Navigating competing demands in monitoring and evaluation: Five key paradoxes

Marijn Faling (iD)
International Institute of Social Studies, Erasmus University Rotterdam, the Netherlands

Greetje Schouten
Knowledge, Technology and Innovation group, Wageningen University & Research, the Netherlands

Sietze Vellema
Knowledge, Technology and Innovation group, Wageningen University & Research, the Netherlands

Abstract
Evaluation in complex programs assembling multiple actors and combining various interventions faces contradictory requirements. In this article, we take a management perspective to show how to recognize and accommodate these contradictory elements as paradoxes. Through reflective practice we identify five paradoxes, each consisting of two contradicting logics: the paradox of purpose—between accountability and learning; the paradox of position—between autonomy and involvement; the paradox of permeability—between openness and closedness; the paradox of method—between rigor and flexibility; and the paradox of acceptance—between credibility and feasibility. We infer the paradoxes from our work in monitoring and evaluation and action research embedded in 2SCALE, a program working on inclusive agribusiness and food security in a complex environment. The intractable nature of paradoxes means they cannot be permanently resolved. Making productive use of paradoxes most likely raises new contradictions, which merit a continuous acknowledging and accommodating for well-functioning monitoring and evaluation systems.

Corresponding author:
Marijn Faling, International Institute of Social Studies, Erasmus University Rotterdam, Kortenaerkade 12, The Hague 2518 AX, the Netherlands.
Email: faling@iss.nl

Introduction

Monitoring and evaluation (M&E) systems by nature operate at the interface of multiple actors and their organizations—think of implementors or accountability officers—each with their own functions and expectations. In this multifaceted landscape, the organization of M&E systems can be diverse in response to the variety of requirements and demands. This inevitably involves design choices and management decisions, including but not limited to the users and uses the system serves (e.g. Patton, 2008). The most notable contradictory demand in M&E relates to the dual function of monitoring for improvement versus evaluating for accountability (Wongtschowski et al., 2016). Navigating such choices entails addressing multiple demands based on differing logics.

Other contradictory demands relate to logics of low-cost and understandable data collection systems versus elaborate, credible and costly set-ups (e.g. Hirschmann, 2003), or objectivity versus subjectivity in evaluation (Raimondo, 2018; Rodriguez and Acree, 2020). M&E approaches are thus expected to incorporate differing demands and remain legitimate in the eyes of multiple stakeholders. This fosters the presence of contradictory logics and demands, each related to its own set of organizing principles and requirements for M&E systems.

Addressing these contradictory logics and demands is important. One of the main reasons for the failure of M&E efforts can be traced to a lack of understanding regarding the purpose and design of the M&E system (Casley and Kumar, 1987: 1). This may result in the selection of methods that are unable to answer evaluation questions, or a mismatch between program characteristics and M&E design (Stern et al., 2012). Recognizing, acknowledging, and making choices regarding various design options in M&E is important for a well-functioning M&E system.

In this article, we approach the contradictory logics and requirements regarding M&E from a management perspective by interpreting them as paradoxes. Paradoxes are contradictory but interdependent elements inherent in organizations that in isolation seem logical but appear inconsistent when combined (Jarzabkowski et al., 2013; Lewis, 2000; Smith and Lewis, 2011). We operationalize this management perspective to the contradictions intrinsic to evaluating in a complex program, with multiple actors, interventions, and contexts. This has led to the identification of five paradoxes in M&E systems dealing with complexity: the acceptability paradox; the position paradox; the purpose paradox; the systematics paradox; and the engagement paradox. Building on our experiences with designing and implementing an M&E system, we demonstrate the empirical manifestation of paradoxes, and shed light on the practice of dealing with such challenges in M&E. As such, we hope to offer a language to open up the discussion about the accommodation of contrasting demands placed on M&E.

Our conceptualization of the five paradoxes is rooted in action research within 2SCALE. This Dutch flagship development program engages in partnering processes situated in fast-growing agribusiness value chains targeting domestic markets in sub-Saharan Africa. 2SCALE is funded by the Dutch Ministry of Foreign Affairs and aims to build partnerships that connect farmers, buyers, and intermediaries, with the aim to enable them to create and grow new businesses, and at the same time to supply quality products to low-income end users (2SCALE, n.d.). Currently the program is building a portfolio of 60 partnerships in eight countries in sub-Saharan Africa.

We characterize this program as a double-complexity program—a rather complex program, set in a complex and uncertain environment. On the one hand, these programs have complex

characteristics: they operate in multiple countries through differing intervention logics, involving a multiplicity of partners from different sectors and with differing backgrounds and values (Schouten and Vellema, 2019; Stern et al., 2012; Van Tulder and Keen, 2018; Vellema et al., 2020). Flexible intervention processes are consequently required to adapt to changing circumstances (Hummelbrunner and Jones, 2013). On the other hand, the environments these programs operate in are characterized by uncertainty and unpredictability; for example, resulting from violent conflict, political turbulence, or erratic weather patterns. This implies feedback loops, multiple related causes, and lengthy time frames, inter alia (Douthwaite et al., 2017; Mayne and Stern, 2013; Pawson, 2013). The coupled complexity of programs and contexts means that pathways to impact are difficult if not impossible to anticipate, and interventions are designed, steered, and adapted along the way (Ling, 2012; Ripley and Jaccard, 2016).

