to be relevant for different types of firms that operate in various economic sectors. The immediate and intermediate outcome areas were related to seven different areas of business management as distinguished by CBI in their company auditing system.

To complete these data, we proposed an online survey to collect data, not only on ultimate outcome indicators (sales, exports, employment), but especially on the intermediate outcomes on knowledge and practices with a combination of self-assessment questions and observable business management practices. The self-assessment questions would give real-time feedback and the observable business management practices would help to

triangulate and validate these perceptions of impact over time. The yearly online survey asked the firm managers to assess the extent to which PSD support contributed to changes in the firm's business management. The survey was tested and adapted to be around 15 to 20 minutes, using the Qualtrics platform.

Case studies. Parallel to this, we designed the case studies in five countries. Considering budget constraints and following advice of the Advisory Committee – 'better do one thing good than two things flimsy' – we refrained from collecting additional data from the supported firms by subcontracted data collection firms, but focused on semi-structured interviews with beneficiaries, experts, and other key stakeholders.

2.3 Phase 3: implementation

Between 2015 and 2018, the PRIME evaluation methods were implemented. On the one hand, this meant collecting and analysing data; on the other hand, this meant regular interaction between researchers and staff from CBI and PUM. The research team had at least monthly interaction with the Consultation Group, quarterly face-to-face meetings with the Advisory Committee, and bi-annual meetings with the Programme Board. PRIME produced monitoring reports after each survey round or country visit, quarterly newsletters for wider audiences, and yearly research briefs for CBI and PUM. Reporting on such data on a regular basis was new for both organisations, and in the case of PUM, led to a special section in their yearly reports.

For PUM, out of 5,353 firms that were invited to take part in one or more surveys, 2,779 completed them. Similarly, for CBI, the online survey was sent to all firms that had received support from CBI in the previous three years. The number of firms that responded to this online survey was 318 in 2014; 369 in 2016; and 348 in 2017. Overall, the response rates – between 30 per cent and 52 per cent – were considered good for online surveys and were largely thanks to the intensive follow-up by PUM and CBI staff.

The qualitative case studies reflected the diversity of the sectors and economic conditions in which PUM and CBI operate. They made use of interviews with beneficiaries, experts, and other key stakeholders in the different sectors and countries. In doing so, the case studies helped to illustrate the programme's effects in terms of types of business knowledge and practices that changed, using insights from the survey. More importantly, it helped to identify enablers and barriers of effectiveness of the support modalities and how the support related to other sector-level innovation processes in the country.

After the first survey, there was a need to adjust the methodology. There was an idea to compare the performance during the programme with the trends three years before the support had started; however, this proved over-optimistic. It became clear

that PUM and CBI did not manage to generate these 'three-years-before' data on the key performance indicators (ultimate outcomes), and that the survey could not fill the gap due to fatigue and recall bias in these estimates.

Moreover, in the inception phase, we assumed that the cohorts of firms that were selected by PUM and CBI would be 'on average' similar. For PUM this assumption proved plausible (van Rijn *et al*. 2018a). However, for CBI it became clear that the number and type of supported firms depended on the sector and countries that were prioritised in each four-year period, which made the inter-cohort comparisons of indicators unreliable (van Rijn *et al*. 2018b). The solution found was to use a pooled regression with the time of participation of a firm as a covariate. This indicator captures the effect of the time after the firm has had the first contact with the support programme.

Additional to the online survey, CBI decided to collect some key performance metrics through support staff in-country. The so-called 'certified results' efforts managed to collect rich data about exports of the firms during a session where CBI staff looked at the financial reports together with the firm owner; the export data were presented in a reporting format signed off by both. The decision to conduct this resource-intensive data collection was not coordinated with the PRIME researchers who had argued in the design phase that this type of data could be too far out of the sphere of direct influence of CBI, and that capturing data on changes in business practices was more important from a resource-efficiency standpoint. In the end, however, the certified results data on exports were highly valuable for the final impact evaluation.

2.4 Phase 4: transitioning

After the funding for PRIME ended, in 2018, the collaboration between WUR and PUM continued, though at a lower intensity and within a different scope. The core objective was to continue to provide independent impact monitoring based on online surveys with PUM's M&E system. The case studies were discontinued as PUM did not receive sufficient insights on impact from these resource-intensive studies.

WUR continued with their support to extract insights from PUM's data by applying econometric techniques. Data collection with the online survey and quality assurance of these data shifted entirely to PUM. This resulted in a yearly expanding and increasingly rich time-series about impact. The increase in data points available enabled the research team to better estimate the lagged effects of the business coaching on SME performance. The key deliverable was to provide PUM with the externally validated ('certified') outcome and impact estimate in PUM's Annual Report, in some way similar to the accountant statements in the financial report.

CBI decided not to continue with PRIME. Reasons were related to the investments in terms of time and money, compared with the perceived benefits of this additional data collection to the existing efforts and the certified results exercises. However, during a 2021 interview with CBI, one interviewee did indicate that, as an indirect effect of PRIME, the organisation became more aware of the importance of good technical systems with good indicators, and that this had led to significant investment to improve these systems. Although CBI did not use the tools developed in PRIME, they strengthened their data systems in response to the experience with PRIME.

Moreover, the perceived benefit of the PRIME programme – namely, assessing the causal effect on exports – became less evident after 2018, as the accountability requirements for PSD had changed. The new guidelines for results reporting indicate that net effects do not always need to be shown (DGIS-RVO 2017); when a PSD programme could show a significant contribution to a complex change process, it was allowed to report the total change generated without estimating the attributable part of this total change. CBI, with activities in a specific sector and with firms, sector organisations, and governments, could show their impact more easily than PUM, and the 'certified results' became sufficient to report their contribution.

3 User feedback

To get a clearer picture of the experiences with the real-time monitoring and evaluation in the PRIME partnership, various in-depth interviews were held with researchers and policymakers involved in the programme. In total, 19 interviews were held in September 2017, with three follow-up interviews in March and April 2021. Moreover, in September 2017, a joint workshop was held with 15 participants who had been involved in phases three and four, to reflect on the use of the information for management decisions.

The interviews indicated that the PRIME partnership had helped both CBI and PUM to increase their accountability to donors by creating trust in their strengthened M&E systems. The partnership was positively appreciated by the Ministry of Foreign Affairs, which helped to secure continued donor funding. A staff member of PUM said: 'We have opened our organisation for a bunch of scientists. It was a bit of a gamble for us, you don't know in advance what comes out of it.'[5] Respondents indicated that support from researchers was and remains essential in designing new questions and analysing the data. Both CBI and PUM indicated that shifting data collection to the implementing organisation, without external researcher involvement, might also negatively impact on the credibility of the results.

The workshop and interviews also gave insight into user feedback in relation to the use of the real-time evaluation approach for

learning: (1) helping implementing organisations to become more impact-oriented; (2) using results from the real-time evaluation in management decisions; and (3) balancing associated workload for implementing staff with the learning from the real-time data.

3.1 Making organisations more impact-oriented

According to respondents, the real-time evaluation approach of the PRIME partnership has made their respective organisations more impact-oriented: 'PRIME has brought more focus on regular impact monitoring. The focus is now more on the quality of the mission, rather than the number of missions.'[6]

The interaction with researchers helped both organisations to sharpen their intervention logic and better define the different impact pathways of their organisations and associated outcome areas. PRIME helped them to broaden their perspective on measuring results and to go beyond the traditional focus on monitoring outputs by including more, and more appropriate, indicators at the intermediate outcome level: 'Now there is much more focus on intermediate outcomes. In each project we now report on intermediate outcomes.'[7]

The academic perspective of the researchers helped the staff of both organisations to think more critically about what to measure and how to organise data collection and data management in a way that assists them in producing portfolio-level reporting. As one respondent of an implementing organisation put it: 'Due to PRIME, we have invested much more in ICT [information and communication technology] and the role of our organisation in doing data collection and analysis.'[8]

3.2 Use of impact data in management decisions

It is clear from the interviews that findings from PRIME were used to shape discussions on the future direction of PUM and CBI activities. A respondent from CBI indicated:

> *Many of the conclusions of the PRIME study were integrated in our latest five-year strategy. In line with the conclusions of PRIME, the strategy proposed to move beyond merely European markets, adopt more digital ways of working, pay attention to gender and youth and identify larger companies to work with.*[9]

Another example comes from PUM. In 2016, the external evaluation of PUM activities (van der Windt *et al*. 2016) used PRIME data on changes in business practices to suggest that PUM should shift its portfolio more to micro and small businesses. However, PUM used the PRIME data on business performance to argue that while impact on the business practices of micro and small companies was indeed higher, in terms of employment and turnover, the effect of PUM was higher on larger companies. The discussion shows that the PRIME data were useful for

strategic decision-making and enabled a better reflection on effectiveness.

However, as areas for improvement, CBI and PUM indicated that the communication of results by the team of researchers was often too technical and complex, and the results did not relate directly enough to the day-to-day practices of the organisations to influence more operational decision-making. It was suggested that more visually attractive and more simply written research products in an early phase of implementation could have been helpful. As one of the implementing organisations put it in the evaluation workshop in 2017: '[In our organisation] you need to present your material on a serving platter, in an attractive and accessible way for the results to be used.'[10] Based on this feedback, the research team in 2018 dedicated more attention to the visual layout and readability of the final reports, and included a separate chapter with recommendations to the management in the subsequent reports.

Moreover, several respondents felt that the PRIME partnership could have embedded the researchers in the offices of CBI and PUM. More personal interaction could have increased the degree to which PRIME data and analysis were used by people in both organisations. 'For a partnership to work, you need to see each other regularly. You need regular discussions with programme managers for it to come to life.'[11] At the same time, one respondent stressed that true **ownership is also required**: 'People will not take it seriously and/or use PRIME. This is especially true now because PRIME was "invented" by people that are not working in the organisation any more'.[12]

Another interviewee from PUM said that these moments of interaction were also important for them as M&E officers to help them to become more visible in their own organisation: 'I need PRIME to connect with the rest of the organisation for it to receive support. This means connecting to people from knowledge management, management accounting and business development.'[13] Other respondents also indicated that more frequent sharing of results on both sides would have improved the level of engagement and learning in the partnership. One suggestion was taken up in phase four with PUM, where at the start of each year a meeting was organised to identify certain strategic themes for which the online survey could be used to collect additional information. As a result, additional topics such as gender, food security, and indirect effects were integrated into the last versions of the online survey.

3.3 Reasonable balance between workload and learning

The above two points sketch out the benefits of PRIME. These benefits need to outweigh the costs, especially when the learning, and not the accountability to the donor, becomes the main goal. Both the online survey and the case studies needed support

from staff. Several respondents from CBI and PUM indicated that the implementation of the PRIME research activities was too heavy a burden for some of their colleagues and partners in the field. There were suggestions to reduce the data collection to the online survey only, eliminating the qualitative part of the real-time evaluation. In contrast, other respondents found the qualitative case studies to be the most relevant for their work, while the results of the quantitative survey were felt to be more challenging to interpret and translate into action. Especially in CBI, a feeling emerged that PRIME was complicated and time-consuming, and had insufficient value for managing the different sector and country programmes; this was one of the reasons why CBI did not engage in phase four of the PRIME partnership. Furthermore, it was harder for CBI to translate the research outputs to the day-to-day work practice and decision-making processes. The support is so diverse that it is difficult to learn from average overall trends. Disaggregation of the econometric results was limited due to the relatively small number of firms involved in each sector.

4 Conclusions

This article presents insights from a theory-based impact evaluation of business coaching and export promotion that navigated the needs of the stakeholders in PSD programmes for learning and accountability, and that provided real-time information for adaptive management.

PRIME succeeded in its aim of improving the reporting of the PSD programmes' contribution to higher-level impact areas (export, employment), and of quantifying the importance of this contribution in a credible way, as demanded by the donors at the time. The key elements that made it convincing were the clear charts with appropriate indicators, the yearly time-series, and the sophisticated econometric analysis by the researchers. This contributed to a more convincing programme evaluation and a more informative report for donors.

Aside from this accountability aim, the ambition of PRIME was to improve effectiveness of the programmes by supporting monitoring, evaluation, and learning processes on an ongoing basis and feeding these with regular (real-time) insights during implementation. While the PSD organisations and policymakers benefit from the theoretical perspective and rigour of the theory-based approach, followed by a theory-based evaluation, the data collection on an ongoing basis has a cost in terms of financial and human resources. The accountability requirement to report net effects at impact level provided a clear incentive to invest in more rigorous survey-based evaluation approaches. When the funder's accountability requirements shifted towards a more qualitative approach to assess contribution and additionality, this incentive became less apparent for CBI.

When learning is concentrated at portfolio level, as was the case in PUM, the data collected in the online survey on firm practices and performance need to be aggregable and, therefore, more general. For CBI, the sector managers required more granular data than PRIME provided, and this explains why they considered the associated workload too high. Making sector-focused versions of a survey could have been a golden midway for CBI as well as PUM, including some questions and indicators that are applicable to a certain sector or country only, which could increase its relevance for the staff and experts involved in the support. The online survey modules with perception questions and contribution scores proved a flexible tool for creating sector-specific versions (see Ton, Taylor and Koleros, this *IDS Bulletin*).

Another important lesson learned in PRIME is that the monitoring and evaluation should connect to the PSD implementers' everyday reality – both their work processes and their information needs. For this purpose, it is critical to regularly involve not only the M&E staff, but also mid-management, such as country or sector managers, and communication staff. More frequent encounters or workshops can assist evaluators, researchers, and policymakers to engage more in joint sense-making of the evaluation results. Regularly 'harvesting' the data needs and key questions that PSD programmes are facing may help to ensure a better match between the research analysis and the reality of the organisations, and may improve the ownership of the evaluation.

The common ground of these two strategies – flexibility in data collection and increased interaction with mid-management – is the search to improve the usefulness and timeliness of theory-based evaluations and to find an appropriate balance between accountability and learning. PRIME helped to navigate this search, took steps in the right direction, but also showed that the road to 'real' real-time monitoring and evaluation is still a long one.

Notes

* We would like to acknowledge the contribution to the design and implementation of the PRIME programme of all the stakeholders involved. Also, we would like to thank Martijn Ramaekers and Andy Wehkamp for their feedback on a draft version of this article.

1 Fédes van Rijn, Senior Researcher, Wageningen University and Research, Netherlands.
2 Haki Pamuk, Senior Researcher, Wageningen University and Research, Netherlands.
3 Just Dengerink, Independent consultant, Food Systems, Netherlands.
4 Giel Ton, Research Fellow, Institute of Development Studies, University of Sussex, UK.
5 PUM staff, interview, 18 September 2017.
6 PUM staff, evaluation workshop, 26 September 2017.
7 CBI staff, interview, 31 March 2021.

8 PUM staff, interview, 26 September 2017.
9 CBI staff, interview, 31 March 2021.
10 PUM staff, interview, 26 March 2017.
11 WUR staff, interview, 12 March 2017.
12 PUM staff, interview, 29 March 2021.
13 PUM staff, interview, 18 September 2017.

References

Apgar, M.; Hernandez, K. and Ton, G. (2020) 'Contribution Analysis for Adaptive Management', *Briefing Note*, London: Overseas Development Institute

Bates, M.A. and Glennerster, R. (2017) 'The Generalizability Puzzle', *Stanford Social Innovation Review* 15.3: 50–54

Blattman, C. (2008) 'Impact Evaluation 2.0', presentation to the Department for International Development (DFID), London, UK, 14 February

Chen, H-T. (1994) *Theory-Driven Evaluations*, London: SAGE

Davey, C. *et al*. (2018) *Designing Evaluations to Provide Evidence to Inform Action in New Settings*, CEDIL Inception Paper 2, London: Centre of Excellence for Development Impact and Learning

DCED (2017) *The DCED Standard for Measuring Achievements in Private Sector Development: Control Points and Compliance Criteria*, London: Donor Committee for Enterprise Development

DGIS (2011) *Protocol Resultaatsbereiking en Evalueerbaarheid in PSD*, The Hague:

DGIS–Directie Duurzame Economie (DDE)–Dutch Evaluation Office (IOB)

DGIS-RVO (2017) *15 Methodological Notes – Instructions for Calculation, Validation and Reporting of Performance Indicators*, The Hague: Dutch Ministry of Foreign Affairs

Giordano, N. (2017) *Monitoring, Evaluation and Learning: Adaptive Management to Achieve Impact Results*, Care International blog, 18 January

Grimm, M. and Paffhausen, A.L. (2014) *Interventions for Employment Creation in Micro, Small and Medium-Sized Enterprises in Low- and Middle-Income Countries: A Systematic Review*, Frankfurt: KfW Bankengruppe

Harms, J.; Ton, G. and Maas, K. (2014) 'Literature Review: The Impact of Advisory Services and Export Promotion on SME Performance', *PRIME Policy Brief 2,* Wageningen: Wageningen University and Research

Khandker, S.R.; Koolwal, G.B. and Samad, H.A. (2010) *Handbook on Impact Evaluation: Quantitative Methods and Practices*, Washington DC: World Bank

Mayne, J. (2001) 'Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly', *Canadian Journal of Program Evaluation* 16.1: 1–24

Piza, C. *et al.* (2016) 'The Impact of Business Support Services for Small and Medium Enterprises on Firm Performance in Low- and Middle-Income Countries: A Systematic Review', *Campbell Systematic Reviews* 12.1: 1–167

Ruyter de Wildt, M. de; Maas, K.; Ton, G. and Harms, J. (2013) *Pioneering Real-Time Impact Monitoring and Evaluation in Small and Medium Enterprises (PRIME) – Final Report Phase I: Programme Design*, Wageningen: Wageningen University and Research

Schulpen, L. and Gibbon, P. (2002) 'Private Sector Development: Policies, Practices and Problems', *World Development* 30.1: 1–15

Ton, G. (2012) 'The Mixing of Methods: A Three-Step Process for Improving Rigour in Impact Evaluations', *Evaluation* 18.1: 5–25

Ton, G.; Vellema, S. and Ge, L. (2014) '**The Triviality of Measuring Ultimate Outcomes: Acknowledging the Span of Direct Influence**', *IDS Bulletin* 45.6: 37–48 (accessed 18 October 2021)

van der Windt, N. *et al.* (2016) *Evaluation of PUM Netherlands Senior Experts 2012–2015: An Independent Evaluation Study Commissioned by the Netherlands Ministry of Foreign Affairs*, Rotterdam: Erasmus University Rotterdam

van Rijn, F. *et al.* (2018a) *Verification of PUM's Intervention Logic: Insights from the PRIME Toolbox*, The Hague: Wageningen Economic Research

van Rijn, F. *et al.* (2018b) *Verification of CBI's Intervention Logic: Insights from the PRIME Toolbox*, The Hague: Wageningen Economic Research

White, H. (2009) 'Theory-Based Impact Evaluation: Principles and Practice', *Journal of Development Effectiveness* 1.3: 271–84

This page is intentionally left blank

Monitoring Systemic Change in Inclusive Agribusiness*

Sietze Vellema,[1] Greetje Schouten[2] and Marijn Faling[3]

Abstract Evaluations of private sector development programmes look at what changed to the workings of the system, and whether these changes are scalable, resilient, and sustainable. We present an evaluation lens that primarily qualifies changes to the systemic nature of food provisioning in markets. It converts theoretical frameworks into 'antennae' receptive to early signs of systemic effects of inclusive agribusiness that fosters food and nutrition security. The tools for this theory-informed approach were developed and applied in 2SCALE, a Dutch-funded programme aiming to incubate inclusive agribusiness and contribute to food and nutrition security goals in Africa. The article reflects on what to monitor to detect early signs of systemic effects and how monitoring can be embedded in unfolding business and partnering processes. It concludes that taking a theory-informed approach gives directionality to strategising and planning, and enhances capacities of partners in inclusive business projects to lead actions towards realising systemic effects.

Keywords inclusive development, partnerships, food and nutrition security, Africa, evaluation.

1 Introduction

Involving the private sector in achieving the public goals of food and nutrition security has led to a variety of market-led programmes that aim for inclusive agribusiness. Cross-sector partnering has been given increased prominence as a pathway to achieve the Sustainable Development Goals (Stibbe, Reid and Gilbert 2019). Sustainable Development Goal (SDG) 17, framed by the United Nations, reinforces this instrumental notion of partnerships as a key vehicle for achieving the goals overall.

However, the capacity of partnerships to contribute to these sustainable development outcomes is far from self-evident (Vellema, Schouten and van Tulder 2020). Partnering processes

The *IDS Bulletin* is published by Institute of Development Studies, Library Road, Brighton BN1 9RE, UK
This article is part of *IDS Bulletin* Vol. 53 No. 1 February 2022 **'Theory-Based Evaluation of Inclusive Business Programmes'**; the Introduction is also recommended reading.

navigate the unfolding interactions among diverse actors in a joint endeavour towards development goals and simultaneously aim to realise impacts in changing complex systems that are rather unpredictable (van Tulder and Keen 2018). Accordingly, the process of monitoring and evaluation (M&E) should be able to move along with the evolving dynamics of interventions and strategies induced by partnerships, and offer information and reflection or sense-making moments that support adaptive management.