We argue that in double-complexity programs paradoxes are salient. Programs that face double complexity are messy to understand. M&E approaches within these programs should anticipate outcomes that are unknown, untangle multiple related pathways to impact, and incorporate dynamic contextual contingencies (Douthwaite et al., 2017). At the same time, M&E systems must remain actionable and practicable, while being constrained by available resources. Reflecting on the design and implementation of an M&E system in the context of a double-complexity program enabled us to identify different evaluation paradoxes, which may be relevant to M&E more widely.

In the next section, we elaborate on the notion of paradoxes in management studies and its relevance for M&E. Subsequently, we explain the action research context which the identified paradoxes stem from. Then, we discuss five prevailing paradoxes within M&E by identifying the two contradictory logics that compose each paradox, narrating how these are amplified by double complexity, discussing the possible tensions following from the paradox, and illustrating how each paradox is accommodated in the M&E approach of 2SCALE.

Paradoxes in monitoring and evaluation

A significant part of challenges related to M&E concerns the construction and management of organizational structures and processes, including organizational priority-setting, stakeholder management, and allocation of resources. We therefore consider it worthwhile to approach the navigating of competing demands and logics from a management perspective. This perspective may support the acknowledgment and appreciation of differing demands in the design and management of evaluation practice. We build upon insights from paradoxical thinking, which enables to explicitly acknowledge and accommodate the apparently contradictory logics inherent to evaluation practice. In this section, we use insights from the management literature around paradoxical thinking, to inform the navigation of contradictory logics in evaluation scholarship and practice.

Paradoxes are part and parcel of everyday life. Contradictory yet interdependent logics relating to short- and long-term objectives (Reinecke and Ansari, 2015; Slawinski and Bansal, 2015) and social and economic value creation (Lewis and Smith, 2014; Margolis and Walsh, 2003) inter alia are inherent to organization and management (Smith and Tracey, 2016). Paradoxes refer to contradictory but interdependent elements of organizations that, in isolation, seem logical, but appear to be oppositional and inconsistent in combination, while usually persisting throughout time (Jarzabkowski et al., 2013; Lewis, 2000; Smith and Lewis, 2011). As a result of paradoxes, tensions between contradictory logics may arise.

These tensions may distort relationships and reduce functioning and effectiveness of organizations. The presence and persistence of paradoxes stems from an increasingly interconnected world, which has exacerbated a multitude of complex interwoven systems, each with its own set of goals, functions, and expectations.

Although paradoxes are everywhere, some contexts and situations are more prone to paradoxes than others. M&E of double-complexity programs forms a breeding ground for paradoxes (Jarzabkowski et al., 2013). First, M&E systems by nature operate at the interface of multiple organizations and (sub)systems, including but not limited to the accountability system and the project implementation system. Interorganizational approaches of partnerships give rise to constellations of inherently differing experiences, resources, expertise, and interests (Vangen, 2017). Whereas M&E is required to be neutral and objective, any attempt to evaluation is imbued with human values and subjectivity (Raimondo, 2018; Rodriguez and Acree, 2020). Organizations operate vis-à-vis this multiplicity that places a multitude of simultaneous—and seemingly opposing—demands on them (Lewis and Smith, 2014). The interactions between (sub)systems with their own functions and expectations trigger paradoxes.

Second, M&E approaches to double-complexity programs are even more prone to paradoxes. Given that partnership processes and intervention areas are complex, any attempt to assess change processes in these circumstances is logically to be fraught with difficulties. For instance, many of the issues that development programs attempt to address are referred to as wicked. The inherent nature of these problems—for example, being intractable, not having a stopping rule—complicates the attempt to evaluate impact and makes the evaluation of their resolution problematic (Dentoni et al., 2018; Termeer and Dewulf, 2018). The interrelated causes of development problems constitute the need to capture outcomes in multiple environments with varying logics. These need to be married in an overarching M&E system.

In the context of market-led interventions by public–private partnerships for food and nutrition security, the paradoxes that M&E systems face are therefore likely to be numerous, salient, and persistent (Smith and Lewis, 2011; Smith and Tracey, 2016; Vangen, 2017). Whereas contradictory logics are often viewed as requiring immediate solutions—to be avoided or resolved—a paradox lens approaches them as accommodable. Based on the notion that contradictory logics are persistent and inherent within organizational systems, they should be embraced and incorporated into organizational structures (Smith and Tracey, 2016; Waldman et al., 2019). Paradox theory hence provides a lens to understand the nature, responses, and implications of paradoxes in management. It presupposes there is a benefit in explicitly acknowledging, exploring, and embracing contradictory logics, which leads to creative solutions that accommodate them (Lewis, 2000; Stadtler and Van Wassenhove, 2016; Vangen, 2017). A paradox lens enables us to approach persistent paradoxes in M&E as accommodable.

Designing and managing M&E: The case of 2SCALE

This article is rooted in action research of the Partnerships Resource Center (PrC) with the 2SCALE program, implemented by a coalition of International Fertilizer Development Center (IFDC), SNV, and Bopinc. 2SCALE is a large-scale partnership program that focuses on incubating inclusive agribusiness fostering food and nutrition security. The program engages in partnering processes situated in agribusiness value chains targeting domestic markets in

sub-Saharan Africa. 2SCALE is funded by the Dutch Ministry of Foreign Affairs and aims to build partnerships that stimulate inclusive development by connecting farmers, buyers, and intermediaries, to enable them to create and grow new businesses, while simultaneously supplying quality products to end users, including base-of-the-pyramid (BoP) consumers. 2SCALE is currently building a portfolio of approximately 60 partnerships in eight African countries: Burkina Faso, Mali, Cote d'Ivoire, Ghana, Nigeria, Niger, Ethiopia, and Kenya. This means that the program operates in different country contexts, addressing different root causes of problems, and is characterized by a multiplicity of different types of actors from different organizations and backgrounds. The incubation of inclusive agribusinesses and stimulating the transformation of the ways in which affordable and nutritious food is brought to accessible markets means the program operates in complex and dynamic settings.

2SCALE and its public donor, the Dutch Ministry of Foreign Affairs, tasked the PrC to provide strategic support through action research and knowledge brokering, and to design, manage, and implement the approach to M&E. The action research approach taken in this article was characterized by an interactive inquiry moving between actions in relation to the design, management, and implementation of the M&E approach and conceptualization and analysis of empirical patterns. Through cycles of action, learning in action, and reflection, this approach creates a continuous process of knowledge development as new understandings emerge (McNiff and Whitehead, 2002). This is what Schön (1983) referred to as "reflective practice." In a reflective conversation between the action researcher and the situation he or she confronts, the action researcher engages in sense-making through framing a complex, uncertain, and messy situation. The next step is to address the situation as suggested by the frame. This will in turn create new challenges as the frame does not exactly fit the empirical situation, and henceforth the action researcher engages in reframing. In the consequent cyclic and continuous process of acting, reinterpreting, and reframing, the action researcher's understanding of and approach to empirics becomes increasingly refined (Schön, 1983).

In the development and management of the M&E system, we simultaneously conceptualized and learned about our approach to the challenges that arose in various phases of the process, from contemplating the goal of the system, to developing, operationalizing, and managing the system. During the various reflective discussions, we gradually came to realize that these challenges all pertained to the existence of contradictions among different elements or logics. This realization inspired us to explicate and map these, to acknowledge them and take them into consideration in decision-making. Subsequently we tried to approach these contradictory elements as accommodable. This informed an explicit exploration of combining the different and seemingly incompatible logics into the design of the M&E system. The literature on paradoxical thinking provided us the lens to conceptualize and delineate the logics, which we subsequently framed as paradoxes. Existing academic and gray literature on research and evaluation inspired the refinement of the competing logics underlying each paradox. This approach has helped and continues to help us to refine our understanding of the evaluative challenges when operating in double complexity and informs the management of the evaluation system of 2SCALE.

Paradoxes in the evaluation of double-complexity programs

Our engagement with the design and management of the 2SCALE M&E system led us to identify five paradoxes that are part and parcel of evaluating programs operating in double

Table 1. Paradoxes in M&E of double-complexity programs.

Paradox	What is it about	Contradicting logics
Purpose	How to determine the function of M&E system?	Accountability: M&E to answer to principles' requirements Learning: M&E to inform values, assumptions, and directions underlying program
Position	How to position the evaluator vis-à-vis the program it assesses?	Autonomy: evaluator as autonomous from evaluated Involvement: evaluator as engaged in program under evaluation
Permeability	How to regulate the interference of /interaction with the surrounding environment in the M&E system?	Openness: M&E as "open system," characterized by interactions with environment Closedness: M&E as "closed system," characterized by stability and limited disruption
Method	How to systematize the M&E system?	Rigor: M&E as certain, reliable, and comparable to safeguard efficiency Flexibility: M&E as agile and adjustable to facilitate relevance
Acceptance	How to create an M&E system that is acceptable against various interests?	Credibility: M&E as believable and appropriate Feasibility: M&E as accessible (limited spending and comprehensible design)

Note: M&E = monitoring and evaluation.

complexity. We present the paradoxes following the common logic for establishing an M&E system: We start with the purpose of the system by introducing the paradox of purpose; then move on to roles and responsibilities by presenting the paradox of position and the paradox of permeability; and conclude with paradoxes that relate to design: the paradox of method, and the paradox of acceptance (Table 1). Each paradox stems from two distinct and interrelated logics operating at the same time. These logics seem self-evident in isolation but appear inconsistent when combined. We introduce each paradox by presenting the two competing logics, and their pertinence considering double complexity. We present our strategies for accommodating the paradoxes in the 2SCALE M&E system and highlight the ongoing journey of responding to contradictory logics.