In addition, the use of public resources also demands that partnerships are accountable for their claimed impacts on food and nutrition security, which comprise contributions to systemic changes (Global Panel on Agriculture and Food Systems for Nutrition 2016; Haddad *et al.* 2016; Posthumus *et al.* 2018a, 2018b).

The dynamics of partnering processes in combination with the systemic nature of transformative processes provides challenges for M&E. Therefore, the scope of M&E in partnerships working on the combined goals of inclusive agribusiness and food and nutrition must capture the systemic nature of the transformation processes. At the same time, it must consider the boundaries of the span of influence of the concrete activities implemented by partners working towards inclusive agribusiness in complex and rapidly changing market environments.

In response, literature has increasingly focused on assessing the assumed systemic or system-level change (Ramirez *et al.* 2018; Dentoni, Pinkse and Lubberink 2021). The approach to monitoring inclusive agribusiness presented in this article complements M&E approaches that look for systemic change through replication or responses by actors outside the sphere of the intervention programme, such as the Adopt-Adapt-Expand-Respond (AAER) approach (Nippard, Hitchins and Elliott 2014; Taylor and Lomax, this *IDS Bulletin*). This line of inquiry is central to the Donor Committee for Enterprise Development (DCED) standard – a framework for enhancing the quality of M&E of private sector development programmes – that defines systemic change as 'a modification to how a system works, and what happens as a result' (Kessler 2021: 5).

The DCED standard uses three criteria to identify whether a change is indeed systemic: **sustainability** – i.e. it should be able to continue without input from the project under evaluation; **scalability** – i.e. it should be capable of benefiting increasing numbers of people over time; and **resilience** – i.e. it should be able to adapt to changing conditions (*ibid.*). For example, the pragmatic approach to assessing system change outlined by Posthumus *et al.* (2020) combines an 'intervention lens' with a 'helicopter lens'. The 'intervention lens' is meant to assess the scale, sustainability, and impact of the changes introduced by programme interventions, while the 'helicopter lens' is designed to assess responses or changes in the broader system.

From a donor perspective, effects on the broader sector or system are important because, as clarified by Ton (2021), direct support to specific business-driven programmes is only considered legitimate when it addresses constraints in a way that benefits other (competing) firms. This article complements these pragmatic approaches to assessing systems changes by presenting a theory-informed approach, which aims to qualify systemic changes associated with inclusive agribusiness and to use early signs of systemic effects in their deliberations. It aims to enhance M&E practice, both in terms of what to monitor in order to detect early signs of systemic effects and how monitoring can be embedded in business-led partnering processes.

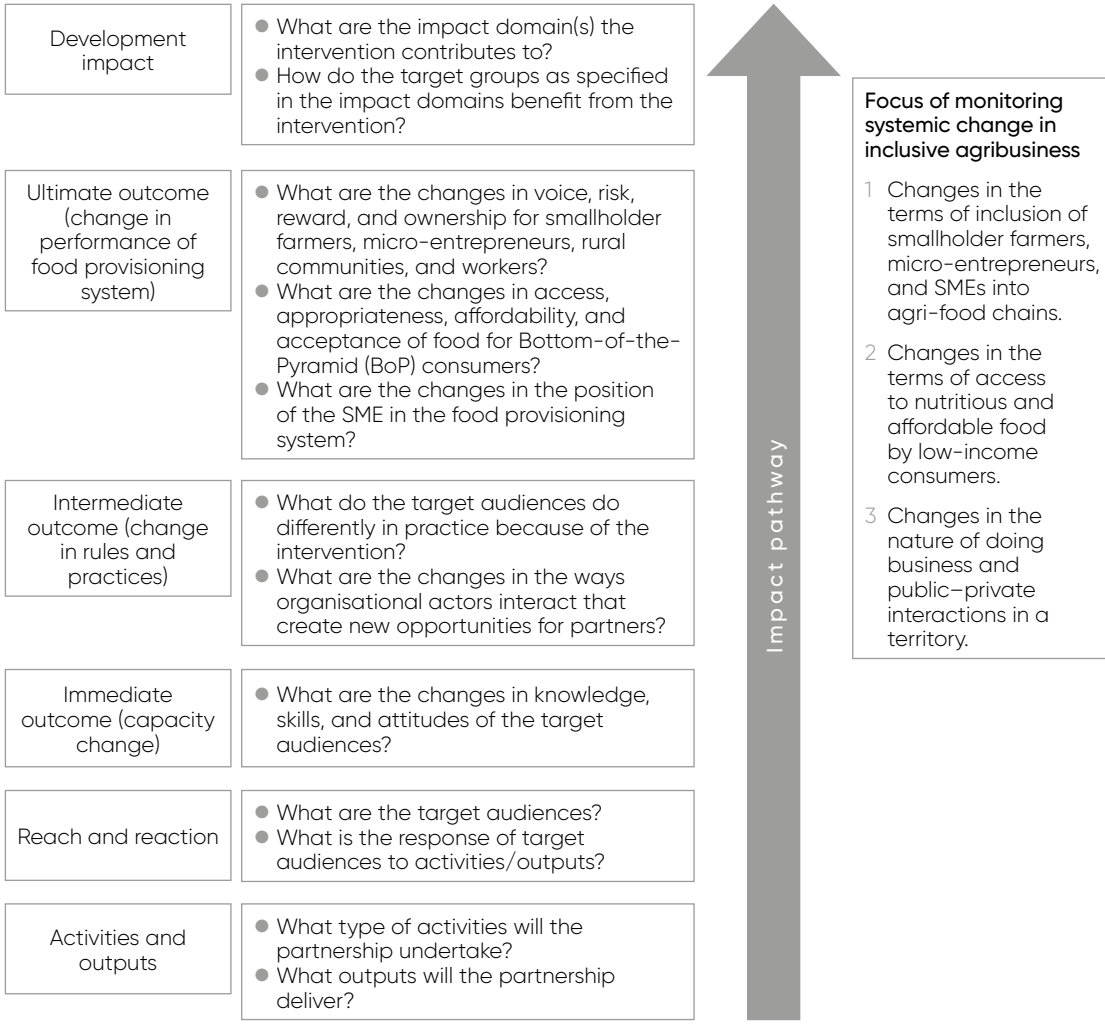
To that end, we use experiences with the design and implementation of M&E in the 2SCALE programme,[4] which supports partnerships that aim to incubate inclusive agribusiness that fosters food and nutrition security in Africa. The 2SCALE programme works in partnership with so-called 'business champions', i.e. companies open to fitting an inclusiveness agenda into their commercial strategy as players in their markets. The 2SCALE programme has a strong focus on small and medium-sized enterprises (SMEs), and its partnership facilitators support collaborating businesses. Besides agribusinesses, partnerships supported by 2SCALE can involve producers' organisations, governments, and a variety of technical and financial service providers. The M&E system in 2SCALE combines several theoretical frameworks to develop 'antennae' that are able to detect early signals of systemic change resulting from the actions and changes in practices of inclusive agribusiness which foster food and nutrition security. The use of these antennae opens space to make the search for systemic change an integral part of reflection and strategising processes of 2SCALE's partners.

The article first presents, in Section 2, the generic format used for describing impact pathways developed and implemented in the 2SCALE programme, which captures systemic changes as part of the ultimate outcomes. Section 3 describes how this theory-informed understanding of systemic change is embedded in processes of facilitated learning and reflection, and becomes part of adaptive management. Then, Section 4 presents examples from 2SCALE that clarify how a theory-informed approach to monitoring systemic change, as an integral part of learning and reflection processes, guides adaptive management. The article ends with a discussion (Section 5) on how this approach to monitoring systemic change fits the dynamics of intervention programmes that strive for inclusive business models, which contribute to food and nutrition security while navigating complex market environments.

2 Capturing systemic change

Theory-informed antennae that are sensitive to signals of systemic change are integrated into the design of the M&E system and concentrate on the tandem of food provisioning

Figure 1 Generic format for impact pathways used in 2SCALE



Source Authors' own, adapted from Partnerships Resource Centre/2SCALE (2021).

and the nature of doing business. Food provisioning is the core function of the system in which the business partners operate. Companies sourcing, processing, and/or distributing food are central to this system in that they make the connection between the production and consumption of food (Reardon 2015; Liverpool-Tasie *et al*. 2020). The companies involved operate in the context of an area-specific history of competition and coordination when doing business, and of interactions between business and state (Whitley 1999; Helmsing and Vellema 2011). It is assumed that the actions of the business partners are directed towards inclusive development, and that linking smallholder producers and poor consumers to these agribusiness companies can be considered as systemic (Birney 2021): the actions refashion

rules and practices ingrained in doing business, and construct or modify social relations in food provisioning. Accordingly, the scope of monitoring systemic change is defined by a set of mutually constituting practices that are functional to food provisioning and embedded in a spatially bounded business system of interacting private and public sector actors.

To discover the systemic effects, the M&E system in 2SCALE integrates evaluative thinking based on contribution analysis and Action Research. The M&E system uses a generic format for impact pathways (IP) that includes the distinction between types of outcomes made in contribution analysis (Ton *et al.* 2019; Ton 2021). Central to the IP-format is a sequence of immediate, intermediate, and ultimate outcomes (see Figure 1). The IP-format envisions a change process that goes from changes in the capacities and skills of target audiences (immediate outcomes), which subsequently are supposed to lead to changes in practices, rules, and interactions between partners and target audiences (intermediate outcomes). The IP-format logically links these changes. The immediate and intermediate outcomes give direction to finding traceable changes in the system of food provisioning and the nature of doing business. These systemic changes are operationalised as intended ultimate outcomes. The IPs explicate how the partners envision contributing to the ultimate outcomes, and how this process can be monitored.

Considering a defined set of ultimate outcomes as signs of systemic change requires qualifying the systemic effects of inclusive agribusiness. The M&E system builds upon an institutional perspective on inclusive development in the context of food provisioning; it combines multiple theory-informed frameworks to further qualify systemic changes and capture these as ultimate outcomes of inclusive agribusiness in food provisioning (Schouten and Vellema 2019; Vellema *et al.* 2020). In contrast to the immediate and intermediate outcomes in the IP that are defined by the partners, the Action Research team delineated a specific set of ultimate outcomes linked to scholarly literature. This offers three categories of systemic change to focus on.

First, the M&E system aims to capture changes in the terms on which smallholder farmers or micro-entrepreneurs are involved in commercial activities: the 'terms of inclusion' of smallholder farmers and other economic actors (Vermeulen and Cotula 2010; Thorpe 2018; Chamberlain and Anseeuw 2019). Second, it looks for changes in the conditions under which low-income consumers access their daily food: the so-called 'terms of access' (Thorpe and Reed 2016; Lashitew, Bals and van Tulder 2020; London 2020). These two categories are key to inclusive development and represent systemic change in food provisioning.

In addition, a third key assumption underlying 2SCALE is that realising inclusive development in food provisioning requires

inclusive agribusinesses to take the lead in doing 'business as unusual' (2SCALE 2019). The 'unusual' business practices are expected to lead and direct the transformation of the nature of doing business, which partly determines whether changes in the terms of inclusion and terms of access materialise. The M&E system considers SME leadership in 'business as unusual' as a third category of systemic change. It focuses on whether the partnership is able to change the nature of doing business in a commercially viable way, and whether their endeavours to enhance inclusiveness of agribusiness attract reinforcing responses by other private and public actors in the business environment.

The Action Research component of the M&E system is tasked to search with partners for signs of systemic effects, and the team engages in and supports data collection to track the consequences of the partnering processes. Realising that capturing systemic change is not an easy task, the M&E/Action Research team takes the lead. The M&E/Action Research is composed of one member based in each country, who interacts closely with the 2SCALE partnerships facilitators and supports data collection and processing. It also involves three part-time Action Researchers based in the Netherlands, who connect M&E to scientific knowledge on inclusive development and partnerships, and design and revise the systematics used for the M&E tools. A prime task of the team is to conduct interviews with partners and key stakeholders, translate the conceptualisation of the three categories of systemic change into palatable questions, and proactively look for possible signs of systemic change in the situated actions of the partnerships. Continuous interactions of the team with the partnership facilitators enables an intentional focus on inclusive agribusiness.

To monitor systemic change, the three categories of ultimate outcomes as introduced above have been further operationalised into a protocol (see Table 1). First, the Action Research looks for changes in the terms of inclusion of smallholder farmers (SHFs) and micro-entrepreneurs. The 'terms of inclusion' are specified in four dimensions: ownership, voice, risk, and reward (Vermeulen and Cotula 2010; Chamberlain and Anseeuw 2019). This operationalisation enables a nuanced understanding of the actual conditions under which SHFs and micro-entrepreneurs are included in business practices. It goes beyond measuring prices and income effects and tries to detect the institutional and procedural features of inclusion, by looking at SHFs and micro-entrepreneurs' voice in decision-making procedures, and the way risks and rewards are divided among (business) partners (Thorpe 2018).

Second, the 'terms of access' are further detailed into four dimensions: affordability, acceptability, availability, and appropriateness (Thorpe and Reed 2016). To address

Table 1 Operationalisation of ultimate outcomes for monitoring systemic change in 2SCALE impact pathway

Terms of inclusion of SHFs and micro-entrepreneurs	**Ownership:** the division of assets such as land and processing facilities between SHFs and/or micro-entrepreneurs on the one hand and the company/lead firm on the other hand.
	Voice: the ability of SHFs and/or micro-entrepreneurs to influence key business decisions. This includes their weight in decision-making processes, arrangements for review and grievance, and mechanisms for dealing with asymmetries in information access.
	Risk: the division of risks between SHFs and/or micro-entrepreneurs on the one hand and the company/lead firm on the other. These risks derive from uncertainties in production, changes in demand of consumers and supply of producers, and wider political and reputational risks.
	Reward: the division of economic costs and benefits between SHFs and/or micro-entrepreneurs on the one hand and the company/lead firm on the other. This includes price-setting and finance arrangements.
Terms of access for Bottom-of-the-Pyramid (BoP) consumers	**Affordability:** the alignment between the cost (and the associated price) of a product, and the consumer's willingness and ability to pay for the product. This dimension is thus determined by the household's cash flow on the one hand, and the cost of developing, producing, marketing, and distributing the food on the other hand.
	Acceptability: the alignment between the characteristics of the food offered and the daily diets in the social and cultural context of the consumer. This dimension is thus determined by the choice and design of a product offered on the one hand. On the other hand, it is determined by the consumer's customs, taste, and habits. It is determined by the consumer's ideas and convictions about food, by preparation time and other preparation requirements, and by awareness of the product and its benefits.
	Availability: the alignment between the location where the food is provided and the place where the intended consumer is located. This dimension is thus determined by the market or channel through which the product is sold.
	Appropriateness: the alignment between the quality and safety of the product offered and the consumer's needs and knowledge regarding quality, quantity, and frequency of consumption. This dimension is thus determined by the consistency of the nutritional quality and safety of the product on the one hand, and the (knowledge about the) quantity, frequency, and the way in which the product is prepared and consumed on the other hand.
Leadership in inclusive agribusiness	**Inclusive business model:** Leadership in inclusive agribusiness means anchoring development objectives into a commercially viable venture. This relates to the restructuring of an organisation's business model that is typically performed in order to combine commercial viability and development impact.
	Clustering: Leadership becomes visible when a company embeds its operations in a wider cluster of economic actors. This dimension refers to economies of scale and developing joint capacities to be competitive.
	Crowding in: Leadership means bringing others on board and mobilising their capacities and resources. This dimension deals with collaborative arrangements led by SMEs towards achieving joint goals that upgrade the sector, industry, or more broadly, food provisioning in an area.

Source Partnerships Resource Centre/2SCALE (2021).

undernutrition, poor households need to have access to nutritious food provided through accessible market channels. Thereto, the food product needs to be affordable; it needs to be available by being offered in market channels geographically proximate to low-income consumers; consumers need to be aware of the product and its characteristics and need to accept the food socially and culturally; and the product needs to be safe, and valuable and credible to the consumer's situation, in the sense that it is aligned to the low-income consumers' dietary needs (Wertheim-Heck, Vellema and Spaargaren 2015).

Third, the dimension of 'leadership in business as unusual' connects the above process of inclusive development to the nature of doing business. This becomes visible in sustained business operations or investments before and after the support, and assumes that business and inclusive development, while unusual in the sector, can be configured in a commercially viable way. It implies monitoring how and whether business entities (including professional cooperatives) drive and diffuse the inclusive business agenda, by means of three types of processes derived from business literature by the M&E team: promoting and spreading inclusive business practices, clustering of value chain actors functional to inclusive agribusiness, and crowding in of actors in the wider public and private networks.

Monitoring the emergence of 'business as unusual' as systemic change looks for how the entanglement of competition and collaboration affects the nature of business practices and the consequential relations of the inclusive agribusiness with other businesses or micro-entrepreneurs in the value chain (Ayakwah, Sepulveda and Lyon 2018). In addition, it directs attention to modes of clustering of economic actors, other than farmers, in the proximity of the leading inclusive agribusiness (Geldes *et al.* 2017; Gebru *et al.* 2019). Clustering of interdependent business practices is assumed to create conducive conditions for developing joint capacities, creating economies of scale, and collaboratively articulating the potential of inclusive agribusiness.

Finally, it relates to signs of crowding in (Fowler and Dunn 2014; Nippard *et al.* 2014) reflected in public and private actors adjusting their practices in reaction to the workings and emerging institutional features (Lawrence, Hardy and Phillips 2002) of inclusive agribusiness realised by the partners. These actors may reorganise, take on new roles and responsibilities, or develop their own offers, in a manner that is supportive to and may even accelerate the realisation of inclusive agribusiness.

3 Embedding M&E in partnering processes

The M&E system in 2SCALE combines a systematic and flexible use of IPs, which enables M&E to follow partnerships that are navigating diverse interests and dynamic market environments. The design of the M&E system recognises that partnering and

change processes are unique and context-specific, which implies that each partnership represents a time- and place-specific case of situated action that generates change (Vellema *et al.* 2013). Accordingly, the approach to M&E is partly based on a case-based analysis of partnership-specific IPs, which are re-specified regularly in a participatory manner. To make the M&E system fit the situated dynamics of partnerships, efforts were made to simplify the tools without compromising compliance with the DCED standard for result measurement:[5] for example, focus on only two or three IPs, use a linear but flexible format, rely strongly on self-reporting, and limit the number of indicators for which data are collected.

In 2SCALE, partnerships are supported to construct a theory of change with two or three IPs, with the intention to centre learning and reflection on strategic choices rather than start from a long list of actions (Faling, Vellema and Schouten 2020; Vellema *et al.* 2017). The aim of integrating systemic change as theory-informed ultimate outcomes in the IPs is to provide some directionality; however, the specification of what contributes to change processes leading to systemic changes in food provisioning and agribusiness is the responsibility of the partners.

In addition, the design recognises that ultimate outcomes are usually visible at the edge of the span of influence of the partnership. Hence, monitoring ultimate outcomes enables partners to specify how their collaboration contributes to systemic change and to recognise how this combines with external influences. Evidence for the partnership's contribution largely results from, and conversely informs, the continuous process of facilitated reflection and governance meetings of the partnerships supported by 2SCALE.

Partners frame and revise their strategies towards inclusive agribusiness using the IP-format (see Figure 1). Each partnership starts with a Diagnostic and Design (D&D) workshop and subsequently organises annual Reflect and Adapt (R&A) workshops. The framing and refinement of IPs result from these participatory processes. The partners and the M&E team jointly identify so-called 'Markers for Change' (M4Cs), which are qualitative and quantitative indications of progress and achievement linked to each immediate, intermediate, and ultimate outcome. The M&E team collaborates closely with the partners and the 2SCALE partnership facilitator to collect and process the evidence linked to each M4C. This informs annual R&A workshops organised with the partners, where a wider group of partners and stakeholders discusses progress made and reflects on the strategic choices underlying the IPs. Questions posed include: is the strategic orientation of the IPs still correct; is there reason to redirect or revise the IPs; is there more clarity about how the actions and partnering contribute to inclusive agribusiness for fostering food and nutrition security?

Embedding M&E in facilitated processes aims to enable partners to use the monitoring of outcomes in their reflections and deliberations. Understandably, the specification of the ultimate outcomes, which qualify the systemic changes underlying inclusive agribusiness, becomes more refined when the partnership matures. The interactions with the M&E team, conducting interviews with partners and partnership facilitators based on a protocol attuned to the approach outlined above, and the discussions during the subsequent R&A workshops, help to situate the reported systemic change. The format for an IP funnels the activities and outcomes to observable systemic changes.