The paradox of purpose—Balancing learning and accountability

The first paradox we identify is the paradox of purpose. This paradox is generally well-known within the evaluation community and refers to contradiction between two classic functions of M&E systems: that of learning and that of accountability (Reinertsen et al., 2022). This contradiction arises from the dual purpose that M&E systems usually serve. On the one hand, M&E systems are supposed to demonstrate a program's successes and achievements for accountability, whereas on the other they serve to identify a program's mistakes to inform learning. This results in the ostensibly opposite requirements of using a system to demonstrate rights versus wrongs.

The first logic is the logic of accountability, which usually refers to organizational mechanisms through which agents answer to their principles (Bovens, 2010; Schoenefeld and Jordan, 2019). Development programs that are funded through public funding (or voluntary donations

or membership fees) must demonstrate their achievements to the donor to provide feedback on whether promises have been kept and how money has been used. This is an "upward" accountability to demonstrate a program's achievements with public funding. Because taxpayers are not aid beneficiaries, there is no feedback loop that would enable receivers of products and services to discipline and control the providers of those products and services. Consequently, reliable information about the results of aid programs is essential to compensate for the remoteness and inaccessibility of aid recipients' experiences (Picciotto, 2018). The effectiveness of aid programs is frequently equated with the achievement of a set of predefined goals and objectives. Considering this, the purpose of the M&E system is to collect data and oversee the achievement of the predefined goals and objectives, through collecting reliable information about the effectiveness of aid programs. The role of M&E in facilitating accountability and overseeing a program's progress toward its predefined targets is especially important in complex and often opaque environments. This environment of complexity, whereby multiple actors are operating, and impacts are to be realized through erratic causal processes involving feedback loops and tipping points, where implementation is realized through multiple different parties and through implementing partners, calls for extra attention to M&E systems that function to enable accountability of programs toward their principals, usually donors.

The second logic is the logic of learning. Learning concerns reflection on vision, strategy, actions, and context to inform readjustments of interventions, and rethink the values and assumptions underlying programs (Guijt, 2011). From that perspective, the purpose of the M&E system is to support practitioners with the right information to enable this reflection. This includes offering regular and timely feedback to program staff to facilitate a continuous development loop and enable adaptive management. The purpose of the M&E system from this logic is thus to support practitioners in program implementation by jointly exploring and unfolding not only successes and achievements, but more importantly also mistakes and errors. The purpose of M&E systems to facilitate learning is considered vital in programs that operate in double complexity means that both programs and the issues they target are unpredictable and ultimately uncontrollable (Ramalingam, 2013; Ramalingam and Jones, 2008). Linear and reductionist approaches will fail to capture actual pathways to change, and consequently need to be replaced with more dynamic, reflexive, and responsive approaches (Archibald et al., 2018). M&E systems as such fulfill the critical function of providing the knowledge and information to inform reflective practice and facilitate adjustments in program implementation along the way.

Determining the purpose of an M&E framework might thus lead to tensions. Whereas the purpose of accountability would lead to M&E aiming to display best results without being open about mistakes, an M&E system with the purpose of facilitating learning would highlight room for improvement, stand by practitioners, in order to facilitate adaptive management.

In the M&E system of 2SCALE, accommodating the paradox of purpose came forward in the flexible use of impact pathways complementary in parallel to monitoring a limited set of Universal Impact Indicators (UIIs) that predominantly served accountability purposes. Each 2SCALE partnership reports on eight UIIs based on a lean measurement of proxies and calculations by the M&E-team (see Section "The paradox of acceptance—combining credibility and feasibility"). The use of impact pathways came from evaluative thinking based on contribution analysis. For each individual partnership a set of two to three impact pathways is developed. Each pathway distinguishes between different types of outcomes typical for contribution analysis: immediate, intermediate, and ultimate outcomes. As such, it envisages

change processes to evolve from changes highly attributable to the actions and activities of the partnerships to changes that also entail action and activities of external influences. For each result level, the partners and the M&E-team jointly identify and collect data on so-called Markers for Change (M4Cs), qualitative and quantitative data points that indicate progress on the different result levels. The use of IPs and M4Cs facilitates a reflexive process among partners, integrating different views on how change happens and enabling the verification of assumed change process. Annually, Reflect & Adapt (R&A) workshops are organized for each partnership. Based on M&E data around impact indicators and M4Cs, partners and stakeholders discuss progress toward objectives, reflect on the direction of the partnership's intended change process, and decide on revisions in the intervention strategy, intertwining the logic of learning in the modus operandi of the partnerships. IPs and M4Cs simultaneously serve accountability purposes by clarifying the plausible contribution of the partnerships to impact as reported on through the UIIs. As such, the format functions simultaneously to provide direction, foster learning, and account for contributions.

However, balancing learning and accountability needs continuous attention. Especially in the beginning phases, 2SCALE's program management team focused largely on getting data on the UIIs for accountability reasons to secure potential additional results-based funding by the donor. This resulted in a rather linear interpretation of how impact would manifest over time and a use of the M&E system to *prove* the effectiveness and impact of the program. This also resulted in lesser attention for documenting change processes using the M4Cs by 2SCALE staff as they were focused on performing toward key performance indicators related to targets. This deflected the program from looking critically at itself to see how, where, and why impact did or did not emerge to *improve* the program. However, to enlarge the potential impact of the program, the latter is of course crucial. Therefore, this paradox requires constant interactions of the M&E-team with program management as well as the donor to stress the importance of learning to improve the program and potentially enlarge its impact, rather than just present the most beneficial image of the program based on brushed-up impact indicator data. A main task for the M&E-team has been to keep this on the agenda and create space for reflection and adaptation.

The paradox of position—Balancing involvement and autonomy

The second identified paradox, the paradox of position, concerns the position of the evaluator vis-à-vis the program it assesses. The paradox emerges because of competing demands regarding the position of the evaluator. On the one hand, the evaluator needs to maintain autonomy vis-à-vis the evaluand to ensure objectivity of the evaluator and resulting reliability of the evaluation. On the other hand, evaluators should be involved and aware of what is going on within a program to facilitate the use of insights for learning purposes.

The first logic of autonomy has traditionally been a key feature of the evaluator's role. To make reliable claims about effectiveness, evaluations need to be considered as "objective" and need to operate autonomously from the program under evaluation (Guenther and Falk, 2007; World Bank Group, 2019). Autonomy prevents evaluators from being subjected to internal dynamics and political pressures of the program they evaluate (Conley-Tyler, 2005). An evaluator that is too closely involved may refrain from being critical, to avoid negative consequences (Mapitsa and Chirau, 2019). External evaluators are traditionally considered more impartial and straightforward in their conclusions and recommendations (Braskamp et al., 1987; Weiss, 1972). In situations of double complexity, where evidence is incomplete

or contradictory, and where different stakeholders may hold different perspectives on the situation at hand (Head and Alford, 2013), autonomy of the evaluator safeguards the incorporation of different viewpoints, to prevent the most powerful from manipulating the process or results, particularly when M&E results are used for adaptive management. Higher levels of trustworthiness are found to correspond with partnerships' eagerness to learn from evaluation findings (Mapitsa and Chirau, 2019). Autonomy of the evaluator is consequently considered important both to retain independence from the program as well as to safeguard trustworthiness of evaluation findings.

The second logic of involvement is increasingly gaining ground. With monitoring and learning becoming more important, the involvement of evaluators has become increasingly appreciated. Involved evaluators tend to have a more nuanced understanding of the priorities and politics of programs, a better view on the usefulness of information, and are better positioned to construct an approach that is appreciated throughout different management layers of programs. Findings by internal evaluators are therefore more likely to be utilized by the program under evaluation (Mapitsa and Chirau, 2019; Shapiro and Blackwell, 1987). Internal evaluators may also smoothen learning processes, as close relations between evaluator and program staff limits resistance against evaluation results (Kniker, 2011). Furthermore, internal evaluators can continuously keep an eye on the follow-up of M&E results, even after the results have been delivered (Mapitsa and Chirau, 2019). Sharing a stake in the success of the program might motivate evaluators to be extra serious about their evaluation task (Chen, 2015). Particularly when M&E systems are related to double-complexity programs, the timeliness of the evaluation is important (Johnson, 1998). Because double complexity means programs must be able to adapt to constantly changing circumstances, M&E results are most useful if conducted in a well-timed and timely manner. As such, they can inform adaptive strategizing by the program. Evaluators must consequently be able to swiftly collect, analyze, and report on available data to the decision-makers of the program, so they can quickly adjust their actions and procedures (Chen, 2015; Mapitsa and Chirau, 2019). All in all, involvement of the evaluator tends to make the M&E process cheaper, faster, and more efficient.

Determining the position of the evaluator in the program under evaluation might thus lead to tensions. On the one hand, the evaluator should remain autonomous to retain independence and safeguard objectivity of findings, whereas on the other hand the evaluator should be closely involved to advance efficiency and usefulness of the M&E endeavor.

To accommodate the paradox of position, the M&E-team in 2SCALE operates to a certain extent autonomously from the program to remain reliable and stay at a distance from program politics. Simultaneously, we aim to align our inputs to the dynamics of the program, to enhance relevance of and receptivity to M&E insights. One example of how we try to balance involvement and autonomy lies in the organization of staff positions. Most of the team is paid by 2SCALE and is positioned in country offices, together with implementing staff. They work in close collaboration with country teams, to have a detailed and nuanced understanding of local partnerships and program dynamics. The other part of the M&E-team is positioned in the Netherlands and is employed by Dutch universities, as independent third parties. While this part of the team works relatively autonomous from the program, there is active engagement with 2SCALE management at program level to understand their needs and facilitate the use of M&E for adaptive management. To further safeguard autonomy, the M&E approach is reviewed by independent experts. When 2SCALE requested the M&E-team to produce so-called "success cases" to be used in promotional efforts, members of the local

M&E-team initially agreed to it. However, after discussion, the part of the team employed outside of the program objected to it, because it would conflict with their integrity as academic researchers. Since this part of the team is not dependent on the program for their salaries, it is easier to set these types of boundaries, and, accordingly, the entire team was able to reject this request.