4 Strategising towards systemic change: examples from 2SCALE

The theory-informed framing of ultimate outcomes (see Table 1) helped to find (early) signs of unfolding systemic changes, which informed priority-setting by partners. The qualification of systemic changes that can be associated with working on inclusive agribusiness is useful as a search device: where are partnerships moving towards and how do they get there? Using ultimate outcomes both to capture systemic change and to inform adaptive management and priority-setting brings the notion of systemic change closer to actual choice-making by partners. This section presents two examples of how the approach to M&E connects to adaptive management in partnerships. Both selected examples are geared towards the terms of inclusion at the upstream side of the agri-food chain, although our lens equally aims to capture signs of systemic change towards the consumer end of the chain. The first example describes how monitoring the terms of inclusion for smallholder farmers informs adaptive management. The second example displays how signs of systemic change enabled partners to capture and demonstrate systemic effects of partnerships that are navigating dynamic business environments.

4.1 Example 1: refashioning terms of inclusion

The first example presents a reflection process informed by monitoring the initially formulated IPs. It exemplifies how M&E supports partnerships to adapt their strategy and include other target audiences in efforts to refashion the terms of inclusion of smallholder farmers.[6]

Central to the partnership is an agro-processing company in Kenya, which decided to create a fortified food division for institutional buyers, such as schools or hospitals, with nutritious food products. Later, they expanded to low-income consumer markets. The main ingredient that the company used was imported soy. The strategy of the partnership aimed to shift to local sourcing of soybeans from smallholder farmers. Therefore, the partnership adopted a strategic focus on the terms of inclusion of smallholder farmers, as main suppliers of soybeans. The narrative below integrates the reflections among partners on the four aspects of inclusion: voice, ownership, risk, and reward.

In the first phase of the partnership, the company adopted a leading role in organising the sourcing from smallholder farmers. Most of the activities (almost 80 per cent) focused on training farmers and providing technical services and access to seed and other inputs. For these activities, the partnership spent 55 per cent of the available 2SCALE budget and almost 80 per cent of the financial contribution of the company. Access to seed was what farmer representatives labelled as the reward of the partnership. However, in the reflections, the business clarified that its buying capacity was limited. Therefore, smallholder farmers were hesitant to plant soybean and purchase quality seeds: the farmer representatives expressed that the uncertain purchase of their produce was considered a major risk, both for farmers and for the partnership.

Failure to guarantee a market and offer fair compensation to farmers would jeopardise the realisation of inclusive business objectives. Smallholder farmers owned land and produce, and the company owned the processing facility. Ownership of the means and resources underlying commercial transactions, (i.e. working capital to make the actual purchase and make payments, logistical infrastructure for aggregation, and transportation equipment of the soybeans), were less clearly embedded in the partnership. This absence was also reflected in one of the initial IPs of the partnership, which centred strongly on increasing productivity at farm level.

During the R&A workshop and subsequent governance meetings of the partnership, farmer representatives were enabled to voice their concerns about the low offtake by the company, which emerged as a source of tension in the partnership. This intensified because farmers claimed that prices offered by the company were not competitive and payments were delayed, which made buying planting materials and entering into an agreement with the company less rewarding. Seemingly, the limited buying capacity of the agro-processing company generated risks for smallholder farmers who had decided to plant soybean based on the assumed access to a reliable buyer. Participants in the R&A workshop confirmed that the company was not able to buy and in response the partnership facilitator teamed up with others to find a buyer for the soybeans. The participants identified the exclusive reliance on a single buyer as problematic and opened space to redirect their deliberation to the terms of inclusion of smallholder farmers and aggregators, and particularly the risks and unintended effects of the inclusive business model.

As a consequence of the R&A workshop, partners decided to refocus their strategy and involve aggregators that were closer to farmers, with the capacity to navigate the seasonal fluctuation of prices, demand and supply, and could create access to alternative market channels. Moreover, these aggregators were able to organise direct payments, which proved to be

more complicated for the company central to the partnership, which was located outside the rural communities. A practical consequence of the reflection by partners was to engage with aggregators in realising inclusive agribusiness; this involved payment modalities attuned to the situation in rural communities, transparent ways of coping with marketing risks, and taking into account the capacities and interests of farmers, as the primary owners of the soybeans. Consequently, the partnerships opted not only to focus on the inclusion of smallholder farmers but to enlarge the scope of the strategy to include intermediate aggregators. This also brought new challenges – namely, how to handle risks and rewards in a way that was favourable to smallholder farmers.

A revised IP reflected this change in strategic direction. The revised IP started with supporting the skill and capacity development of aggregators to engage with smallholder farmers who were facing risks as an immediate outcome: the number of aggregators involved and experiences with multiple arrangements shared in this process were chosen as indicators. As an intermediate outcome, which focuses on changes in practices, rules, and relationships in the business, the partnership aimed for establishing a network with an aggregation centre and linked mini-aggregation hubs. These hubs also offered space for building hubs to distribute seeds.

The performance of this network would become visible in the share of soybean sourced, which was selected as an indicator for the intermediate outcome. Ultimately, this was assumed to contribute to changing the terms of inclusion of smallholder farmers, who had reliable access to aggregation centres that were able to offer attractive prices for their produce and ensure access to affordable seeds and inputs. The partners agreed to track the number of aggregators selling to the company as a measure for this ultimate outcome, and to investigate the quality of their relationship. The process of reflection and revising IPs helped partners to identify centres of aggregation as the preferred target audience for refashioning the terms of inclusion of smallholder farmers.

4.2 Example 2: reconfiguring business relations

The second example reveals how the delineated ultimate outcomes guide partnerships in capturing systemic change; it zooms in on the emerging configuration of leading inclusive agribusiness, clustering other economic actors around the enterprise, and crowding in of public and private sectors in the area. By monitoring the reconfiguration of business relations and the nature of doing business, it is possible to focus on leadership of the main business partner in attracting other businesses to the unfolding process of making agribusiness inclusive.[7]

The core business of the Nigerian business partner was to supply large food and beverage companies with high quality

cassava starch. The company managed a 400-hectare nucleus farm and a processing factory with a capacity of 50 tonnes/day. At the start of the partnership with 2SCALE, it worked with 200 farmers contracted as out-growers. The company decided to expand its sourcing basis and increase the number of smallholder farmers included in its business. Initially, the interventions focused on training, input provisioning, and tractor services aimed to increase productivity and yields of newly contracted smallholder farmers.

The emerging coordination between company and farmers created a breeding ground for complementary initiatives by a network of transporters. Timely delivery of cassava roots is essential for processing companies. However, farmers working with the company complained about overcharging and low reliability of the transporters, while the transporters complained about the high costs due to an array of levies and taxes at local government level and at checkpoints. This hampered the expansion of the sourcing base, which was crucial for making inclusive agribusiness commercially viable. Interestingly, the collaboration and coordination between company and farmers appeared to be of interest for transporters.

One of the tangible steps taken by the company was to offer transporters and truck drivers a sticker showing that they worked for the company. A condition was to register as a transporter and comply with the relevant regulations. Eventually, more than 30 transporters registered with the company. This clustering of the transporters in the business of sourcing and processing cassava modified the nature of doing business, mainly due to increased transparency in the computing of transport fees. The fees shifted from payment per trip to payment per weight, and factored in distances between clusters of farmers and the company. A weighing bridge at the site of the company increased the transparency of costs for transportation for both transporters and farmers.

Moreover, the reduced transportation costs were one of the reasons that made inclusion in the business model attractive for farmers. The transporters themselves experienced reduced delays at checkpoints. Building relationships with transporters enlarged the cluster of actors who supported making inclusive agribusiness commercially viable, and enabled the company to connect to agribusiness clusters involving a growing number of smallholder farmers. The leading business in the partnership recognised the alignment with transporters as a crucial element of inclusive agribusiness, and directed their actions towards realising this.

The emerging coordinated actions in the cluster generated a process of crowding in of public and private actors, which reduced costs even more. Local government agreed to charge transporters per day and no longer per trip. The partnership supported transporters in accessing credit for repairs and maintenance, and even for buying new trucks. This set of mutually constituting

practices enabled the company to expand its network of supplying farmers from 200 to 2,000 smallholder farmers.

Capturing the process of reconfiguring business relations as systemic change is appreciative of the navigation process of a business taking the lead in realising inclusive agribusiness. Close monitoring of ultimate outcomes may impel a strategic focus of the partnership's interventions on the newly regulated nature of the local transportation network. It recognises clustering and crowding in as systemic effects of leadership of the company with an inclusive agribusiness agenda, which, in this case, changed the terms of inclusion of both smallholder farmers and transporters.

5 Discussion and conclusion

The evaluation lens and IP-format developed and used in the 2SCALE programme encouraged partners to discover, qualify, and monitor the systemic effects of working on inclusive agribusiness through the tools presented in this article. The use of theory to delineate a set of ultimate outcomes associated with inclusive agribusiness, i.e. terms of inclusion, terms of access, and leadership in business as unusual, offered partnerships 'antennae' to detect and monitor early and emerging signs of systemic change resulting from their actions. Integrating such directionality in M&E encouraged partners to alter the terms of inclusion of suppliers of food and terms of access of buyers of food.

Methodologically, we propose that combining contribution analysis with Action Research helps to make M&E, embedded in partnering processes, actionable and forward-looking. Including an Action Research component explicitly connected M&E to existing knowledge, and opened conceptual and methodological space to collect evidence on emerging systemic effects of unfolding partnering and problem-solving processes (Burns 2007, 2014). This implies a systematic approach of evaluation that helps to go beyond the traditional focus in impact evaluations to assess average effects for a defined set of target groups (Ton 2021).

The Action Research engages with unfolding partnering processes and refrains from prescribing or inducing actions; instead it is supportive of setting priorities and outlining strategies directed towards inclusive agribusiness (Greenwood and Levin 2007; Vellema 2012). The examples from 2SCALE illustrate how this may lead to re-strategising by partnerships in a practical, iterative, and reflexive way (Apgar, Hernandez and Ton 2020). Moreover, it makes M&E appreciative of the work of partners and partnership facilitators to diffuse the new practices, rules, and interactions beyond the boundaries of the lead business partner and the partnership (Lawrence *et al.* 2002; Lawrence, Suddaby and Leca 2009; Zietsma and McKnight 2009; Vellema and van Wijk 2015). Consequently, monitoring is tasked to capture systemic change as emergent from the choices made by partners navigating complex market environments.

The format for IPs including systemic changes as ultimate outcomes supported partners to use the format as a heuristic device to recognise and appreciate the contribution of their partnership to transformative processes. The examples from 2SCALE show that this way of monitoring the quality of systemic change has the potential to enable partners to adjust, refine, and focus their actions, and to re-strategise by shifting to other target audiences or setting different priorities in the light of systemic change.

Eventually, it may help to reinforce or catalyse processes that are plausibly generating systemic change. However, our experience in 2SCALE also indicates that the magic is not just in the format. It is not self-evident that partnerships allow space for reflexive and systemic-oriented deliberations. In early phases, after brokering partnerships, deliberations concentrated on immediate actions and preferred solutions, and tended to stay away from creatively navigating towards inclusive agribusiness. And, the theoretical nature of delineating signs of systemic change potentially overwhelms users, including data-oriented M&E staff, which therefore requires constant translation of theory to become relevant to practice.

Moreover, the pressure for programme management and the donor to use M&E primarily for accountability reporting in the domain of food and nutrition security made it difficult to keep the monitoring of unfolding and at times whimsical change processes on the agenda. The M&E team had to make continuous efforts to keep the monitoring process as close as possible to the core of the envisioned and context-specific change processes, which was not easy to make commensurate with evaluating the achievement of generic impact targets. Integrating the insights from monitoring early signs of systemic change in carefully facilitated sense-making moments seems to be a productive way to reiterate the programmatic direction towards inclusive agribusiness and to inform deliberations about where and how to intervene in the system.

We conclude that a theory-informed qualification of plausible systemic effects of inclusive agribusiness deepens the understanding of what generates change and contributes to the transformation of food provisioning. The flexible and theory-informed format for the description of and reflection on IPs, central to the approach to M&E presented in this article, fits processes of learning and adaptive management. Reflecting on the contributions of the partnerships' actions to the ultimate outcomes (understood through the following labels: terms of inclusion, terms of access, business leadership) encourages partners in business-driven intervention strategies to envision and appreciate how their interventions and activities were able to reshape the wider system of food provisioning.

Notes

* The authors would like to thank the 2SCALE programme for its support to this research. We acknowledge the invaluable contribution of the M&E team, their engagement in discussions, and their central role in data collection. We thank the Centre for Development Impact for their critical engagement in the design of the M&E approach. iCRA has been a key partner in composing the manuals that outline the design of D&D and R&A workshops central to learning and reflection in 2SCALE. A special thanks to the partners for their openness and willingness to engage in the Action Research and the M&E. The authors would finally like to thank the reviewers who helped to improve the article.

1 Sietze Vellema, Associate Professor, Knowledge, Technology and Innovation group, Wageningen University, the Netherlands.

2 Greetje Schouten, Senior Research Fellow at the Partnerships Resource Centre, Rotterdam School of Management, the Netherlands.

3 Marijn Faling, Partnerships Resource Centre (PrC), Erasmus University Rotterdam, the Netherlands.

4 The **2SCALE programme** is implemented by the International Fertilizer Development Center, BoP Innovation Center, and SNV Netherlands Development Organisation and is funded by the department of Inclusive Green Growth of the Directorate-General of International Cooperation of the Dutch Ministry of Foreign Affairs (DGIS/IGG). 2SCALE brokered and supports 65 business partnerships in eight countries in West and East Africa, and in four sub-sectors: staples, fresh produce, oil seeds, and animal products. The Partnerships Resource Centre, Rotterdam School of Management, is a strategic knowledge partner of the programme and leads the M&E.

5 For more information on the DCED standard, see DCED (2021).

6 Based on detailed reporting of 90+ reported activities and responses of participants (February 2020–April 2021) and data collected during an R&A workshop (26–27 January 2021), and preparatory interviews with the lead business, 2SCALE facilitator, and key informants (business coaches, company owners and managers, farmer representatives, representative county-level Ministry of Agriculture, and commercial bank representative).

7 Based on document analysis (e.g. minutes and reports on the partnership) and interviews with 2SCALE staff and the lead business partner after support by 2SCALE ended (source: Jay-Yina (2021)).

References

2SCALE (2019) *Business as Unusual: Insights from the 2SCALE Project,* Amsterdam: International Fertilizer Development Center, BoP Innovation Center, iCRA and KIT Royal Tropical Institute

Apgar, M.; Hernandez, K. and Ton, G. (2020) 'Contribution Analysis for Adaptive Management', *Briefing Note*, London: Overseas Development Institute

Ayakwah, A.; Sepulveda, L. and Lyon, F. (2018) 'Competitive or Cooperative Relationships in Clusters: A Comparative Study of Two Internationalising Agro-Processing Clusters in Ghana', *Critical Perspectives on International Business* 14.2–3: 230–51

Birney, A. (2021) 'How Do We Know Where There is Potential to Intervene and Leverage Impact in a Changing System? The Practitioners Perspective', *Sustainability Science* 16.3: 749–65

Burns, D. (2014) 'Systemic Action Research: Changing System Dynamics to Support Sustainable Change', *Action Research* 12.1: 3–18

Burns, D. (2007) *Systemic Action Research: A Strategy for Whole System Change*, Bristol: Policy Press

Chamberlain, W.O. and Anseeuw, W. (2019) 'Inclusiveness Revisited: Assessing Inclusive Businesses in South African Agriculture', *Development Southern Africa* 36.5: 600–15

DCED (2021) ***Introduction to the DCED Standard***, Donor Committee for Enterprise Development (accessed 25 October 2021)

Dentoni, D.; Pinkse, J. and Lubberink, R. (2021) 'Linking Sustainable Business Models to Socio-Ecological Resilience Through Cross-Sector Partnerships: A Complex Adaptive Systems View', *Business and Society* 60.5: 1216–52

Faling, M.; Vellema, S. and Schouten, G. (2020) ***Impact Pathways: Monitoring and Evaluating Partnerships for Sustainable Development***, Rotterdam: Partnerships Resource Centre, Erasmus University (accessed 19 October 2021)

Fowler, B. and Dunn, E. (2014) *Evaluating Systems and Systemic Change for Inclusive Market Development: Literature Review and Synthesis*, Leveraging Economic Opportunities Report 3, Washington DC: United States Agency for International Development

Gebru, K.M.; Rammelt, C.; Leung, M.; Zoomers, A. and van Westen, G. (2019) 'Inclusive Malt Barley Business and Household Food Security in Lay Gayint District of Northern Ethiopia', *Food Security* 11.4: 953–66

Geldes, C.; Heredia, J.; Felzensztein, C. and Mora, M. (2017) 'Proximity as Determinant of Business Cooperation for Technological and Non-Technological Innovations: A Study of an Agribusiness Cluster', *Journal of Business and Industrial Marketing* 32.1: 167–78

Global Panel on Agriculture and Food Systems for Nutrition (2016) *Food Systems and Diets: Facing the Challenges of the 21st Century*, London: Global Panel on Agriculture and Food Systems for Nutrition

Greenwood, D.J. and Levin, M. (2007) *Introduction to Action Research: Social Research for Social Change*, London: SAGE

Haddad, L. *et al.* (2016) 'A New Global Research Agenda for Food', *Nature* 540.7631: 30–32

Helmsing, A.H.J.(B) and Vellema, S. (2011) 'Governance, Inclusion and Embedding: Raising the Issues', in A.H.J.(B) Helmsing and S. Vellema (eds), *Value Chains, Social Inclusion and Economic Development: Contrasting Theories and Realities*, New York NY: Routledge

Jay-Yina, K. (2021) 'Crowding Processes and Alignment Capacity for Inclusive Business, The Psaltry Case', unpublished research report for PrC-2SCALE

Kessler, A. (2021) *Assessing Systemic Change: Implementation Guidelines for the DCED Standard*, Donor Committee for Enterprise Development

Lashitew, A.A.; Bals, L. and van Tulder, R. (2020) 'Inclusive Business at the Base of the Pyramid: The Role of Embeddedness for Enabling Social Innovations', *Journal of Business Ethics* 162.2: 421–48

Lawrence, T.B.; Hardy, C. and Phillips, N. (2002) 'Institutional Effects of Interorganizational Collaboration: The Emergence of Proto-Institutions', *Academy of Management Journal* 45.1: 281–90

Lawrence, T.B.; Suddaby, R. and Leca, B. (2009) 'Introduction: Theorizing and Studying Institutional Work', in T.B. Lawrence, R. Suddaby and B. Leca (eds), *Institutional Work: Actors and Agency in Institutional Studies of Organizations*, Cambridge: Cambridge University Press

Liverpool-Tasie, L.S.O. *et al.* (2020) 'A Scoping Review of Market Links Between Value Chain Actors and Small-Scale Producers in Developing Regions', *Nature Sustainability* 3.10: 799–808

London, T. (2020) *The Base of the Pyramid Promise: Building Businesses with Impact and Scale*, Stanford CA: Stanford University Press

Nippard, D.; Hitchins, R. and Elliott, D. (2014) *Adopt-Adapt-Expand-Respond: A Framework for Managing and Measuring Systemic Change Processes*, Springfield Centre Briefing Paper, Durham: Springfield Centre

Partnerships Resource Centre (PrC) / 2SCALE (2021) 'Monitoring Progress and Direction of Impact Pathways: Guide for 2SCALE', (Version 4), unpublished internal document

Posthumus, H.; Shah, R.; Miehlbradt, A. and Kessler, A. (2020) *A Pragmatic Approach to Assessing System Change*, Boekel: Hans Posthumus Consultancy, Springfield Centre, Miehlbradt Consulting, and Donor Committee for Enterprise Development

Posthumus, H.; de Steenhuijsen-Piters, B.; Dengerink, J. and Vellema, S. (2018a) *A Decision-Support Tool for the Design of Food and Nutrition Security Programming*, Amsterdam: KIT Royal Tropical Institute

Posthumus, H.; de Steenhuijsen-Piters, B.; Dengerink, J. and Vellema, S. (2018b) *Food Systems: From Concept to*

Practice and Vice Versa, Wageningen Economic Research Memorandum, Wageningen: Wageningen University and Research and KIT Royal Tropical Institute

Ramirez, M.; Bernal, P.; Clarke, I. and Hernandez, I. (2018) 'The Role of Social Networks in the Inclusion of Small-Scale Producers in Agri-Food Developing Clusters', *Food Policy* 77: 59–70

Reardon, T. (2015) 'The Hidden Middle: The Quiet Revolution in the Midstream of Agrifood Value Chains in Developing Countries', *Oxford Review of Economic Policy* 31.1: 45–63

Schouten, G. and Vellema, S. (2019) 'Partnering for Inclusive Business in Food Provisioning', *Current Opinion in Environmental Sustainability* 41: 38–42