Balancing autonomy and involvement, however, remains a recurring issue. For example, the active engagement with program management leads to an increased understanding of programmatic logics and internal dynamics. Like discussed in Section "The paradox of purpose—Balancing learning and accountability" at the start of the program, 2SCALE was predominately interested in measuring impact at the level of impact indicators, mostly for reasons of accountability. Active engagement with the program led the M&E-team to dedicate most of their time and resources to deliver insights into the impact indicators. This focus came at the expense of monitoring the processes leading to this impact to understand how impact is actually realized within 2SCALE. Our autonomous position was that impacts cannot be claimed without showing the contribution and necessity of the program at different steps of the impact pathway. However, partnership facilitators concentrated mainly on activities while management focused strongly on impact targets: closing the gap entailed an autonomous effort and independent action research by the team. This shows the persistent nature of the paradox of position, which requires continual attention to accommodate both the logic of involvement and the logic of autonomy.

The paradox of permeability—Accommodating openness and closedness

The third paradox we have identified is the paradox of permeability. It is induced by two contradictory logics regarding the engagement of the environment in the M&E system. On the one hand, the logic of openness dictates that the M&E system benefits from practitioners' engagement in the design of and contributions to the M&E system, to secure relevance of the evaluation. On the other hand, the logic of closedness prescribes that the M&E system need to be isolated from its environment to deliver the required results in an efficient manner.

The logic of openness means that the M&E system should have an open and receptive attitude toward the ideas, suggestions, and contributions of its surroundings. Different opinions and alternative ways of doing things should be acknowledged, and a tolerance for errors incorporated, to demonstrate that the system takes its surrounding environment seriously (INTRAC, 2017). It is based on the principle that to survive, a system must have an appropriate relationship with its environment, characterized by interactions and consequent adaptations. Feedback as such is considered essential for survival. This helps the system to remain relevant vis-à-vis a dynamic environment. Engagement of practitioners in what is to be monitored and through which indicators (data points) helps to enable partnerships to tailor results to maximize usefulness for day-to-day management and validate assumptions on which intervention logic has been based (INTRAC, 2019b). In complex environments when dealing with multiple partners, partnerships need to be able to adapt their intervention logic along the way, to navigate complex environments, learn from experience, and discover new opportunities. As such, they should enable fine-tuning of strategic planning of the partnership. As practitioners are closest to dynamic realities on the ground, they should have a say in the design and execution of the evaluation system.

The second logic of closedness means that the evaluation system has a responsibility to efficiently produce services and products related to the M&E system, including figures and results that indicate insight in the progress of programs. To do so reliably the M&E system needs to act as a relatively closed system. In other words, the evaluation process needs to maintain a certain level of stability and freedom from interruption to deliver on its deliverables (Huey et al., 1990). It is therefore needed that the M&E-team keeps control over the M&E process, maintains an overall responsibility for the evaluation, and is in the position to guide the assessment. Particularly in complex settings, where multiple partners have an opinion regarding the way M&E needs to be conducted, and M&E systems are tasked with a multitude of processes that need to be monitored, some level of control and stability are indispensable to maintain an actionable and efficient M&E system.

In summary, whereas the logic of openness requires the evaluation system act dynamically in interaction with its environment, the logic of closedness dictates that the system isolates itself from external influences, to deliver on its core tasks and responsibilities in an efficient manner.

In 2SCALE, the M&E system started as an open system during the design phase. Program, thematic experts, and M&E engaged in focused discussions around how to formulate the impact indicators, to demonstrate the aggregated contributions of the program to food and nutrition security and inclusive agribusiness. This led to clearly formulated and mutually agreed upon indicators. The indicators were laid down in the document "monitoring and evaluation at partnership level," to be used as a guideline throughout to track the program's contribution to its overall objectives. However, there have been two examples where the program proposed an adjustment of the measurement method of the indicator. For one we opened the system, for the other we remained closed.

The first consideration of permeability was around the impact indicator for smallholder farmers with increased yields and/or net incomes. Initially it was agreed that this indicator would be measured through volumes of produce sourced by the business champion or its commercial partners. However, several partnerships indicated situations where farmers' yields increased, but business champions were unable to source all the produce, for instance, due to a lack of processing capacity. This led to a call from program management to adapt the indicator. They argued that the current indicator would not capture an increase in yields, while the program in practice did increase the volumes to be marketed. It was thereupon decided to adapt the indicator. Our estimation was that if we would keep the original indicator, with program eager to demonstrate results there was a risk that activities would be mainly targeted toward increasing the capacity of the business champion, at the expense of other actors and processes. Therefore, it was decided to adapt the indicator to become an estimate of the produce sold by the smallholder farmer, differentiated for each off-taker.

The second instance of program appealing to the permeability of the system was around the impact of smallholder farmers, micro-entrepreneurs, and small and medium-sized enterprises (SMEs) having improved access to financial services. After repeated engaged discussions between thematic expert and M&E, it was decided to measure this impact using a dual indicator of newly added clients making use of financial services, combined with the total value of financial services offered. The program turned out to provide data mainly on the latter part of the indicator, the total value of financial services offered. As M&E-team, we were not satisfied with this and argued in favor of the program complementing these figures with data on new clients. This led to a strong discussion between program and M&E-team about the way of

measuring, whereby the program pleaded for the total value of financial services to be the core indicator for the impact domain on farmers, entrepreneurs, and SMEs with access to financial services. In this case, M&E decided to not adapt the indicator to the program's wishes. M&E considered the indicator of the value of financial services offered insufficiently representative of target groups with improved access to financial services and feared it could lead to the program focusing only on increasing the value of financial services to the existing customer base at the expense of making these services available to new people. These two examples show that the permeability of the M&E system requires continuous consideration and demonstrate this paradox requires case-by-case judgments based on the consequences of adapted measurements and indicators.

The paradox of method—Combining rigor and flexibility

The paradox of method concerns the contradiction arising from the competing logics regarding the way the M&E system should be designed to deliver optimal results. On the one hand, the M&E system needs to accommodate rigorous systematics to safeguard reliability and enable comparability over time and between projects and contexts. On the other hand, M&E systems need to incorporate flexibility to accommodate a dynamic and non-linear reality. This creates a seemingly contradictory demand from M&E systems to be characterized by a flexible and adjustable design, while simultaneously being rigorous to assure quality and comparability.

The logic of rigor has dominated the M&E discourse for decades and M&E systems generally and traditionally tend to be founded on certainty. This approach is based on a set of assumptions about the world such as the belief that relevant data exists and can be measured, knowledge is unquestioned, problems can be deducted to largely independent and measurable parts, and causality can be considered largely linear and predictable in nature (Archibald et al., 2018). Though increasingly criticized for painting a simplistic picture of reality, this approach serves an important function of assessing and demonstrating a program's results. To shed light on a program's progress, it is essential to collect data points that are comparable, both over time (to measure progress toward impact) and between programs (to be able to compare and aggregate results). This requires that M&E is carried out in a scrupulous and meticulous manner, to safeguard the quality, trustworthiness, and value of insights. As such, a certain level of uniformity and rigor is needed, for instance, through using a shared set of indicators, uniform reporting templates, and joint databases to report on results. Perhaps the best-known manifestation of this approach is the randomized controlled trial (RCT), and the experimental approach to development economics based on counterfactual thinking (see, for instance, work from Nobel laureates Duflo, Kremer and Banerjee (Banerjee et al., 2016). This rigorous systematics enables to collect data on multiple projects at different moments in time, and analyze them at an aggregate level (Simister, 2019). Failing to incorporate a certain level of rigor raises the risk that the M&E system becomes incoherent (Simister, 2019). A rigorous systematics serves as a common point of reference when operating in double complexity. It enables to provide solid aggregable empirical proof about the impact of development interventions. In a situation characterized by double complexity it might be more difficult to align data (due to differing or changing contexts, flexible intervention strategies, and untransparent causal pathways inter alia) and rigorous systematics helps to organize and compare program results.

The second logic of flexibility refers to the need for M&E systems to be able to accommodate agility, authenticity (of context, program, phase), and changing facts and circumstances.

Flexibility of M&E systems is consequently essential to capture the particularities of individual projects in discrete moments in time. If the reality in which programs operate is acknowledged as unpredictable and erratic, M&E systems should accommodate this reality. Progressive insight, evolving organizational requirements, and changing circumstances are highly likely to change the purposes for carrying out M&E. A certain level of flexibility enables programs to continue to report on objectives or indicators that are up-to-date, useful to practitioners and continuously linked to implementation reality. A too rigid M&E system runs the risk of becoming cumbersome, detached from reality, and too bureaucratic (INTRAC, 2019a). M&E systems situated in double complexity thus need to be flexible enough to cope with constant change, while needing to be designed with potential shifting priorities in mind. Rigid M&E systems may fail to capture the emerging patterns of change or learning which would ideally become the renewed focus for M&E, to inspire the steering and adaptive management of programs characterized by double complexity. An example of a method that fits the logic of flexibility is outcome harvesting, which "does not measure progress towards predetermined outcomes or objectives, but rather collects evidence of what has been achieved, and works backward to determine whether and how the project or intervention contributed to the change" (Wilson-Grau and Britt, 2012: 1). Sufficiently flexible M&E systems can be redirected to capture the change processes that both practitioners and M&E are interested in, considering the program's objectives.

Determining the systematics of the M&E system might thus have to adhere to two seemingly competing demands. On the one hand, the M&E system requires rigidity to safeguard reliability and comparability, whereas on the other hand the system needs to be flexible so that it does not wipe out the unique insights and features of each program that are not only worth documenting in themselves but are also essential to inform adaptive management.

The M&E system of 2SCALE is fundamentally flexible, pursuing rigor using general formats and procedures. The impact indicators advance rigor by enabling the aggregated assessment of impact in a manner comparable over time and unit. However, rigor in M&E is not only about assessing the scale of impact, but also about assessing the necessary role of the program in realizing the outcomes. The contribution analysis approach based on IPs and M4Cs (Section "The paradox of purpose—Balancing learning and accountability") fulfills that role by clarifying the plausible contribution of partnerships to impact. The IPs that we use to this end are unique to each partnership and flexible over time, but within certain boundaries. Like displayed in Figure 1, there is a general format for IPs. Within this format, partners are free to identify partnership-specific relevant outcomes. For each outcome level partnerships must regularly report on qualitative and quantitative M4Cs, each unique to and collaboratively identified by the partnership. The M&E system requires all partnerships to reflect annually on their progress and direction, adjusting IPs and M4Cs if necessary. To this end, the M&E-team provides guidance on both the data collection prior to the R&A workshop and the reporting format following the R&A workshop.