Stibbe, D.T.; Reid, S. and Gilbert, J. (2019) *Maximising the Impact of Partnerships for the SDGs*, Oxford: The Partnering Initiative and United Nations Department of Economic and Social Affairs

Thorpe, J. (2018) 'Procedural Justice in Value Chains Through Public–Private Partnerships', *World Development* 103: 162–75

Thorpe, J. and Reed, P. (2016) ***Addressing Market Constraints to Providing Nutrient-Rich Foods: An Exploration of Market Systems Approaches***, IDS Evidence Report 172, Brighton: Institute of Development Studies (accessed 19 October 2021)

Ton, G. (2021) 'Development Policy and Impact Evaluation: Learning and Accountability in Private Sector Development', in H. Zafarullah and A.S. Huque (eds), *Handbook of Development Policy*, Cheltenham: Edward Elgar

Ton, G. *et al.* (2019) ***Contribution Analysis and Estimating the Size of Effects: Can We Reconcile the Possible with the Impossible?,*** CDI Practice Paper 20, Brighton: Institute for Development Studies (accessed 25 October 2021)

van Tulder, R. and Keen, N. (2018) 'Capturing Collaborative Challenges: Designing Complexity-Sensitive Theories of Change for Cross-Sector Partnerships', *Journal of Business Ethics* 150.2: 315–32

Vellema, S. (2012) 'Action Research and Institutional Change', in S. Nederlof and R. Pyburn (eds), *One Finger Cannot Lift a Rock: Facilitating Innovation Platforms to Trigger Institutional Change in West Africa*, Amsterdam: KIT Royal Tropical Institute

Vellema, S. and van Wijk, J. (2015) 'Partnerships Intervening in Global Food Chains: The Emergence of Co-Creation in Standard-Setting and Certification', *Journal of Cleaner Production* 107: 105–13

Vellema, S.; Schouten, G. and van Tulder, R. (2020) 'Partnering Capacities for Inclusive Development in Food Provisioning', *Development Policy Review* 38.6: 710–27

Vellema, S. *et al.* (2017) ***Do Theories of Change Enable Innovation Platforms and Partnerships to Navigate Towards Impact?***, KIT Working Paper 2017-13, Amsterdam: KIT Royal Tropical Institute (accessed 25 October 2021)

Vellema, S.; Ton, G.; de Roo, N. and van Wijk, J. (2013) 'Value Chains, Partnerships and Development: Using Case Studies to Refine Programme Theories', *Evaluation* 19.3: 304–20

Vermeulen, S. and Cotula, L. (2010) *Making the Most of Agricultural Investment: A Survey of Business Models that Provide Opportunities for Smallholders*, London/Rome/Bern: International Institute for Environment and Development/ Food and Agriculture Organization of the United Nations/ International Fund for Agricultural Development/Swiss Agency for Development and Cooperation

Wertheim-Heck, S.C.O.; Vellema, S. and Spaargaren, G. (2015) 'Food Safety and Urban Food Markets in Vietnam: The Need for Flexible and Customized Retail Modernization Policies', *Food Policy* 54: 95–106

Whitley, R. (1999) *Divergent Capitalisms: The Social Structuring and Change of Business Systems*, Oxford: Oxford University Press

Zietsma, C. and McKnight, B. (2009) 'Building the Iron Cage: Institutional Creation Work in the Context of Competing Proto-Institutions', in T.B. Lawrence, R. Suddaby and B. Leca (eds), *Institutional Work: Actors and Agency in Institutional Studies of Organizations*, Cambridge: Cambridge University Press

Assessing Contributions Collaboratively: Using Process Tracing to Capture Crowding In*

Marijn Faling[1]

Abstract If inclusive business is to realise wide and sustained development impacts, it is likely to depend on crowding in of other public and private actors. Assessing the contribution of inclusive business to crowding in is difficult because the phenomenon usually only manifests after project completion, and the complex operating environment complicates the process of evidencing a link between intervention and outcome. With donors placing increasing emphasis on demonstrating impact, innovative approaches to assess crowding in are needed. This article presents an adapted form of process tracing to assess the contribution of inclusive business to crowding in. It reports on the contribution of CREATE, an inclusive agribusiness project, to the crowding in of malting companies in Ethiopia's barley value chain. Though predominantly focusing on demonstrating a programme's contributions to crowding in, the article offers suggestions for how this process tracing-based exercise may support the fostering of inclusive agribusiness practices more broadly.

Keywords crowding in, inclusive business, impact evaluation, process tracing, value chain intervention, Ethiopia.

1 Introduction

Private sector engagement in development is gaining traction. This has encouraged the formation of inclusive agribusiness models – that is, ways of doing business that aim to improve the livelihood of smallholder farmers through integration in value chains in commercially viable ways (van Westen *et al.* 2019). Such approaches are often implemented in collaboration with other stakeholders in the value chain. The assumption is that through uniting the resources of public, private, and non-governmental stakeholders, development outcomes will exceed the outcomes that could be achieved by individual actors. Inclusive business models are assumed to enable wide-scale and sustained results

The *IDS Bulletin* is published by Institute of Development Studies, Library Road, Brighton BN1 9RE, UK
This article is part of *IDS Bulletin* Vol. 53 No. 1 February 2022 **'Theory-Based Evaluation of Inclusive Business Programmes'**; the Introduction is also recommended reading.

(OECD n.d.; Hestad 2021), which not only benefit direct partners and target audiences but also bring changes in broader market systems (Schouten and Vellema 2019; Thorpe 2014).

Entities concerned with private sector development, such as the Donor Committee for Enterprise Development, Springfield Centre, and Building Effective and Accessible Markets (BEAM) Exchange, view crowding in as central to realising systemic change. Crowding in refers to the phenomenon whereby other public and private actors in the system adjust their practices in a manner that supports the intervention's development objectives (Fowler and Dunn 2014; Nippard, Hitchins and Elliott 2014).

There are several features of crowding in that pose challenges for assessing it in evaluations. Usually when programmes close, systemic results like crowding in have only begun to materialise. This is because crowding in usually only manifests over longer time frames, beyond the temporal scope of an intervention (Kessler 2021). Validating the role of an intervention in stimulating crowding in requires the evaluator to be explicit about the link between intervention and outcome; this is in order to demonstrate that the effect is a consequence of the intervention and not something happening by chance or because of other developments (Mayne 2012). In the complex environment in which these programmes are implemented, convincingly demonstrating the contribution of a particular private sector development (PSD) programme to crowding in is difficult.

Meanwhile, donors and commissioners of impact evaluations are increasingly demanding an assessment of programmes' contributions (Befani and Stedman-Bryce 2017). Besides serving accountability purposes, these evaluations may facilitate learning about effective processes of crowding in of inclusive agribusiness. Consequently, practitioners as well as researchers are piloting approaches to assess PSD contributions to processes of systemic change (Posthumus *et al.* 2020).

This article describes a collaborative exercise with the Community Revenue Enhancement Through Agricultural Technology Extension (CREATE) partnership, a collaborative private-sector engagement project in Ethiopia, during the period 2015–20. It focused on including smallholder farmers in the malt barley supply chain for beer production and the food market. Key partners included Heineken, the European Cooperative for Rural Development (EUCORD), the International Finance Corporation (IFC), and the Dutch Ministry of Foreign Affairs (MoFA). The project's triple objectives were improving the wellbeing of 20,000 smallholder farmers, reducing reliance on imports, and contributing to food security. Its main interventions centred on local barley production and on connecting farmers to the value chain. CREATE claims to have contributed to the investments of two European malting companies that started operating malting plants in Ethiopia

early in 2021. CREATE interpreted these as furthering its inclusive agribusiness objectives. Together with the CREATE partners, the collaborative exercise set out to find and assess evidence for this claim.

The approach was based on process tracing, adjusted in several ways to make it suitable for a relatively resource-constrained collaborative evaluation around future emergent outcomes. Adjustments included assessing the probative value of emergent future events instead of past events; and basing process tracing on existing data without additional data collection.

Section 2 of this article discusses the basics of process tracing, as well as the adjustments to tailor process tracing to assess contribution claims. Section 3 demonstrates how process tracing was applied to the case of CREATE. The exercise is discussed with conclusions in Section 4.

2 A process tracing approach to assess contribution collaboratively

Process tracing is an approach of causal analysis used for in-depth (multi-)case studies (Beach and Pedersen 2019). Although it has existed as a methodology in social sciences for some time, particularly history and political science, it is increasingly used in theory-based impact evaluation (Stern *et al.* 2012; Wauters and Beach 2018). Process tracing is used to explore and test causal inferences by critically analysing the sequence of events that have unfolded. It is based upon a mechanistic understanding of causality. It is a tool to unpack and critically assess a causal process consisting of interlinked mechanisms between an independent cause C (e.g. a PSD programme) and the dependent outcome O (e.g. the impact).

Mechanisms are often unobservable. Process tracing therefore distinguishes between hypotheses about causal mechanisms, and the observable and testable manifestations of the existence of those mechanisms in reality (Beach and Pedersen 2019). We cannot get full certainty about the existence of mechanisms, therefore process tracing helps to increase or decrease our confidence in the hypotheses about reality, in light of limited available information (Befani and Stedman-Bryce 2017; Fairfield and Charman 2017). The goal of process tracing is to approach the hypotheses like a detective and to look for the 'evidence' that convincingly demonstrates that a certain mechanism has taken place (Punton and Welle 2015a).

There are various forms of process tracing, depending on the nature and aim of the exercise (Beach and Pedersen 2019). Theory-testing process tracing assesses whether a hypothesised mechanism links intervention and outcome. Theory-building process tracing starts with empirics and is concerned with finding the mechanism that links intervention and outcome. Outcome-explaining process tracing involves collecting multiple causal

mechanisms to explain a certain outcome of interest (Wauters and Beach 2018). Although all versions differ in their approach, they share some common characteristics in the way they look for and analyse pieces of evidence.

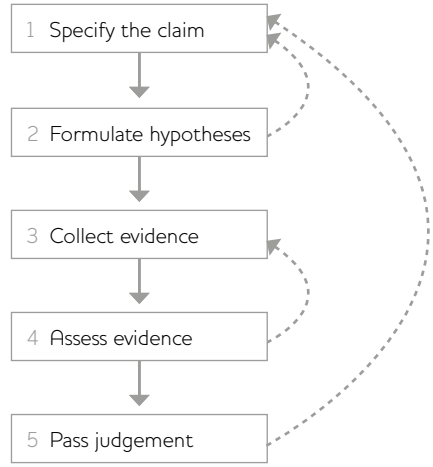
To determine whether the collected data would usefully serve as evidence, each piece of potential evidence is assessed according to the indicators of **certainty** and **uniqueness**. Certainty relates to whether we have to find the data for the hypothesis to be true, whereas uniqueness relates to whether there are alternative explanations for the presence of the piece of evidence (Beach 2017; Bennett 2015; Rohlfing 2012). The function of potential pieces of evidence for confirming or disconfirming hypotheses is determined by a combination of the certainty and uniqueness of evidence. The certainty of evidence is high when the evidence needs to be found to confirm our hypothesis. If certainty is low, evidence is not necessary to prove our hypothesis. The uniqueness of evidence is high when it is sufficient to confirm our hypothesis, whereas if the uniqueness is low, evidence leaves room for other explanations and does not prove that an intervention contributes to the impact (Beach and Pedersen 2019; Punton and Welle 2015b).

A single piece of evidence can underpin several hypotheses, while sometimes multiple data sources together form a piece of evidence. The evaluator should always question what the evidence found means, and whether it can be trusted. Imagine a farmer stating that their improved yields are the result of the support received from programme X. The reliability of this piece of evidence depends on the context and the motives of the farmer. If the farmer's statement is the result of an interview by a practitioner from programme X, it is likely that the farmer does not want to disparage the programme. In that case, the evidence reveals little about the phenomenon of interest, and so additional evidence is needed to validate the hypothesis. Combined, the indicators of certainty and uniqueness and the assessment of reliability prompt the following questions for each piece of potential evidence (Beach and Pedersen 2019):

- Can we trust the source (reliability)?
- What does the evidence tell us (what is it evidence of)?
- Is it necessary to find this evidence for the hypothesis to hold (certainty)?
- If the evidence is found, are there any alternative explanations that may still disconfirm the hypothesis (uniqueness)?

The exercise described in this article used process tracing in a customised manner. The article briefly discusses the steps involved and illustrates these in more depth in the subsequent section.[2]

Figure 1 The steps of the process tracing exercise



Source Author's own.

1 The first step has been a collaborative brainstorm session to identify, specify, and describe the factors that allegedly contributed to the outcome and to identify possible alternative explanations.

2 In the second step, this contribution claim and the proposed alternative explanations were used to develop a set of hypotheses. These first two steps roughly follow a theory-building starting point, which aims to identify and conceptualise C(ause) and O(utcome), to enable the testing of their presence (Beach and Pedersen 2019). It was necessary to rely on existing evidence collected by programme staff that could provide an indication of the likelihood that the outcome of interest would occur. The drafting of hypotheses involved several rounds of formulating and discussing with the partners to arrive at the ultimate hypotheses to be tested.

3 The third step was undertaken in a more collaborative manner. We engaged in a search for existing data to serve as potential evidence to establish confidence in the formulated hypotheses. Next, to process information available from the project, the author dug into the existing academic and grey literature in search of evidence in the form of earlier studies around similar or comparable themes.

4 In the fourth step, the author subjected the collected evidence to the four identified questions to critically assess it and determine confidence in the contribution claim. During this step, the author again consulted with the partners several times to identify and collect additional empirical fingerprints that could further strengthen the confidence in the set hypotheses.

5 In the final step of passing a judgement, the author reviewed all the evidence by confirming or disconfirming the overall claim that CREATE has contributed to crowding in. Confirming the claim means the evaluator has sufficient confidence in the contribution claim. Not confirming the claim does not necessarily mean that there was no contribution; it means that there was not sufficient evidence to confirm the contribution claim.

Figure 1 sketches the steps of the exercise.

3 Applying process tracing

This section illustrates how the steps described above were used to demonstrate how the approach works for the evaluation of the contribution claim of CREATE about crowding in of the two malting companies to the benefit of smallholder inclusion.

CREATE aimed to commercialise farming based on contracts, supplying a package of high-yielding seed varieties and other agricultural inputs, alongside cultivation techniques such as row-planting and crop rotation. The objective of CREATE was to increase productivity and income – and thus wellbeing – of smallholder farmers; providing a secured market for their produce by connecting farmers to the malt barley value chain; contributing to food security; and reducing Heineken's and the country's reliance on imports (Heineken 2018). After the implementation period of CREATE, two malting companies decided to open malting factories in Ethiopia. CREATE claims that their activities have attracted these malting companies to invest. They consider this development to be a systemic effect of their project and supportive of realising the project's inclusive agribusiness objectives.

The claim can be broken down into two separate overarching claims: (a) that the investments by Boortmalt and the Soufflet Group (Soufflet hereafter) can be causally linked to the CREATE project; and (b) that these investments support the original approach and objectives of CREATE towards inclusive agribusiness.

The following sub-sections follow the steps as described in Section 2 with discussion of each claim developed by the hypotheses. The article illustrates per hypothesis how we identified and assessed one of the pieces of evidence, and how the criteria of reliability were applied, what the evidence demonstrates, and what are the **certainty** and the **uniqueness** of the evidence. Table 1 illustrates the evaluation of all pieces of evidence.

3.1 Causally linking the investments by Boortmalt and Soufflet to the CREATE project

The first part of the claim about CREATE's role in attracting investments translates into the following hypothesis:

H1: CREATE has contributed to attracting investments of malting companies in Ethiopia's malt barley value chain.
The hypothesis formed the basis for a discussion about how the partners perceived that CREATE had contributed to the crowding in of malting companies, and what data could be used as evidence to demonstrate CREATE's contribution to this development. Three sets of potential evidence were identified in collaboration with the partners.

One of the pieces of data that partners identified is a public statement on video in which the Senior Operations Officer of the International Finance Corporation (IFC) explains that IFC has made an equity investment of US$20m in the malting company Soufflet Ethiopia, a subsidiary of the Soufflet Group. The reason for making the equity investment, as explained in the video, was partly because of the CREATE programme that through its positive results demonstrated the opportunities in terms of potential capacity of malt barley production in Ethiopia.

When assessing this piece of evidence, the first question concerns the **reliability** of the source. The video comes across as an authentic video in which we see the Senior Operations Officer of IFC explaining the reasons behind IFC's equity investment. The video has been published by IFC, and therefore it is concluded that the data source itself can be trusted.

This leads to the second question of **what the evidence tells us**. The data are an indirect piece of evidence in the sense that it demonstrates that CREATE's success in increasing the productivity of quality malting barley attracted investments that have financially supported the opening of Soufflet, one of the malting plants. Regarding certainty, we would not necessarily need to find this piece of evidence for the hypothesis to hold. IFC could have invested in Soufflet without publicly stating their rationale for doing so.

Further, technically Soufflet could have invested without an equity investment by IFC. Regarding the uniqueness of evidence, it needs to be certain that there are no plausible alternative explanations for finding this evidence that are unconnected to the contribution of CREATE. It could well be that IFC would praise CREATE, even without it being the real reason for making the equity investment. Because the evidence is neither certain nor unique, it is insufficient to confirm that CREATE motivated the malting companies to invest.

Therefore, in this exercise, the partners were brought together to discuss whether there would be additional evidence that could rule out any 'bragging' factor on the part of the IFC. The partners came up with an internal presentation by the Senior Operations Officer of IFC to the IFC board in which he presents the success of CREATE and raises the opportunity of investing in Soufflet

following CREATE's success. This piece of evidence is rather reliable, as an internal presentation would not be influenced by the potential need of keeping in mind a wider audience. It comes across as an authentic source as it contains the name of the official involved in CREATE and bears the IFC logo.

The presence of this piece of evidence makes it much less likely that the link between the equity investment and CREATE was just a promotional talk, and hence increases the uniqueness of the evidence. Consequently, combined, these pieces of evidence gave sufficient confidence that Soufflet has been attracted at least partly as a result of CREATE.

3.2 Linking the investments by the malting companies to improvements in the wellbeing of smallholder farmers

For the second part of the claim, it is necessary to assess whether the investments can be considered an indication of crowding in. Because this is an outcome that is only starting to emerge and has not come to fruition yet, it is not possible to know for sure whether crowding in will effectively occur. Instead, it is possible to test the probability that the investments can be considered as plausible indicators for the future crowding in. This also implies that these malting companies would need to support the inclusive agribusiness objective of improving the wellbeing of smallholder farmers:

H2: The investments by malting companies contribute to improving the wellbeing of smallholder farmers.

The partners identified several pieces of evidence that could potentially underpin hypothesis H2 (Table 1). The piece of evidence that could potentially strengthen this hypothesis entails data that indicate the existence and the nature of a follow-up barley value chain development project by Heineken, EUCORD, IFC, and one of the malting companies, Soufflet. The new programme, Barley Organization of Supply and Training in South East and Central Oromia (BOOST), will run from 2020 to 2023. The piece of evidence demonstrates how the project aims to enhance the productivity of farmers and the quality of their produce through access to improved seed varieties and other inputs, and by building capacity of barley producers. It aims to source 80,000 tonnes of barley annually from 55,000 mostly smallholder farmers (CREATE n.d.; Otuki 2021).

We first assessed the **reliability** of the source. It is an official project proposal, and there is public coverage of the project by several sources that are known to critically scrutinise the assumptions for investments in new development programmes. This means that the piece of evidence is considered as reliable.

What does the evidence reveal? Since it is a three-year project, the piece of evidence shows that in the coming years, Soufflet, together with other BOOST partners, will aim at improving the

economic wellbeing of smallholder farmers, through increasing farmers' productivity and product quality through access to improved seed varieties and other inputs, and technical and agronomic capacity-building measures. Their aim is to source 80 per cent from smallholder farmers. Although projects do not always manage to deliver the intended results, this is likely to do so, because of the involvement of CREATE partners and their experience and networks, which enhances the likelihood that BOOST will succeed. BOOST will not be a direct continuation of CREATE. Because the follow-up project is implemented in a different geographical location, it is unlikely that the follow-up project will claim outcomes that in fact have been produced by CREATE in the past, and not by BOOST's support activities.[3]

Regarding **certainty** of the piece of evidence, given the widely held view that cross-sector partnerships are required to advance inclusive (agri)business approaches (Schouten and Vellema 2019), it is likely that Soufflet would engage in this collaborative initiative when it wanted to work on improving the wellbeing of smallholder farmers. There are no likely scenarios in which we would not find this evidence. Furthermore, since the malting company is partly reliant on the project for its malt barley supply, there are limited incentives for the maltster to leave the partnership. This means that the **uniqueness** of the evidence is high and therefore considered sufficient on its own to confirm the hypothesis.

3.3 Alternative explanations for the investments by the malting companies

It is useful to think about possible alternative explanations for the hypotheses, especially to put CREATE's contribution in perspective. For instance, obviously CREATE has not been the only programme targeting Ethiopia's barley value chain, and other value chain initiatives may have led to rising production and productivity as well. Furthermore, the government of Ethiopia has adopted a long-term strategy to promote the development of smallholder farmers and the agricultural sector, with the malt barley value chain as one of the target areas (Lavers 2011), and it may have had activities in the area that explain the outcome. Note that these alternative hypotheses are not necessarily rival hypotheses. Confirming either of the alternative hypotheses does not necessarily lead to disconfirming the main hypotheses about CREATE's role in the process. A plausible alternative explanation about the crowding in of malting companies is therefore that other value chain initiatives have attracted Soufflet and Boortmalt to invest in Ethiopia's malt barley value chain:

H3: Other initiatives have contributed to the crowding in of malting companies.

One of the pieces of evidence that could confirm this hypothesis is a 2019 Annual Report by the Agricultural Transformation Agency (ATA 2019), an initiative of the Ethiopian government to promote agricultural sector transformation. The report states

Table 1 Assessment of evidence

#	Hypothesis	Evidence	Evidence of	Reliability	Certainty[4]	Uniqueness[5]	Evaluation[6]
H1	CREATE has contributed to attracting investments from malting companies Soufflet and Boortmalt in Ethiopia's malt barley value chain	• IFC statement regarding Soufflet • IFC internal presentation	CREATE spurred investments that co-facilitated the investments by Soufflet	High. Video comes across as authentic, publication by IFC emphasises reliability of the source.	Low	High	Partly confirms H1
		• Internal email conversation in which SECOBRA (barley breeding organisation of which Soufflet and Boortmalt are shareholders) requests to share right to the traveller barley variety introduced by CREATE with malting companies	Part of CREATE's interventions (introduction of new seed varieties) are appreciated by malting companies	High. Data contains an original email conversation. No signs this conversation was manipulated in any way.	Low	Low	Does not confirm H1
		• Quantitative data about CREATE's successes in terms of rising production and productivity • Signing MoU between Ethiopian government and malting companies	Investment plans occurred sequentially after rising productivity of CREATE, as an indication that the willingness to invest occurred after the project has demonstrated positive results	Moderate. Success of CREATE is most strongly illustrated in project documentation, which may have used calculations that could exaggerate success of the project. However, additional sources confirm the rising production and productivity.	High	Low	Does not reject H1

cont./

Table 1 Assessment of evidence (cont.)

#	Hypothesis	Evidence	Evidence of	Reliability	Certainty	Uniqueness	Evaluation
H2	The investments by malting companies Soufflet and Boortmalt will likely improve the wellbeing of smallholder farmers	● BOOST project coverage	Soufflet aims to improve smallholder integration into the value chain during 2020–23	High. Other sources confirm existence and objectives of the partnership.	High	High	Partly confirms H2
		● MSc thesis on supporting farmers in the malt barley value chain in Ethiopia	Boortmalt relies on a similar inclusive agribusiness model as Soufflet	Moderate. Authenticity of master's theses is generally rather difficult to determine.	Low	Low	Does not confirm H2
		● ATA 2019 Annual Report claims that malting companies' investments will improve the livelihoods of 10,000 farmers	Government is optimistic about the impact of malting companies on farmers' wellbeing	High. Official report published on ATA's website.	Low	Low	Does not confirm H2
H3	Other value chain initiatives have contributed to the crowding in of malting companies	● ATA report covering government's efforts around attracting malting companies	Ethiopian government has contributed to attracting investments Boortmalt and Soufflet	High. Official government documentation published on ATA website.	Moderate	High	Confirms H3
		● Project page Sourcing for Growth (S4G) partnership	Other initiatives have contributed to improving productivity and quality of malt barley	High. Other sources confirm existence and objectives of the partnership.	Low	High	Confirms H3
H4	The investments by malting companies Soufflet and Boortmalt are unlikely to realise any substantial improvements in the wellbeing of smallholder farmers	● Academic article discussing how wellbeing improvements through barley value chain integration in Ethiopia depend on economic status farmer	BOOST project likely to focus on farmers with certain economic and geographical characteristics, excluding older and more remote farmers	High. Academic articles that go through peer-review process may be expected to contain reliable information.	Low	Low	Does not confirm H4

Source Author's own.

how the efforts of the ATA have led to agreements with Soufflet and Boortmalt to establish malting plants. The website of the ATA mentions how the agreement signed with Boortmalt in 2017 grants the malting company land permits to build its factory (ATA 2017); and similar arrangements were made with Soufflet in 2018 (ATA 2018).

The **reliability** of this source of evidence is high, as the information is provided in official ATA communication channels published on their website. The evidence would demonstrate that the Ethiopian government has contributed to attracting investments by Boortmalt and Soufflet.

The **certainty** of the evidence is moderate. On the one hand, we would expect the government to report on its successes in terms of attracting foreign direct investments to the Ethiopian agricultural sector, especially given the government's priority to boost agricultural modernisation. However, on the other hand, it would also be likely that the government would report on attracting investments in more general terms, meaning that we would not find explicit coverage of government spending at the level of detail of individual organisations. Therefore the certainty of this piece of evidence can be considered as moderate.

The **uniqueness** of this piece of evidence is rather high. It is very unlikely that an official government report would report on investments made by the government if there had been none. Furthermore, given the fact that the government administers all land in Ethiopia, it is very likely that the land permits would have been issued by the Ethiopian government. This means that this piece of evidence confirms the hypothesis that in addition to CREATE, other initiatives, more particularly initiatives from the government, have also contributed to the crowding in of the malting companies.

In addition, it is useful to verify alternative hypotheses to assess what the investments by the malting companies will lead to. Because the events that we are looking for have only begun to emerge, the alternative explanations are more likely, including the probability that the outcome develops in a different direction:

H4: The investments by malting companies are unlikely to realise any substantial improvements in the wellbeing of smallholder farmers.

For this hypothesis, it was not possible to rely on existing data from the partners. However, there is a variety of academic literature that discusses malt barley value chain projects in Ethiopia. One of these articles reports how the wellbeing improvements of malt barley value chain integration in Ethiopia seem to be dependent on the socioeconomic status of farmers; more specifically, that value chain integration initiatives tend to exclude older farmers and farmers who live in remote areas (Gebru *et al.* 2019).

The reliability of this piece of evidence is evaluated as high. The article is published in a well-known journal by a trustworthy publisher and every submission goes through a double-blind peer review process. The evidence would mean that the BOOST project might equally focus on farmers with certain economic and geographical characteristics, like large commercial farmers.

As it is not necessary to find this academic article for the hypothesis to be true, the **certainty** of the evidence that there will be no inclusive agricultural development is considered as low. This kind of finding about selective involvement of farmers in contract farming arrangements is usually highly dependent on the set-up of value chain interventions and the local context in which these interventions are implemented. We may have found this piece of evidence without the hypothesis needing to be true. We therefore assess the uniqueness of the evidence to be low as well. In conclusion, this piece of evidence is insufficient to confirm H4.

By systematically assessing the collected evidence for the different hypotheses, it can be concluded that we can partly confirm the claim that CREATE has contributed to the crowding in of malting companies (see Table 1). More precisely, we can claim with confidence that CREATE has contributed, alongside other factors, to the crowding in of Soufflet and that this likely contributes to part of CREATE's objectives, including improving the economic wellbeing of smallholder farmers.

4 Discussion

This article reports on an exercise to assess the reliability of the pieces of evidence to support a claim that a programme contributed to systemic changes beyond the temporal and spatial boundaries of the programme. Based on a collaborative approach to identify and critically assess the evidence, it was possible to confirm the hypotheses about the crowding in of at least one of the malting companies. This means that following this process tracing approach, CREATE can claim with confidence that the project has contributed to the crowding in of other actors, more specifically the malting company Soufflet, into CREATE's inclusive agribusiness approach. The establishment of the subsidiary Soufflet Ethiopia is likely to contribute to improving the wellbeing of smallholder farmers. The exercise has helped to advance insight about the likelihood that the process of development as pursued by CREATE will continue beyond project termination.

Though far from a done deal, this scrutiny of the pieces of evidence helped us to become more precise and certain about the contribution of an inclusive agribusiness. The guidance of process tracing helps to approach the formulation of contribution claims and the selection and assessment of evidence in a structured way, by making use of the criteria of necessity that the piece of evidence would be present and sufficiency of the evidence for the

claim to be true. As such, the application of this approach enables the evaluator to increase the robustness and conceptual precision of contribution claims (Befani and Stedman-Bryce 2017).

The exercise demonstrates that process tracing, although frequently presented as time- and resource-intensive (Hay 2016), can also be undertaken in a simpler way. By making use of available evidence and expert judgements of practitioners to identify this evidence, the approach of process tracing becomes achievable, even with limited resources. This approach also opens the door to more robust collaborative evaluation approaches. An often-cited risk with collaborative evaluation approaches is that the evaluator becomes too engaged, leading to bias in the findings (Mapitsa and Chirau 2019; Braskamp, Brandenburg and Ory 1987). The explicit guidance offered by process tracing approaches helps to collect and assess data offered through a process tracing approach, and functions to improve independence and critical scrutiny when assessing contributions collaboratively.

Although an exercise such as this one seems to be capable of enhancing our confidence in a programme's contribution to crowding in, this is just one experience of how to use a process tracing approach in a collaborative context. Much work remains to be done. While the approach has demonstrated (a) that CREATE has contributed to the crowding in of malting companies, and (b) that one of the malting companies is likely to contribute to continuing and widening the benefits for smallholder farmers, this exercise tells us little about the precise pathways and activities through which CREATE has fostered these investments. This is an important void that needs to be addressed.

The central objective of process tracing is to unpack the causal process that links cause and outcome, by looking for evidence along the causal chain. The unpacking of the causal process between intervention and outcome can be done more granularly than has been possible in this article. Tracing the causal process is particularly important for monitoring and learning processes, as these require timely feedback on progress and direction of programmes, and an understanding of the processes through which (combinations of) strategies and processes contributed to the results (Rogers and Macfarlan 2020). More granular insights would support practitioners in developing a sensitivity to recognise crowding in, to help them strategise to reach outcomes that are beyond their direct sphere of influence.

The exercise presented in this article could serve as a first iteration and starting point to further understand the pathways through which CREATE has triggered crowding in (Taylor, Torugsa and Arundel 2018). As concrete follow-up to this research, existing theoretical knowledge about pathways towards crowding in could be used to propose new hypotheses that are empirically

testable in a subsequent round of process tracing, with the objective of better understanding the particular pathways through which crowding in can be fostered and nurtured.

Furthermore, incorporating exercises like these in follow-up programmes will help to broaden the regular focus on outputs of activities to include the more systemic outcomes of programmes in monitoring and evaluation efforts. Becoming aware of the signs of systemic changes will help practitioners to track processes outside regular result frameworks and log frames. Articulating and critically evidencing a programme's contribution claim enables practitioners and evaluators to set boundaries of what needs to be focused upon, both programmatically and in terms of monitoring. Systemic changes can as such be incorporated into programme management so that it helps practitioners to continue fostering this process, in order to nourish the continuation and widening of inclusive agribusiness practices.

Notes

* The author thanks the Community Revenue Enhancement Through Agricultural Technology Extension (CREATE) partners for their invaluable contribution to the research. The research was made possible through the grant of the Dutch Ministry of Foreign Affairs (MoFA) to the Partnerships Resource Centre. The author furthermore thanks the editors Sietze Vellema and Giel Ton for their feedback, which helped improve the focus and approach of the manuscript.

1 Marijn Faling, Partnerships Resource Centre (PrC), Erasmus University Rotterdam, Netherlands.

2 The exercise was funded by MoFA to explore innovative approaches of enhancing public accountability and facilitate learning processes regarding the ways in which private sector engagement modalities contribute to systemic change. The partnership was selected as a typical case; selection was based on the alleged contribution of CREATE to the crowding in of malting companies. The exercise was undertaken in 2018–19 in collaboration with representatives from the main partners engaged in CREATE (MoFA, EUCORD, Heineken). It consisted of two rounds of interactive workshops, and various bilateral conversations with the individual partners.

3 CREATE was implemented in Arsi, West-Arsi, and Bale zones; the BOOST project will be implemented in Assela zone.

4 The criterion of certainty relates to whether we have to find the evidence for the hypothesis (H) to be true.

5 This criterion relates to whether there are alternative explanations for the presence of the evidence.

6 (Dis)confirmation of the H is based on the certainty and uniqueness of the evidence.

References

ATA (2019) *Annual Report on Transforming Ethiopian Agriculture*, Addis Ababa: Agricultural Transformation Agency

ATA (2018) '**Malteries Soufflet Signs a Land Lease Agreement with Industrial Parks Development Corporation**', *Ethiopian Agricultural Transformation Agency*, 13 June (accessed 10 July 2021)

ATA (2017) '**Landmark Lease Agreement Signed for International Investment in Malting Factory**', *Ethiopian Agricultural Transformation Agency*, 24 November (accessed 10 July 2021)

Beach, D. (2017) '**Process-Tracing Methods in Social Science**', *Oxford Research Encyclopedias of Politics*, 25 January (accessed 14 October 2021)

Beach, D. and Pedersen, R.B. (2019) *Process-Tracing Methods: Foundations and Guidelines*, 2nd ed., Michigan MI: University of Michigan Press

Befani, B. and Stedman-Bryce, G. (2017) 'Process Tracing and Bayesian Updating for Impact Evaluation', *Evaluation* 23.1: 42–60

Bennett, A. (2015) 'Appendix: Disciplining our Conjectures: Systematizing Process Tracing with Bayesian Analysis', in A. Bennett and J.T. Checkel (eds), *Process Tracing: From Metaphor to Analytic Tool*, Cambridge: Cambridge University Press

Braskamp, L.A.; Brandenburg, D.C. and Ory, J.C. (1987) 'Lessons About Clients' Expectations', in J. Nowakowski (ed.), *The Client Perspective on Evaluation: New Directions for Program Evaluation, No. 36*, San Francisco CA: Jossey-Bass

EUCORD (n.d.) ***BOOST Ethiopia***, CREATE (accessed 24 September 2021)

Fairfield, T. and Charman, A. (2017) '**Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats**', *Political Analysis* 25.3: 363–80, DOI: doi:10.1017/pan.2017.14 (accessed 31 August 2021)

Fowler, B. and Dunn, E. (2014) *Evaluating Systems and Systemic Change for Inclusive Market Development: Literature Review and Synthesis*, Leveraging Economic Opportunities Report 3, Washington DC: United States Agency for International Development

Gebru, K.M.; Rammelt, C.; Leung, M.; Zoomers, A. and van Westen, G. (2019) 'Inclusive Malt Barley Business and Household Food Security in Lay Gayint District of Northern Ethiopia', *Food Security* 11.4: 953–66

Hay, C. (2016) 'Process Tracing: A Laudable Aim or a High-Tariff Methodology?', *New Political Economy* 21.5: 500–4

Heineken (2018) ***Growing with Ethiopia***, CREATE (accessed 24 September 2021)

Hestad, D. (2021) '**The Evolution of Private Sector Action in Sustainable Development**', *IISD Earth Negotiations Bulletin: Brief 12*, Winnipeg: International Institute for Sustainable Development (accessed 31 August 2021)

Kessler, A. (2021) *Assessing Systemic Change: Implementation Guidelines for the DCED Standard*, Donor Committee for Enterprise Development (accessed 31 August 2021)

Lavers, T. (2011) ***The Role of Foreign Investment in Ethiopia's Smallholder-Focused Agricultural Development Strategy***, LDPI Working Paper 2, The Hague: Land Deal Politics Initiative (accessed 31 August 2021)

Mapitsa, C.B. and Chirau, T.J. (2019) 'Institutionalising the Evaluation Function: A South African Study of Impartiality, Use and Cost', *Evaluation and Program Planning* 75: 38–42

Mayne, J. (2012) '**Contribution Analysis: Coming of Age?**', *Evaluation* 18.3: 270–80, DOI: 10.1177/1356389012451663 (accessed 8 October 2021)

Nippard, D.; Hitchins, R. and Elliott, D. (2014) ***Adopt-Adapt-Expand-Respond: A Framework for Managing and Measuring Systemic Change Processes***, Springfield Centre Briefing Paper, Durham: Springfield Centre (accessed 9 September 2021)

OECD (n.d.) '**The Holistic Toolbox for Private Sector Engagement in Development Co-operation**', *Private Sector Peer Learning Policy Brief* 1, Paris: Organisation for Economic Co-operation and Development (accessed 1 September 2021)

Otuki, N. (2021) ***In Ethiopia, Linking Farmers to Prosperity***, International Finance Corporation (accessed 24 September 2021)

Posthumus, H.; Shah, R.; Miehlbrandt, A. and Kessler, A. (2020) ***A Pragmatic Approach to Assessing System Change***, Durham: Springfield Centre (accessed 1 September 2021)

Punton, M. and Welle, K. (2015a) ***Straws-in-the-Wind, Hoops and Smoking Guns: What Can Process Tracing Offer to Impact Evaluation?***, CDI Practice Paper 10, Brighton: Institute of Development Studies (accessed 8 October 2021)

Punton, M. and Welle, K. (2015b) ***Applying Process Tracing in Five Steps***, Annexe to CDI Practice Paper 10, Brighton: Institute of Development Studies (accessed 1 September 2021)

Rogers, P. and Macfarlan, A. (2020) ***An Overview of Monitoring and Evaluation for Adaptive Management***, Monitoring and Evaluation for Adaptive Management Working Paper 1, BetterEvaluation (accessed 1 September 2021)

Rohlfing, I. (2012) *Case Studies and Causal Inference: An Integrative Framework*, Basingstoke: Palgrave Macmillan

Schouten, G. and Vellema, S. (2019) '**Partnering for Inclusive Business in Food Provisioning**', *Current Opinion in Environmental Sustainability* 41: 38–42, DOI: 10.1016/j.cosust.2019.10.004 (accessed 8 October 2021)

Stern, E. *et al.* (2012) *Broadening the Range of Designs and Methods for Impact Evaluation*, London: Department for International Development

Taylor, R.; Torugsa, N.A. and Arundel, A. (2018) 'Leaping Into Real-World Relevance: An "Abduction" Process for Nonprofit Research', *Nonprofit and Voluntary Sector Quarterly* 47.1: 206–27

Thorpe, J. (2014) ***Business and International Development: Is Systemic Change Part of the Business Approach?***, IDS Evidence Report 92, Brighton: Institute of Development Studies (accessed 8 October 2021)

van Westen, G.; Mangnus, E.; Wangu, J. and Worku, S.G. (2019) '**Inclusive Agribusiness Models in the Global South: The Impact on Local Food Security**', *Current Opinion in Environmental Sustainability* 41: 64–8, DOI: 10.1016/j.cosust.2019.11.003 (accessed 8 October 2021)

Wauters, B. and Beach, D. (2018) '**Process Tracing and Congruence Analysis to Support Theory-Based Evaluation**', *Evaluation* 24.3: 284–305, DOI: 10.1177/1356389018786081 (accessed 8 October 2021)

Understanding Behaviour Change in Theory-Based Evaluation of Market Systems Development Programmes

Jodie Thorpe[1]

Abstract Market systems development (MSD) programmes aim to influence private actor behaviour to promote markets that work better for the poor. This article aims to inform theory-based evaluation (TBE) of such programmes, arguing that a stronger analysis of market actor behaviour change is needed. It proposes a 'behaviour change framework' (BCF), building on recent advances in the TBE literature. These focus attention on behaviour change as contingent on the alignment of actor capability, motivation, and opportunity, influenced by the meso and macro contexts. The article applies the BCF to three theory-based MSD evaluations to illustrate its applicability and draw lessons on its use. The BCF can be used to identify evidence gaps and support more compelling explanations of what worked and under what conditions. Such evidence can inform future MSD programmes, and enable them to better stimulate systemic change in line with poverty reduction.

Keywords market system development, theory-based evaluation, behaviour change, motivation, capability, opportunity.

1 Introduction

Growing interest in business as a development actor has led donor agencies, governments, non-governmental organisations (NGOs), and businesses to implement support programmes that promote private investment in economic activities that contribute to development goals (Humphrey *et al.* 2014). One such approach is known as 'market systems development' (MSD). MSD aims to systematically understand and intervene in market systems, in order to:

> identify the underlying causes (rather than symptoms) of weak market system performance in order to realise large-scale change. Intervention should continually strive to

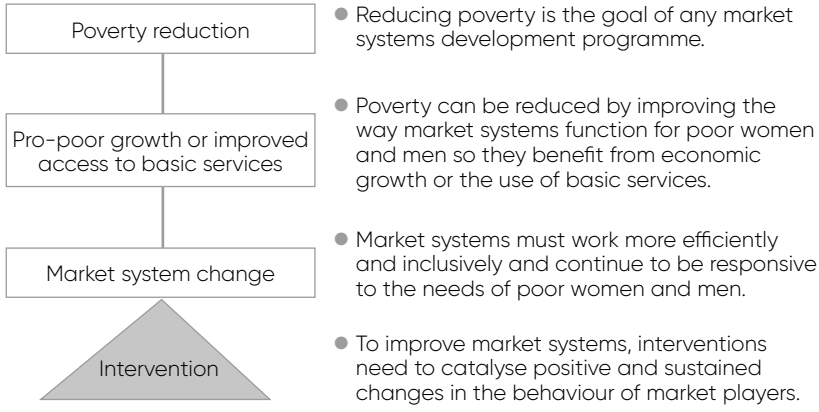
The *IDS Bulletin* is published by Institute of Development Studies, Library Road, Brighton BN1 9RE, UK
This article is part of *IDS Bulletin* Vol. 53 No. 1 February 2022 **'Theory-Based Evaluation of Inclusive Business Programmes'**; the Introduction is also recommended reading.

> leverage the actions of key market players to bring about extensive and deep-seated system change.
>
> Sustainability is a prime concern of market systems development. This means considering not just the existing alignment of key market functions and players but how they can work more effectively and inclusively in the future, **based on the incentives and capacities of market players**.
>
> The approach focuses on **stimulating a change in behaviour of market players** – public and private, formal and informal – so that they are better **able and motivated** to perform important market functions effectively.
> (Springfield Centre 2014: 3, emphasis added)

MSD interventions seek to influence the behaviours of market actors such that they are better aligned with responsible or inclusive business models, catalysing systemic changes towards more inclusive economies and poverty reduction (see Figure 1). However, this impact chain is highly stylised. The complexity of market systems means interventions do not in fact progress via a fixed or linear plan and also depend deeply on context.

Evaluating MSD programmes has proven challenging (ICAI 2014; Creevey *et al.* 2010; Coffey International Development and M4P Hub 2012; Taylor 2013). As a result, there is a lack of robust evidence and analysis showing how development programmes may best stimulate systemic change in value chains and markets (Campbell 2013; Creevey, Dunn and Farmer 2011). In light of this challenge, theory-based evaluation (TBE) has been recommended for MSD programmes (Jenal and Liesner 2017; O'Sullivan 2016; White 2009). In TBE, evaluators are encouraged to elicit and test different causal chains to understand how outcomes are achieved. Critical elements of high-quality TBE

Figure 1 Strategic framework for market systems development programmes



Source Springfield Centre (2014: 5), reproduced with permission.

include deep questioning of multiple sources of evidence and an emphasis on why and how processes being evaluated work or not, including assessing underlying assumptions and the contextual factors that influence these processes (Delahais and Toulemonde 2012; Mayne 2008, 2012; Patton 2012; Ton, Vellema and de Ruyter de Wildt 2011; White 2009).

This article argues that evaluators of MSD programmes would benefit from conceptual frameworks that make it easier to identify and assess market actor behaviour change and its relationship to the meso and macro environments. The next section examines how behaviour change is currently discussed in the TBE literature. It concludes with the presentation of a prototype 'behaviour change framework' (BCF). Section 3 describes the article's methodology for applying this framework to assemble and assess evidence from three existing MSD evaluations. Section 4 presents the findings from this process, leading to a discussion of the potential for the BCF to support future MSD evaluations in Section 5. The article's conclusions follow in Section 6.

2 Behavioural change and theory-based evaluation

This section discusses key concepts relevant to understanding behaviour change in MSD programmes, drawing from literature on TBE. It focuses on the two most mentioned types of TBE: theory of change approaches and realist evaluation (CEE 2012), discussing each in turn.

2.1 Theory of change approaches

Theory of change approaches are based on understanding how programme interventions are intended to function, linking activities to outputs, immediate and intermediate outcomes, and impacts, including the assumptions inherent in these causal chains. Mayne (2015) introduces what he describes as more 'intuitive' labels to be used in these chains. He uses **behaviour change** instead of immediate outcomes, **direct benefits** for longer-term outcomes, and **wellbeing changes** for impacts. In between outputs and behaviour changes, Mayne also introduces two steps: (1) **reach and reaction**, and (2) **capacity change**. Reach and reaction refer to the spread of ideas or incentives to groups targeted by an intervention, and their initial response. In MSD, these groups would be market actors, such as manufacturers, banks, or business service providers. Capacity changes relate to knowledge, attitudes, skills, aspirations, and opportunities (Mayne and Johnson 2015), and are a prerequisite for new actions to be taken.

Further work by Mayne (2018) draws on the COM-B model, a behaviour change system set out in Michie, van Stralen and West (2011), which was developed from a systematic review of behavioural approaches in the health sector. In this system, three elements interact to generate capacity change: (1) **motivation**, or the internal processes which direct behaviour, including

both reflective or analytical processes and more automatic or instinctive habits, norms, and emotional responses; (2) **capability**, including the physical and psychological capacity to act; and (3) **opportunity**, or the external factors (outside the individual) that enable or block behaviours, related to the physical, social, or cultural environment and to systems of rules or incentives, which influence an actor's expectations of reward or punishment. Among these factors, motivation plays a particular role as it involves the choices and habits that energise and direct behaviour (*ibid.*). Both capability and opportunity can have an impact on motivation, such as by promoting new ways of thinking.

Notably, all three elements need to be present to drive capacity and behaviour change (Darnton 2008; Mayne 2018). Programmes therefore need to establish which of these element(s) are preventing desired behaviours, and design interventions to address gaps. Feedback loops are also a crucial component, and there is often a feedback loop from new behaviours to the future capacities of actors (Mayne 2015). For example, new knowledge regarding market opportunities that has been generated as a result of product innovation might motivate further innovation, while poor results may deter it.

2.2 Realist evaluation

Realist evaluation tests hypotheses about which programme interventions work, for whom, and under what conditions. The focus is on causal mechanisms that motivate actor behaviour, and particularly whether and how programme interventions stimulate new behaviours (Pawson and Tilley 1997; Ton *et al.* 2011; Jenal and Liesner 2017). Recognising that these processes are contingent on context, the hypotheses to be tested are expressed in the form of Context-Mechanism-Outcome (CMO) configurations.

In relation to MSD, we can define context to include institutional, organisational, socioeconomic, and cultural conditions and resources affecting specific (groups of) market actors. Outcomes are observable behavioural changes stemming from these actors' decisions, which are influenced by context and by programme interventions. Mechanisms are key to behaviour change. They are the incentive structures that shape actor decisions, and which programme interventions aim to influence. Realist evaluation also recognises feedback loops through which outcomes may influence (strengthen or dampen) causal mechanisms.

There is debate in the literature over the nature of causal mechanisms. While Pawson and Tilley (1997) have explained mechanisms as being related to actor reasoning and resources, others (Westhorp 2018; Ton 2021) define them as working at different levels of social systems. Mechanisms therefore include 'the inner motivations of people and firms' as well as 'the power of structures that shape or constrain their agency' (Ton 2021).

Westhorp argues that multiple constructs are needed in order to assess how and why programme interventions work, particularly when viewed from a whole system perspective.

> Here is the crux of the issue: the causal properties of systems are not solely reducible to the decision-making of people within those systems. The implication for evaluation is equally clear. If programmes are indeed social systems, as Pawson and Tilley have eloquently argued, then the causal properties of the programmes are by definition not reducible solely to the decision-making of the targeted individuals.
> (Westhorp 2018: 8)

Instead, Westhorp suggests that mechanisms operate across different system levels, which may include material (biochemical, physical), individual, social-group, and social-institutional. Her key contribution is to emphasise that **systemic change happens across these levels**, sometimes in different time frames.

> Realism has long acknowledged that mechanisms operate at different levels of the system than their outcomes… It is necessary to look to the sub-systems – of what they are comprised, what they do and how they do it, and what the consequences of their operations are – in order to understand how a system – or some aspect of it – works. However, realism also acknowledges that causation works downwards, as well as upwards.
> (*ibid.*: 5).

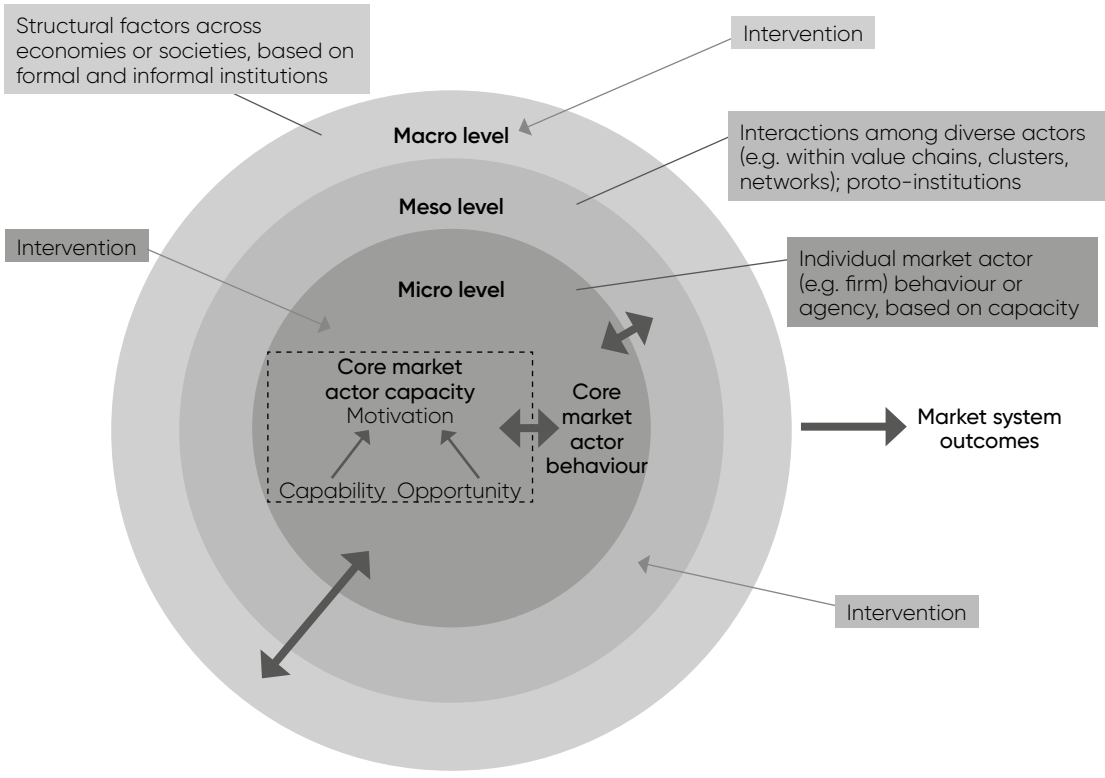
2.3 Behaviour change framework

Based on ideas drawn from both theory of change and realist perspectives, Figure 2 presents a framework for evaluating behaviour change in MSD programmes. This behaviour change framework (BCF) integrates the COM-B model (Michie *et al.* 2011), as presented in Mayne (2018), and Westhorp's (2018) insight that mechanisms operate at different system levels. It has resonances with the framework presented in the introduction to this edition, as well as other work on stakeholder behaviour in value chains (Ton *et al.* 2021; Ton 2021).

At the centre is a market actor, such as a firm or enterprise, that the programme seeks to reach and influence in order to produce behaviours in line with poverty reduction. Actor behaviour is determined by capacity, which is the product of capability + opportunity + motivation. Although the actors in MSD are primarily composite actors like firms, rather than individuals, their behaviours still result from coordinated actions by the individuals involved. Where MSD programmes have multiple components that target different groups of actors, the BCF would be applied separately to each group of interest.

MSD programmes catalyse sustainable changes in market actor capacity by avoiding direct solutions at micro level. Instead,

Figure 2 Behaviour change framework for market systems



Programme mechanisms related to	
Motivation, based on…	Internal decision-making processes; or automatic habits, norms, emotions
Capability, based on…	Physical and psychological capacity
Opportunity, enabled by…	External physical, social or cultural environment; rules or incentives creating expectation of reward/action

Source Author's own, based on concepts from Mayne (2015, 2018), Michie *et al.* (2011), Westhorp (2018). See also Ton *et al.* (2021) and Ton (2021).

they influence the availability or quality of meso and macro level support functions, services, and institutions which in turn influence actor capacity. In realist terms, these meso and macro interventions create mechanisms that motivate actors towards desired behaviours. The meso level involves interactions among diverse actors (e.g. within value chains, clusters, networks, communities), and may comprise proto-institutions, such as voluntary standards or multi-stakeholder initiatives. The macro level involves structural factors that work across economies or societies, based on formal and informal institutions (van Wijk *et al.* 2019). The BCF (Figure 2) maps programme pathways from

Table 1 **Sample of theory-based MSD evaluations analysed**

Programme	Programme aim	Source	Type of evaluation	Evaluation focus	Target market actor assessed
Developing Effective Private Education Nigeria (**DEEPEN**)	Improve the quality of education provided by private schools in Lagos	MacAuslan *et al.* (2018)	Theory of change	Whole programme	Private schools in Lagos
Financial Sector Deepening Trust Kenya (**FSDK**)	Generate sustainable livelihood improvements through better financial sector capacity and operations	Stone, Johnson and Hayes (2010)	Theory of change	Sample of 13 projects (5 micro, 4 meso, 4 macro)	Equity Bank
Oxfam's Gender Transformative and Responsible Agribusiness Investments in South-East Asia (**GRAISEA**)	Improve livelihoods of women and men small-scale producers through more responsible and inclusive value chains and private sector investments	Tobing-David (2019)	Realist	Whole programme	Vietnamese agribusiness

Source Author's own.

interventions to meso- or macro-level outcomes, and ultimately to actor capacity change at micro level.

In the BCF, all three elements of capability, opportunity, and motivation need to be in place in order to generate the desired behaviours and outcomes. This condition is achieved through a combination of pre-existing contextual factors and programme mechanisms. The programme's assumptions and intentions regarding these three elements can also be highlighted (see the table 'Programme mechanisms' at the bottom of Figure 2).

Embedded in these processes are feedback loops through which outcomes may amplify or dampen their causes, indicated as double-headed arrows in Figure 2. Feedback loops often link market actor behaviour and market actor capacity, for example. They may also link micro, meso, and macro levels of the market system, since micro-level changes in the behaviour of actors can also contribute to new meso and macro contexts (Westhorp 2018).

3 Methodology

The rest of this article investigates the applicability and added value of the BCF in TBEs. To do so, it applies the framework to three existing MSD evaluations, identifying, assembling, and re-examining evidence across macro, meso, and micro levels. Publicly available programme evaluations are chosen for this study as they present a comprehensive account of MSD programme results, reflecting on the systemic nature of outcomes

achieved. Such evaluations are frequently used as a core source of learning by programme funders and implementers to understand what works, under what conditions.

3.1 Data sources and sampling

The three evaluations were selected from the Building Effective and Accessible Markets (BEAM) Exchange evidence map (BEAM Exchange 2018), a database of published resources that investigate the connection between MSD interventions and programme results. In this map, evaluations are tagged based on the results level that they illustrate. As of 1 April 2021, the database contained 90 sources which presented 'high confidence' evidence and learning on MSD effectiveness. Twelve are independent, theory-based impact evaluations, of which three were selected for this study (see Table 1). These three were chosen because they (a) primarily illustrate the 'intervention' or 'systemic change' results levels, which were expected to provide a deeper and richer discussion of market actor behaviour change, and (b) represent a variety of contexts, covering the education, finance, and agriculture sectors across three countries in Africa and Asia.

3.2 Data analysis

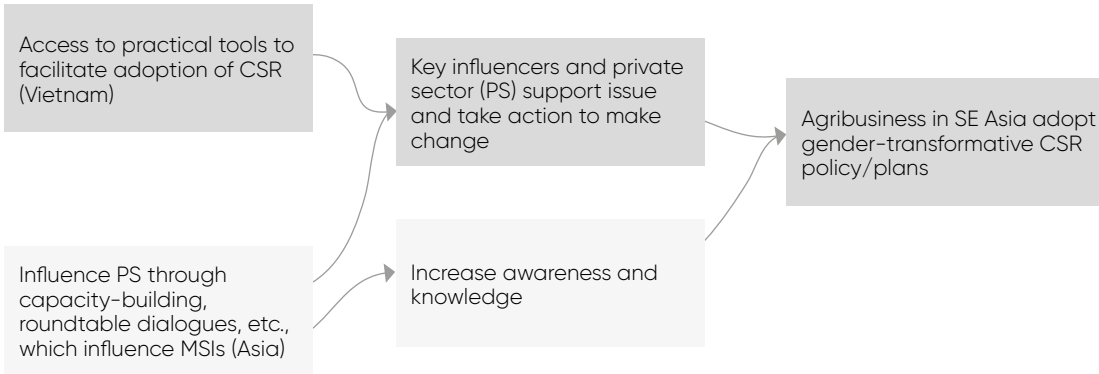
The application of the BCF to these evaluations involved three steps:

1 **Identification of the core market actor** whose behaviour change is the target of the programme (Table 1). In two of the three cases, FSDK and GRAISEA, the evaluations assessed multiple programme elements involving different targets. In these cases, just one core market actor was selected for illustrative purposes, prioritising those where the evaluation offered detailed information across system levels. Once the core actor was identified, relevant evidence on behaviour changes and factors contributing these outcomes were identified within the evaluation.

2 **Application of the BCF** to assemble the evidence on behaviour change. Evidence included the evaluation's conclusions on programme interventions at micro, meso, and macro level and/or their contribution to changes in capability, opportunity, or motivation. As the evaluations rarely used this exact terminology, the definitions in the conceptual framework were used to assign these labels. In addition, the evaluation findings were reviewed to identify insights regarding contextual factors, feedback loops, or programme assumptions relevant to actor capacity. This process resulted in Figures 4 to 6, which are presented in the next section.

3 **Assessment of BCF insights.** The final step involved comparing the change dynamics as described in the evaluations with the insights suggested by the BCF, in order to consider the ways in which the framework could offer enhanced learning for MSD programmes.

Figure 3 GRAISEA's most significant outcome of 'result 3'



Note The darker shade indicates the pathway to the most significant outcome. The original diagram also included trajectories for the Philippines, Myanmar, Thailand, and Cambodia, but these were not labelled as being 'most significant'. For the sake of simplicity and clarity, these have been left out here. MSIs refers to multi-stakeholder initiatives.
Source Author's own, adapted from Tobing-David (2019: 26).

4 Findings

4.1 GRAISEA

Gender Transformative and Responsible Agribusiness Investments in South-East Asia (GRAISEA) was an Oxfam programme that aimed to improve the livelihoods of small-scale producers through catalysing more responsible and inclusive private sector activity. It targeted leading agribusinesses, financial institutions, multi-stakeholder initiatives, and national legislation in support of more sustainable production practices in four value chains across seven Asian countries. The evaluation analysed the programme's contributions in four results areas, identifying and discussing the most significant outcomes that 'theoretically showed the strongest logical link and empirically demonstrated positive results' (Tobing-David 2019: 8). The evaluation explicitly uses a realist approach, exploring strategy effectiveness with respect to seven causal mechanisms.

To illustrate the use of the BCF, this article focuses on what the evaluation terms 'Result 3'. In this component, GRAISEA sought to catalyse Asian agribusinesses to adopt corporate social responsibility (CSR) policies that support small-scale producers and gender equity. The most significant outcome identified by the evaluation was the adoption of gender-transformative CSR policies and plans in Vietnam, with 31 seafood companies adopting gendered CSR guidelines and reporting, and 13 companies reporting full compliance (*ibid*.: 26). Figure 3 reproduces a segment of a diagram from the evaluation report which illustrates this outcome trajectory.

The evaluation concludes that three critical factors contributed to GRAISEA's achievements. Firstly,

> in a country like Vietnam where there is a strong state presence, the government plays a truly defining role. Secondly, international standards imposed by the export market mean that CSR has a commercial value, and lastly, Oxfam in Vietnam made CSR practices more 'practical' by introducing gender CSR Guidelines and Sustainability Index reporting for the companies to experience it.
> (*ibid*.: 38)

Compatibility between companies' values and the programme's goals was another contributory factor (*ibid*.: 36).

In contrast, the evaluation found that programme activities to convene and influence multi-stakeholder initiatives (MSIs) were not impactful. GRAISEA had assumed that large corporates were highly engaged in MSIs, and highly likely to make reference to MSI guidelines in their business strategies. However, there was little evidence that these assumptions held true (*ibid*.: 38).

4.1.1 Applying the behaviour change framework

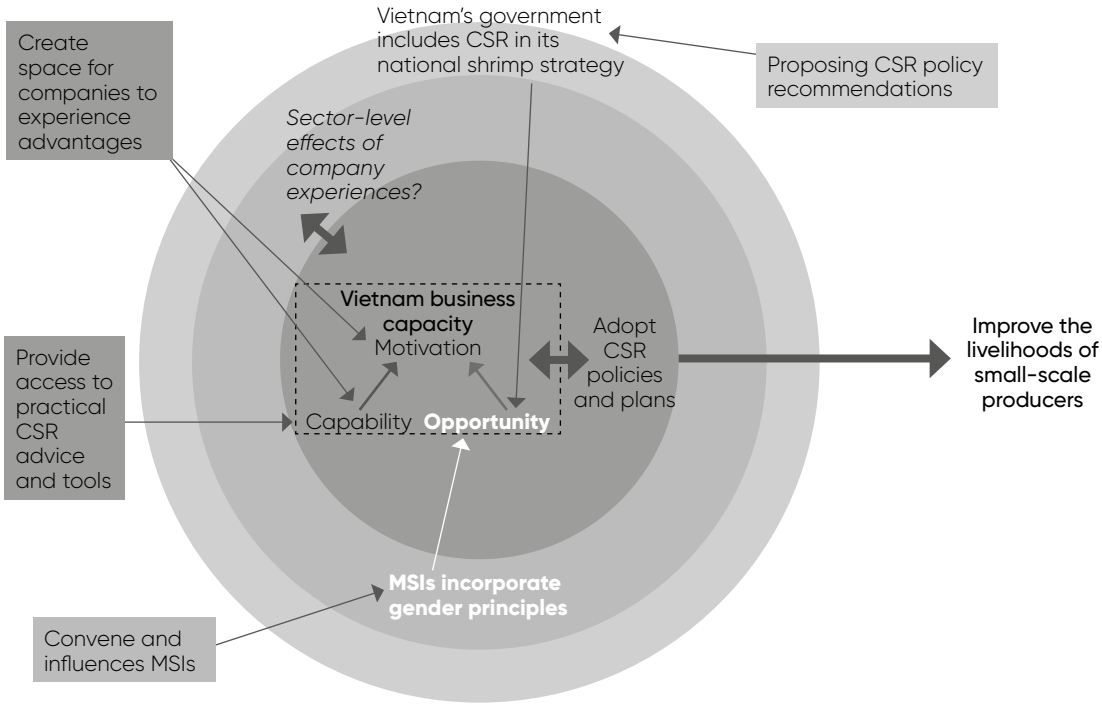
Figure 4 presents GRAISEA's evidence through the BCF lens. There are similarities with GRAISEA's outcome trajectory, especially in framing pathways in terms of target actors and desired behaviours. However, the BCF introduces an additional pathway at macro level, drawing on detail from other sections of the evaluation. Through the BCF, these macro, meso, and micro interventions are assessed with respect to supporting capability, opportunity, and motivation, adding detail from the evaluation report on programme mechanisms and contextual factors.

The BCF shows GRAISEA playing a key role in providing access to knowledge and tools, which foster technical capability for CSR within the private sector. It does raise a question, however, highlighted in italics (in Figure 4), of whether this direct delivery approach has generated sector-wide effects, in line with MSD systems thinking.

Comparing the BCF to Figure 3, the BCF adds most detail with respect to **opportunity**. At macro level, it adds the programme's work on gender-sensitive CSR guidelines and the government's adoption of CSR elements in Vietnam's national shrimp strategy (*ibid*.: 18). In the context of a strong state, this development is likely to have had an important bearing on companies' perceptions of opportunities from CSR, either in the form of rewards or punishments. While this pathway is clear in the BCF, it is only represented in Figure 3 through a reference to 'key influencers'.[2]

At meso level, the BCF (Figure 4) presents a similar picture to Figure 3, with both indicating weak opportunities resulting from GRAISEA's work with MSIs. However, the BCF also draws out an important contextual factor showing that standards imposed in export markets influence opportunities for Vietnamese companies trading outside the country.

Figure 4 Capacity and behaviour change of agribusiness in Vietnam



Motivation, based on…	● having a safe space to experiment ● company culture (commitment to CSR)
Capability, based on…	● having access to CSR advice and tools
Opportunity, enabled by…	● a strong state that promotes CSR adoption ● international standards imposed by the export market ● **collective governance mechanisms, which influence corporate behaviour**

Source Author's summary of factors linked to the adoption of CSR by Vietnamese agribusinesses, based on Tobing-David (2019).

Finally, although not visible in Figure 3, the evaluation identifies ways in which GRAISEA affected companies' **motivations**. This is described as giving 'space for companies to understand and experience its relative advantage' from using CSR tools (*ibid*.: 36), which was further enabled by the trust that Oxfam built with these companies. The evaluation also identifies more intrinsic motivations, notably company awareness and commitment to act responsibly, especially in the case of micro, small and medium-sized enterprises (MSMEs). 'Unless there is a strong drive to adopt responsible business conduct policies and practices, especially ones that recognize the significant role of women, these MSMEs will be less likely to adopt' (*ibid*.: 13).

4.2 FSDK

Financial Sector Deepening Trust Kenya (FSDK) aimed to support the development of an inclusive Kenyan financial sector, building on an earlier programme of technical support. Both programmes were funded by the UK's Department for International Development (DFID).[3] FSDK developed a portfolio of 34 projects targeting policy and regulatory change, sector support services, and retail banking capacity. Together, these projects were intended to impact the capacity and operation of the sector, and to generate sustainable livelihood improvement for poor Kenyans. The evaluation focuses on the validity of the FSDK impact pathways across a sample of these projects using a TBE framing. 'The aim… was, first, to establish the theoretical programme impact pathways and, second, to obtain evidence that can substantiate (or refute) the effective functioning of these pathways in practice' (Stone, Johnson and Hayes 2010: 6).

The evaluation report discusses changes at three levels, which it explicitly defines as macro (policy and regulation), meso (sector support services), and micro (retail). It finds a significant contribution of FSDK at all three levels (*ibid*.: v), and highlights strong synergies, citing Equity Bank as a key example (*ibid*.: 17). FSDK helped Equity, a former building society, to transform into a bank, while its support for MicroSave, a consulting company providing product development support, contributed to the Equity Bank's subsequent expansion. Policy influencing, enabled by DFID, also contributed to this transformation and growth.

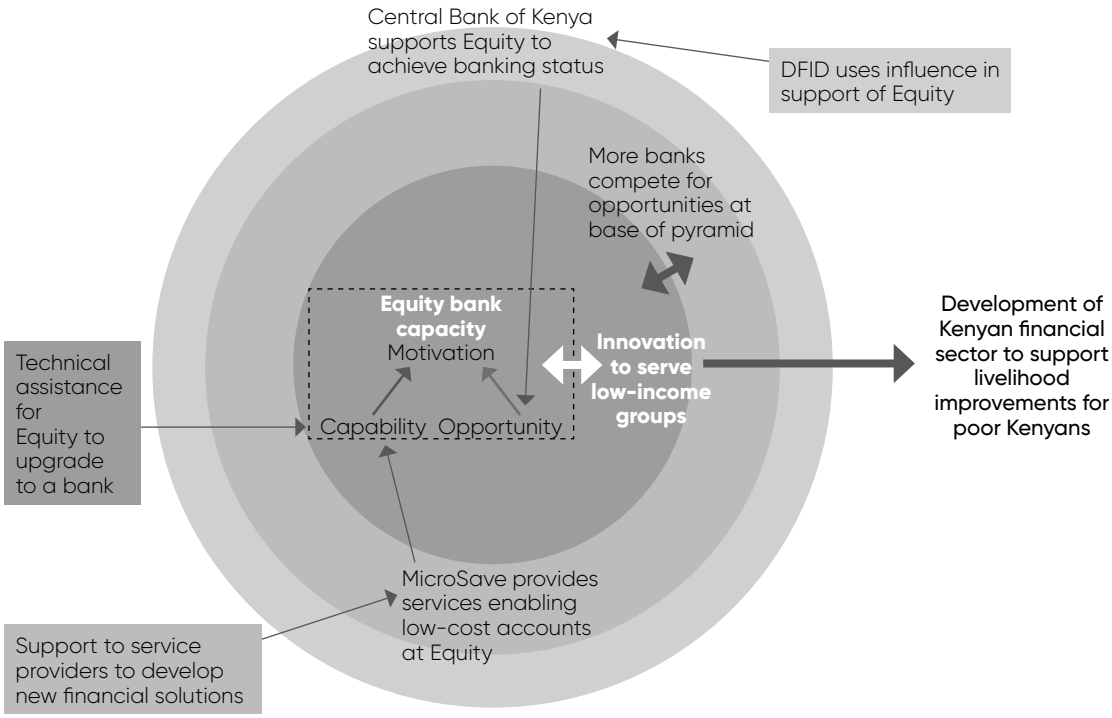
4.2.1 Applying the behaviour change framework

Figure 5 presents evaluation evidence through the BCF lens, using Equity Bank as the example. While the FSDK evaluation contains a very similar visual of micro, meso, and macro levels of support in a nested structure (*ibid*.: 5), it does not extend the use of this visual to present specific outcome pathways, nor does it relate interventions to changes in capability, opportunity, or motivation. The BCF fills these gaps, and in doing so, helps to illustrate and explain synergies across FSDK pathways.

Early macro-level interventions supported by DFID[4] played a role in influencing Kenyan policymakers, smoothing the way for Equity's evolution into a bank, and enabling new growth **opportunities**. Interventions at both micro and meso levels supported new **capabilities**. Micro interventions provided technical assistance for Equity's upgrading. Meso-level interventions targeted MicroSave, enabling it to develop the financial solutions that would inform new product development at Equity Bank. The evaluation presents no specific evidence relevant to **motivations**, which may imply that capability and opportunity together motivate innovation.

On the other hand, the evaluation does highlight micro to meso links, indicated by the double-ended black arrow in Figure 5. It finds that Equity Bank's successes have impacted on the

Figure 5 Capacity and behaviour change of Equity Bank



Motivation, based on…	capability + opportunity
Capability, based on…	access to technical support
Opportunity, enabled by…	a regulatory regime aligned with Equity's transformation (from building society to bank)

Source Author's summary of factors linked to changes at Equity Bank, based on Stone *et al.* (2010).

culture of the wider finance sector, by demonstrating market opportunities for different customer segments (*ibid*.: 9). As a result, mainstream banks are beginning to compete in lower-income markets (*ibid*.: 12).

Despite these successes, the evaluation found that poorer clientele were still missing out.

> Compared with 2006, we found that Equity had clearly more than proportionately increased its outreach to the rural population, women, younger people and the less-educated… [but] it has not clearly achieved outreach to a poorer clientele any more than has the rest of the banking sector. (*ibid*.: 11–12)

This weakness is indicated in Figure 5 by the white arrow between Equity Bank's capacity and serving low-income groups.

Unfortunately, it is not clear from the evidence presented whether the barrier lies with capability, opportunity, or motivation, although the evaluation does question the suitability of Equity Bank's accounts from the perspective of poorer clients.

In future, the BCF could be applied to help evaluators probe such issues more deeply. Is this primarily a capability issue, affecting Equity Bank's product portfolio, as the evaluation seems to suggest? Or is the root cause at the level of market opportunity? The only causal mechanism linked to opportunity for Equity Bank is rooted in its transformation from a building society, but is it realistic to think that this change created new opportunities that motivated Equity to work more closely with 'base-of-the-pyramid' customers? Perhaps other measures such as tax incentives or universal service obligations would be needed? Or perhaps the root cause lies with intrinsic habits and norms that shape the bank's motivation to serve this sector?

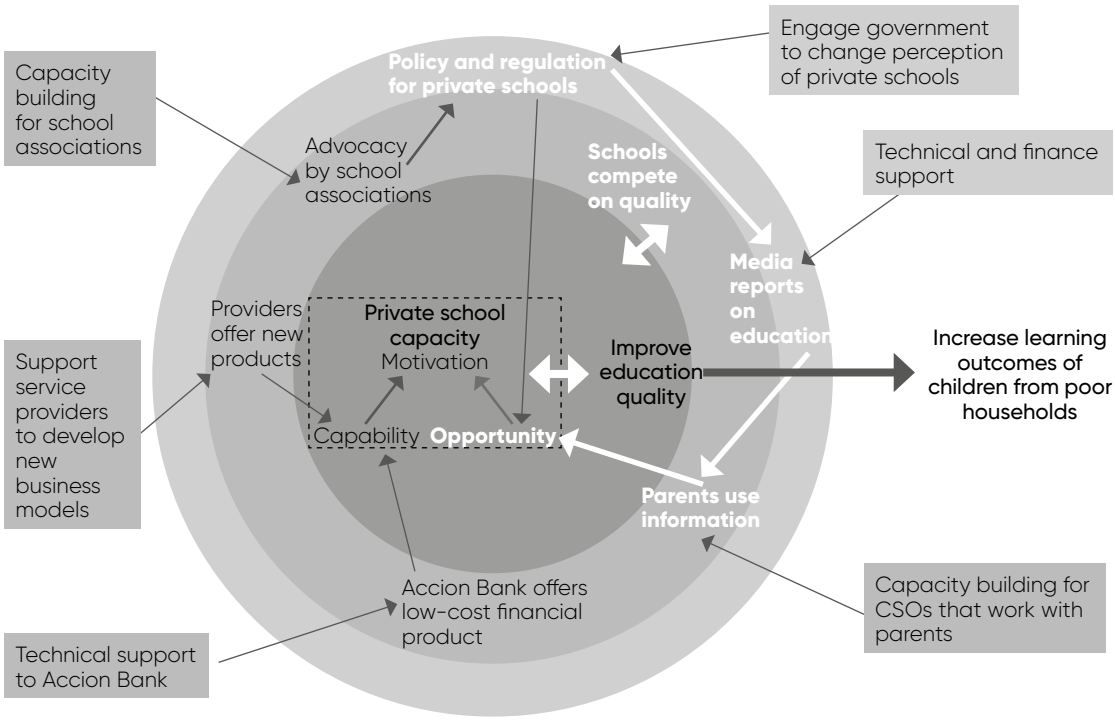
4.3 DEEPEN

Developing Effective Private Education Nigeria (DEEPEN) aimed to improve private school education in Lagos state. It sought to address core constraints, mainly information asymmetries, especially parents' information about school quality; missing support functions, including access to finance and teacher training services; and an unreceptive regulatory regime that left many schools operating informally. By facilitating innovations, DEEPEN intended to improve the quality of education delivered by private schools, particularly those serving poor children. The evaluation 'assesses DEEPEN by following its theory of change and gathering data on the key assumptions and context, as well as expected outputs and outcomes' (MacAuslan *et al.* 2018: 4). It gathered evidence across the programme's four workstreams: (a) rules and standards, (b) information, (c) finance, and (d) school improvement.

The evaluation found that with the exception of the information workstream, DEEPEN made considerable progress. It influenced government perceptions, policies, and practices towards private schools, and supported credit provider Accion Bank to develop a low-cost financial product. Service providers also developed new and affordable school improvement programmes, although these were out of reach of the poorest schools (*ibid*.: iv). However, the evaluation also identified major limitations to DEEPEN's outcomes. There were 'only very modest changes in behaviour in the low-cost schools that were surveyed' (*ibid*.: 10), in terms of improved capacity or better learning conditions. These schools struggled to access credit or pay for improvement programmes, despite increased affordability. Overall, few gains were detected in pupils' learning outcomes.

Failure of the state government to fully implement the Graded Assessment of Private Schools (GAPS) legislation was identified

Figure 6 Capacity and behaviour change of private schools in Lagos



Motivation, based on…	● capability + opportunity
Capability, based on…	● access to technical support ● access to finance
Opportunity, enabled by…	● state governments recognise private schools ● **parents identify/pay for higher quality**

Source Author's summary of factors, based on MacAuslan *et al.* (2018).

as a major factor (*ibid.*: v). GAPS had been intended to rate the quality of private schools, with results made available to parents and the media. DEEPEN expected that these stakeholders would then generate incentives for schools to invest in quality improvement. However, a change of government in 2015 unexpectedly restricted the roll-out of GAPS. According to the evaluation, this severely constrained the impact potential of all of DEEPEN's interventions (*ibid.*: 31).

4.3.1 Applying the behaviour change framework

Figure 6 presents the evaluation evidence viewed through the BCF lens. It provides a new visual representation of the impact pathways and their interactions as described in the evaluation. It also

encourages thinking about how these pathways affect motivation, capability, and opportunity, rather than focusing too narrowly on the GAPS policy failure. The elements highlighted in white indicate multiple breakdowns in mechanisms affecting opportunity, which together explain the modest results achieved by DEEPEN.

On the positive side, DEEPEN's efforts to strengthen schools' **capabilities** through engaging service and credit providers broadly functioned well, although very low-cost schools remained unable to access credit or afford improvement programmes. With respect to **opportunity**, however, DEEPEN had intended to influence both macro and meso environments in ways that would generate rewards for schools investing in improvements. Although DEEPEN did succeed in improving the state government's recognition and support for private schools, positively influencing their external environment, other key market drivers were missing.

As the evaluation identified, the problems started with the failure of the state government to fully implement GAPS, which would have delivered important information on school quality. However, the weaknesses in the intended impact pathways do not stop there. Even when DEEPEN attempted to compensate for GAPS failings by directly giving media outlets technical assistance and financial support for educational programming, there is little evidence that parents were actually listening to the radio for information on education and school quality (*ibid*.: 16). Hence the intended causal mechanism involving parents identifying and paying for higher quality schools is weak. In addition, the findings suggested that the educational programming supported by DEEPEN is unsustainable beyond the end of the programme, as it does not align with the commercial interests of the radio stations (*ibid*.: 16), which will not pursue it.

As for FSDK, there is no discussion of causal mechanisms or contextual factors linked to motivations, suggesting implicitly that capability and opportunity together would be sufficient. In addition, DEEPEN also assumed that competition would be an important causal mechanism to scale up change. However, in contrast to FSDK, competition did not play this expected role and this link is highlighted in white in Figure 6. In part, this finding reflects the weak opportunities already discussed. However, the evaluation also finds that:

> while competition plays a role, and some of the proprietors who were interviewed by the endline evaluation team indeed felt protective of their know-how, there appears to be a high degree of collaboration. This is consistent with the finding that private schools do not always operate on market logic, and that many see themselves more as social enterprises or charitable organisations that are serving an important need. (*ibid*.: 28)

Finally, the BCF also draws attention to a potentially important negative feedback loop through which the outcome (improved school quality) limits or undermines capacity for change, particularly in low-cost schools. In these settings, studies have shown that there is frequent teacher turnover, leaving school proprietors reluctant to invest in teacher training (*ibid*.: 28). The reasons for the turnover are unclear and are likely to be complex. However, to the degree that training enables teachers to access better jobs elsewhere, it would represent a negative feedback loop.

5 Discussion: Evaluating behaviour change in MSD programmes

The findings demonstrate how the BCF can enrich TBEs, by aiding evaluators to visually represent and systematically assess market actor behaviour change. In the case of DEEPEN, the framework helps to focus the findings on the (lack of) opportunities for private schools to improve. With FSDK, it encourages deeper understanding of the intersection of micro, meso, and macro factors as they relate to Equity Bank's capacity. Within the scope of this article, the result is a stronger visual and narrative of Equity's successes. However, the BCF could also be used to seek new evidence to better understand the programme's failures to benefit lower-income groups. In the case of GRAISEA, the BCF brings together programme mechanisms and contextual factors currently discussed across the report and shows how they interact to contribute to CSR adoption in Vietnam.

Based on these findings, this section draws out lessons for evaluators – and by extension for those who use the results of TBEs. These insights relate to the drivers of behaviour change for different private sector actors, the interactions between these drivers at micro, meso and macro levels, and the use of the BCF to capture these dynamics.

5.1 Analysing the drivers of private sector behaviour change

Fundamentally, the BCF provides a framework for those conducting TBEs to bring together and think critically about multiple sources of evidence relevant to assessing market actor behaviour change. The key is recognising that capability, opportunity, and motivation must all be present and aligned with the desired behavioural outcomes (Darnton 2008; Mayne 2018). These factors may either be pre-existing or be catalysed through programme interventions. Capability and opportunity together influence motivation, although programme interventions may also directly contribute.

Across all three evaluations, programmes were generally successful in their technical and financial support to build company capabilities. However, the findings suggest that evaluators should be particularly interested in assessing opportunity for change. For both DEEPEN and GRAISEA, impact pathways targeting opportunity were hampered by problematic assumptions which the BCF helped to highlight. Understanding motivations can also

help to explain how and why change happens. While motivations may be difficult to observe directly, they can be probed with respect to decision-making or preferences.

In this respect, the BCF can be used to enrich what has been termed the 'will-skill' framework within MSD practice (Springfield Centre 2014). According to this framework, MSD interventions may address market actor capability ('skill') or their incentives and motivations ('will'). Where capabilities are high but motivations are weak, programmes can focus on making the case for change, for example, or on reducing companies' perceptions of risk through co-funding investments. However, where actors already have high will and high skill but are not exhibiting the desired behaviour, it implies that obstacles lie in the external landscape (i.e. related to opportunity). The BCF thus offers a will-skill-opportunity framework, and can help programmes and evaluators think more about the meso and macro factors shaping opportunity.

Finally, the BCF encourages evaluators to pay greater attention to feedback loops and particularly the ways in which behaviour change outcomes influence actor capacity. Such dynamics were not explicitly discussed in any of the three evaluations included in this article, despite their prevalence in systemic change processes. However, in the case of DEEPEN, a potential negative feedback loop was identified in which teacher training that was intended to raise the capacity of low-income schools may lead to teachers using their new skills to seek better opportunities elsewhere, returning the school to its low-capacity state. The BCF can prompt evaluators to ask more questions about such loops.

5.2 Whose behaviour?

Studies applying the COM-B system have mostly been concerned with health interventions to change individual behaviours in areas such as smoking (Barker, Atkins and de Lusignan 2016; Gould *et al.* 2017; Suntornsut *et al.* 2016). However, the BCF extends the use of COM-B concepts to the composite market actors that are the focus of MSD programming. For composite actors, choices, decisions, and behaviours reflect 'the joint intended effect of coordinated action as expected by the participating individuals' (Scharpf 1997: 52). These composite actors are influenced not only by an objective ('rational') analysis of self-interest but also by subjective motivations. For example, for GRAISEA in Vietnam, the compatibility of MSMEs' values with CSR activities was identified as an important factor in companies' capacity to adopt CSR.

Given the complexity of MSD programmes, one challenge can be to identify which actor(s) should be the focus of the micro-level behaviour change in the BCF. In the case of GRAISEA, for example, interventions targeted policymakers, MSIs, and agribusinesses. However, the central actor in the BCF is the one whose behaviour is directly affecting poverty outcomes, and

whose incentives MSD programmes seek to change. For GRAISEA, these are Asian agribusinesses. That said, the meso and macro layers are also populated by actors, who could theoretically be analysed using the BCF lens. In the case of GRAISEA, Oxfam built trust with and offered expertise to policymakers, enabling new behaviours, in the form of CSR policy decisions. The designation of micro vs meso and macro within the BCF is fundamentally an analytical choice, shaped by a programme's theory of change and the evaluator's questions.

The BCF can also be used to think in a more granular way about actor behaviour change and particularly the motivations of different target actors. Taking the case of DEEPEN again, the outcomes for schools that serve the poorest children were found to be much more modest than for the others. These schools' motivations were affected by lower financial and technical capabilities, including higher teacher turnover, and they are also likely to face different opportunities than more affluent schools. Another finding from the DEEPEN evaluation is that the motivations of schools that act as social or community enterprises are different from fully commercial providers. Linking this finding to the BCF suggests that collaboration rather than, or in addition to, competition can be an important mechanism to support the scaling of micro-level behaviour changes to the wider sector. Finally, for FSDK, the BCF helped to highlight the evaluation's finding that Equity Bank's ability to serve low-income groups was limited. It could also be used to assess the root causes of this constraint, whether linked to opportunity, capability, or motivation.

5.3 At what level?

Alongside focusing on composite actors, the BCF extends the use of COM-B ideas to systematically capture behaviour change drivers at meso and macro levels. In the case of FSDK, the evaluation had already discussed changes in macro, meso, and micro terms. However, GRAISEA and DEEPEN did not use these designations and Figure 4 and Figure 6 show how their results can be mapped in this way. Admittedly, the more linear outcome trajectories presented in the GRAISEA evaluation (Figure 3) are simpler to understand. However, the price of this simplicity is that many important elements which Figure 4 readily captures are buried in long passages of text.

The BCF can be used not only to represent how macro- and meso-level interventions shape micro-actor behaviour, but also how new micro-level behaviours influence the meso and macro contexts. Both FSDK and DEEPEN, for example, expected scale to emerge through the demonstration effects from micro-level behaviour interacting with competition at sector level, although this mechanism was more effective in the case of FSDK than DEEPEN. These dynamics are represented in the BCF through the arrows linking micro to meso levels.

Using the BCF to lay out and evaluate impact pathways connecting these multiple levels aligns with realist understanding that causal mechanisms of change operate at a different system level than their outcomes (Bhaskar 1997; Westhorp 2018). In this sense, the BCF also aligns with structuration theory (Giddens 1984) and 'actor-centred institutionalism' (Mayntz and Scharpf 1995; Scharpf 1997) which emphasise that social phenomena are the product of the interaction between intentional choices by actors, and the institutional context in which they occur.

6 Conclusion

This article aims to inform the design of theory-based evaluations for market systems development programmes through encouraging a stronger analysis of market actor behaviour change. It develops and tests a new behaviour change framework (Figure 2), which has been informed by ideas discussed in the TBE literature. At the centre is the COM-B model (Michie *et al.* 2011; Mayne 2018), showing that stimulating particular behaviours requires that capability, opportunity, and motivation are all present. In addition, inspired by Westhorp (2018), the BCF shows how micro-level behaviour change needs to be understood with respect to multiple system levels. Dynamics in the macro and meso environment create the conditions for behaviour change, while micro-level behaviours can condition the meso and macro environment for others.

In the three MSD evaluations discussed in this article, interventions and assumptions related to capability, opportunity, and motivation were readily detected. However, the BCF requirement that all three of these elements align with the desired behaviours encourages deeper critical thinking. In this way, the BCF enables evaluators to seek new evidence and/or more compelling explanations of what has worked within MSD programmes, under what conditions, as well as to explain why programmes miss the mark. Considering the findings presented here, programmes and evaluators could pay more attention to whether and how technical capabilities supported by programmes are matched by meaningful opportunities and aligned with companies' conscious and intrinsic motivations.

As the BCF is actor-focused, it encourages evaluators to think much more carefully about whose behaviour is being changed, with due attention to nuanced differences in opportunities, capabilities, and motivations. This also means identifying problematic assumptions with respect to the motivations of different types of enterprises or the opportunities available in different market segments. It was notable that in two of the three cases explored in this article, programme successes did not adequately translate to benefits for low-income groups. If future TBEs could generate better understanding of these dynamics, it would provide valuable insights regarding how MSD programmes can better stimulate systemic change in line with poverty reduction.

Notes

1 Jodie Thorpe, Research Fellow, Institute of Development Studies, University of Sussex, UK.
2 This very general reference to 'key influencers' is likely to have been a simplification to fit the diagram. However, it gives little visibility to the important role of state support, which is described elsewhere in the evaluation as being 'truly defining'.
3 Now the Foreign, Commonwealth & Development Office (FCDO).
4 FSDK's support for Equity's transformation was a continuation of an earlier DFID-funded programme.

References

Barker, F.; Atkins, L. and de Lusignan, S. (2016) '**Applying the COM-B Behaviour Model and Behaviour Change Wheel to Develop an Intervention to Improve Hearing-Aid Use in Adult Auditory Rehabilitation**', *International Journal of Audiology* 55. suppl 3: S90–S98, DOI: 10.3109/14992027.2015.1120894 (accessed 6 October 2021)

BEAM Exchange (2018) ***Evidence Map*** (accessed 18 October 2021)

Bhaskar, R. (1997) *A Realist Theory of Science*, London and New York NY: Verso

Campbell, R. (2013) *Feed the Future Learning Agenda Literature Review: Expanded Markets, Value Chains, and Increased Investment*, Rockville MD: Westat

CEE (2012) *Theory-Based Approaches to Evaluation: Concepts and Practices,* Ottawa: Treasury Board of Canada Secretariat Centre of Excellence for Evaluation

Coffey International Development and M4P Hub (2012) *M4P Evaluation Workshop Executive Summary*, Reading: Coffey International Development

Creevey, L.; Dunn, E. and Farmer, E. (2011) *Outreach, Outcomes and Sustainability in Value Chain Projects*, microREPORT 171, Washington DC: United States Agency for International Development

Creevey, L.; Dunn, E.; Northrip, Z.; Snodgrass, D. and Cogan Wares, A. (2010) *Assessing the Effectiveness of Economic Growth Programs: Private Sector Development Impact Assessment Initiative*, Washington DC: United States Agency for International Development

Darnton, A. (2008) ***Practical Guide: An Overview of Behaviour Change Models and Their Uses***, London: Government Social Research Unit, HM Treasury (accessed 10 June 2021)

Delahais, T. and Toulemonde, J. (2012) 'Applying Contribution Analysis: Lessons from Five Years of Practice', *Evaluation* 18.3: 281–93

Giddens, A. (1984) *The Constitution of Society: Outline of the Theory of Structuration*, Cambridge: Polity Press

Gould, G.S. *et al.* (2017) '**Designing an Implementation Intervention with the Behaviour Change Wheel for Health Provider Smoking Cessation Care for Australian Indigenous Pregnant Women**', *Implementation Science* 12.114, DOI: 10.1186/s13012-017-0645-1

Humphrey, J.; Spratt, S.; Thorpe, J. and Henson, S. (2014) ***Understanding and Enhancing the Role of Business in International Development: A Conceptual Framework and Agenda for Research***, IDS Working Paper 440, Brighton: Institute of Development Studies (accessed 6 October 2021)

ICAI (2014) ***DFID's Private Sector Development Work***, Report 35, London: Independent Commission for Aid Impact

Jenal, M. and Liesner, M. (2017) *Causality and Attribution in Market Systems Development,* London: BEAM Exchange

MacAuslan, I.; Sutoris, P.; Findlay, A.; Durieux, M. and Bahri, S. (2018) ***Developing Effective Private Education Nigeria (DEEPEN) Endline Evaluation Volume 1: Synthesis***, Abuja: Edoren (accessed 7 October 2021)

Mayne, J. (2018) '**The COM-B Theory of Change Model**', Working Paper version 4, unpublished (accessed 6 January 2021)

Mayne, J. (2015) '**Useful Theory of Change Models**', *Canadian Journal of Program Evaluation* 30.2: 119–42, DOI: 10.3138/cjpe.230 (accessed 6 October 2021)

Mayne, J. (2012) '**Contribution Analysis: Coming of Age?**', *Evaluation* 18.3: 270–80, DOI: 10.1177/1356389012451663 (accessed 6 October 2021)

Mayne, J. (2008) 'Contribution Analysis: An Approach to Exploring Cause and Effect', *ILAC Brief* 16, Rome: Institutional Learning and Change Initiative

Mayne, J. and Johnson, N. (2015) 'Using Theories of Change in the CGIAR Research Program on Agriculture for Nutrition and Health', *Evaluation* 21.4: 407–28

Mayntz, R. and Scharpf, F. (1995) 'Der Ansatz des akteurzentrierten Institutionalismus', in R. Mayntz and F. Scharpf (eds), *Gesellschaftliche Selbstregelung und Politische Steuerung*, Frankfurt: Campus

Michie, S.; van Stralen, M.M. and West, R. (2011) '**The Behaviour Change Wheel: A New Method for Characterising and Designing Behaviour Change Interventions**', *Implementation Science* 6.42, DOI: 10.1186/1748-5908-6-42 (accessed 6 October 2021)

O'Sullivan, F. (2016) *Impact Evaluations for Market Systems Programmes*, London: BEAM Exchange

Patton, M.Q. (2012) 'A Utilization-Focused Approach to Contribution Analysis', *Evaluation* 18.3: 364–77

Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*, London: SAGE

Scharpf, F.W. (1997) *Games Real Actors Play: Actor-Centered Institutionalism in Policy Research*, Boulder CO: Westview Press

Springfield Centre (2014) *The Operational Guide for the Making Markets Work for the Poor (M4P) Approach*, 2nd ed., Durham: Springfield Centre

Stone, R.; Johnson, S. and Hayes, J. (2010) ***FSD Kenya Impact Assessment Summary Report***, Nairobi: Financial Sector Deepening Kenya (accessed 7 October 2021)

Suntornsut, P. *et al*. (2016) '**Barriers and Recommended Interventions to Prevent Melioidosis in Northeast Thailand:**

A Focus Group Study Using the Behaviour Change Wheel', *PLoS Neglected Tropical Diseases* 10.7: e0004823, DOI: 10.1371/journal.pntd.0004823 (accessed 6 October 2021)

Taylor, B. (2013) *Evidence-Based Policy and Systemic Change: Conflicting Trends?*, Springfield Working Paper 1, Durham: Springfield Centre

Tobing-David, V.E. (2019) ***Final Evaluation: Gender Transformative and Responsible Agribusiness Investments in South-East Asia (GRAISEA) Programme. Period: 1 April 2015–30 June 2018***, Oxfam Evaluation Report, Oxford: Oxfam International (accessed 7 October 2021)

Ton, G. (2021) 'Development Policy and Impact Evaluation: Learning and Accountability in Private Sector Development', in H. Zafarullah and A.S. Huque (eds), *Handbook of Development Policy*, Cheltenham: Edward Elgar

Ton, G.; Vellema, S. and de Ruyter de Wildt, M. (2011) 'Development Impacts of Value Chain Interventions: How to Collect Credible Evidence and Draw Valid Conclusions in Impact Evaluations?', *Journal on Chain and Network Science* 11.1: 69–84

Ton, G.; Thorpe, J.; Egyir, I. and Szyp, C. (2021) ***Value Chain Governance: Entrance Points for Interventions to Address Children's Harmful Work in Agriculture***, ACHA Working Paper 6, Brighton: Action on Children's Harmful Work in African Agriculture, Institute of Development Studies, DOI: 10.19088/ACHA.2021.001 (accessed 18 October 2021)

van Wijk, J.; Zietsma, C.; Dorado, S.; de Bakker, F.G.A. and Martí, I. (2019) 'Social Innovation: Integrating Micro, Meso, and Macro Level Insights from Institutional Theory', *Business and Society* 58.5: 887–918, DOI: 10.1177/0007650318789104

Westhorp, G. (2018) 'Understanding Mechanisms in Realist Evaluation and Research', in N. Emmel, J. Greenhalgh, A. Manzano, M. Monaghan and S. Dalkin (eds), *Doing Realist Research*, London: SAGE

White, H. (2009) '**Theory-Based Impact Evaluation: Principles and Practice**', *Journal of Development Effectiveness* 1.3: 271–84, DOI: 10.1080/19439340903114628 (accessed 6 October 2021)

This page is intentionally left blank

Glossary

3ie International Initiative for Impact Evaluation [India, UK, USA]
AAER Adopt Adapt Expand Respond
ATA Agricultural Transformation Agency [Ethiopia]
BEAM Building Effective and Accessible Markets (Exchange) [UK]
BCF behaviour change framework
BoP bottom-of-the-pyramid / base-of-the-pyramid
BOOST Barley Organization of Supply and Training in South East and Central Oromia [Ethiopia]
CBI Centre for the Promotion of Imports from developing countries [Netherlands]
CDI Centre for Development Impact [UK]
CGE computable general equilibrium
CMO context-mechanism-outcome
COM-B capabilities, opportunities and motivations for behaviour
CREATE Community Revenue Enhancement Through Agricultural Technology Extension [Netherlands]
CSIS Center for Strategic and International Studies [USA]
CSR corporate social responsibility
D&D diagnostic and design
DCED Donor Committee for Enterprise Development [UK]
DEEPEN Developing Effective Private Education Nigeria
DFID Department for International Development [UK]
DGIS Directorate-General for International Development (of the Ministry of Foreign Affairs of the Netherlands)
DGIS-RVO Ministerie van Buitenlandse Zaken-Rijksdienst voor Ondernemend Nederland [Dutch Ministry of Foreign Affairs-Netherlands Enterprise Agency]
EDRI Ethiopian Development Research Institute
EP Enterprise Partners [Ethiopia]
EQuALS Evaluation Quality Assurance and Learning Service [UK]
ESE Erasmus School of Economics [Netherlands]
EUCORD European Cooperative for Rural Development [Belgium]
FCDO Foreign, Commonwealth & Development Office [UK]
FDFA Federal Department of Foreign Affairs
FSDK Financial Sector Deepening Trust Kenya
GAPS Graded Assessment of Private Schools [Nigeria]
GEMS Growth and Employment in States [UK/Nigeria]
GMP good manufacturing practices
GRAISEA Gender Transformative and Responsible Agribusiness Investments in South-East Asia [Oxfam]
HIPSTER Hawassa Industrial Park Sourcing and Training Employees in the Region [Ethiopia]
ICAI Independent Commission for Aid Impact [UK]
ICT information and communication technology
IDS Institute of Development Studies [UK]
IFAD International Fund for Agricultural Development [Italy]

IFC International Finance Corporation
IFPRI International Food Policy Research Institute [USA]
INUS insufficient, non-redundant, unnecessary, sufficient (factor)
IP impact pathway
IOB Dutch Evaluation Office
ISEAL International Social and Environmental Accreditation and Labelling Alliance [UK]
M4C Marker for Change
M4P Making Markets Work for the Poor
M&E monitoring and evaluation
MEL monitoring, evaluation and learning
MoFA Ministry of Foreign Affairs [Netherlands]
MRM monitoring and results measurement
MSD market systems development
MSI multi-stakeholder initiative
MSMEs micro, small and medium-sized enterprises
MSV mobile seed vendor
NMDP Nepal Market Development Programme
PCAF Private Capital Advisory Fund
PEPE Private Enterprise Programme Ethiopia
PrC Partnerships Resource Centre [Netherlands]
PRIME Pioneering Real-time Impact Monitoring and Evaluation [Netherlands]
PSD private sector development
PUM Programma Uitzending Managers [Netherlands Senior Experts]
QuIP Qualitative Impact Protocol
R&A reflect and adapt
R&D research and development
SDC Swiss Agency for Development and Cooperation
SDG Sustainable Development Goal
SHF smallholder farmer
SMEs small and medium-sized enterprises
TBE theory-based evaluation
ToC theory of change
UN United Nations
USAID United States Agency for International Development
WUR Wageningen University and Research [Netherlands]

institute of development studies

IDS Bulletin

Transforming Development Knowledge

Volume 52 (2021)
No. 1 Building a Better World: The Crisis and Opportunity of Covid-19
No. 2 China and International Development: Knowledge, Governance, and Practice

Volume 51 (2020)
No. 1 Gender and Energy: Opportunities for All
No. 1A Fifty Years of Research on Pastoralism and Development
No. 2 Collective Action for Accountability on Sexual Harassment: Global Perspectives

Volume 50 (2019)
No. 1 Exploring Research–Policy Partnerships in International Development
No. 2 The Political Economy of Food
No. 3 Accountability Amidst Fragility, Conflict, and Violence: Learning from Recent Cases
No. 4 The Belt and Road Initiative and the SDGs: Towards Equitable, Sustainable Development

Volume 49 (2018)
No. 1 Value Chains for Nutrition in South Asia: Who Delivers, How, and to Whom?
No.1A Inclusive Peace and Security
No. 2 Accountability for Health Equity: Galvanising a Movement for Universal Health Coverage
No. 3 Emerging Economies and the Changing Dynamics of Development Cooperation
No. 4 The Millennium Villages: Lessons on Evaluating Integrated Rural Development
No. 5 Youth Employment and the Private Sector in Africa

Volume 48 (2017)
No. 1 Sex Education in the Digital Era
No. 2 Interrogating Decentralisation in Africa
No. 3 Africa's Youth Employment Challenge: New Perspectives
No. 4 Courting Catastrophe? Humanitarian Policy and Practice in a Changing Climate
No. 1A Has Universal Development Come of Age?
No. 5–6 Green Power for Africa: Overcoming the Main Constraints

Periodical ID statement The *IDS Bulletin* (ISSN 0265-5012 print) is published in four issues per year. Mailing agent: Gary Edwards, Institute of Development Studies, Library Road, Brighton BN1 9RE, UK. Tel: +44 (0)1273 606261 **idsbulletin@ids.ac.uk**

Postmaster Send all address changes to *IDS Bulletin*, Gary Edwards, Institute of Development Studies, Library Road, Brighton BN1 9RE, UK. Tel: +44 (0)1273 606261 **idsbulletin@ids.ac.uk**

Journal customer services For ordering information, claims and any enquiry concerning your journal subscription please contact Gary Edwards, Institute of Development Studies, Library Road, Brighton BN1 9RE, UK. Tel: +44 (0)1273 606261 **idsbulletin@ids.ac.uk**

Delivery terms and legal title Prices include delivery of print journals to the recipient's address. Delivery terms are Delivery Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Legal title passes to the customer on despatch.

ISSN 0265-5012 (print), 1759-5436 (online)

institute of development studies

Theory-Based Evaluation of Inclusive Business Programmes

Editors **Giel Ton and Sietze Vellema**
Volume 53 | Number 1 | February 2022

IDS Bulletin
Transforming Development Knowledge

'Informative impact evaluations, going beyond a simple box-ticking exercise, are crucial for the learning of stakeholders who are working to develop new inclusive business models in food and agriculture.'

ISSN 0265-5012 (print), 1759-5436 (online) DOI: 10.19088/1968-2022.100