The M&E-team is closely involved in guiding the partnerships through thematically informing and steering of sense-making processes to safeguard the program's contribution to public goals. To that end, the M&E has been consequently closely involved in the design and incorporation of various measures to keep inclusiveness on the program's agenda. For instance, although partners have a lot of room to identify their own theory of change and related indicators, the M&E-team instructs partners to report on outcomes and indicators related to the terms of inclusion of small producers and workers (e.g. in decision-making, or distribution of
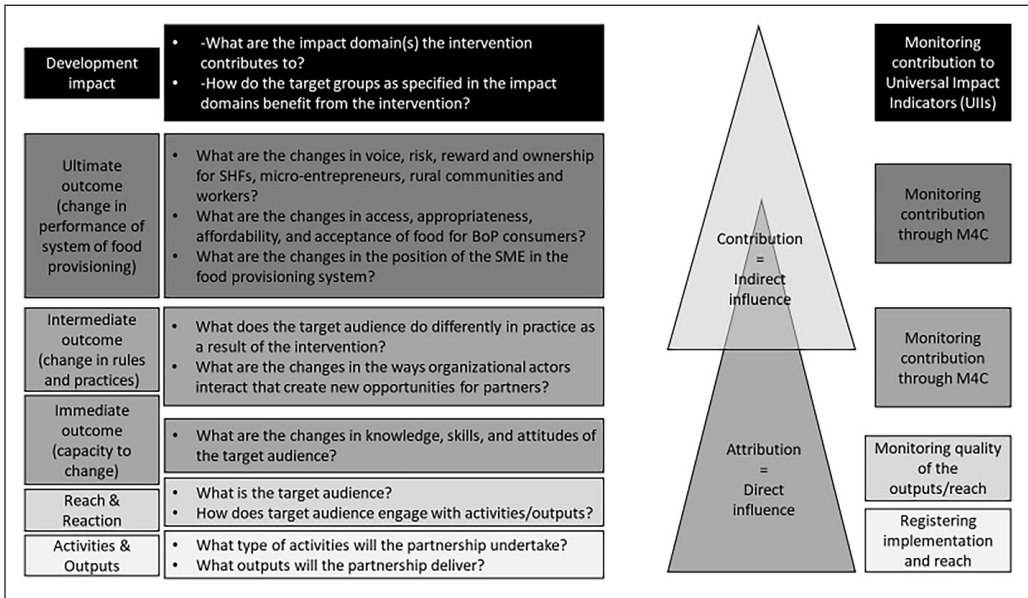
Figure 1. Impact pathways template.

benefits among partners involved) (see also, Vellema et al., 2022). These terms of inclusion are firmly anchored in scholarly literature on inclusive agribusiness in food provisioning.

Although accommodation of this paradox may be fine on paper, reality is more unruly, resulting in vulnerability regarding rigor of the M&E system. Methods providing flexibility are not yet well-established, well-known, and refined. The M&E-team that was largely new to a contribution analysis approach started systematic data collection around partnership-specific combined qualitative and quantitative indicators only in the third year. The M&E-team consequently struggled to collect sufficiently complete evidence on every step in the impact pathway. For example, it was very difficult to describe and systematize what qualitative evidence should look like. This risks the creation of a void in the evidence-based narrative of the contribution analysis that should plausibly link the reported impact to the 2SCALE program. The M&E-team has suggested an external evaluation committee to critically assess contribution claims, functioning as a leverage to ascertain rigor. This committee could select a sample of partnerships and critically assess the provided evidence to see if the contribution claims are sufficiently plausible. This would help safeguard rigor of the approach.

The paradox of acceptance—Combining credibility and feasibility

The paradox of acceptance is a result of contradictions between the logic of credibility and the logic of feasibility. The contradiction exists in the objective of delivering an M&E system that is acceptable from the perspective of different systems, each with their own users and expectations. This leads to the simultaneous striving for and extensive and robust M&E system that requires a vast amount of time, money, and data points; while simultaneously delivering a system that is understandable, lean, and manageable.

The first logic is that of credibility. It refers to the extent to which research—or M&E for that matter—is believable and appropriate (Mills et al., 2010). This relates to whether these represent plausible information on a program and include correct interpretations of existing views (Korstjens and Moser, 2018). Credibility is a necessary condition for M&E systems to be valid from various actor perspectives including partners, donors, and participants in the interventions (King et al., 2013). General research strategies to ensure credibility include prolonged engagement in the field; persistent observation; triangulation and feeding back interpretations and conclusions to participants (Lincoln and Guba, 1985; Tracy, 2010). Some of the strategies for credibility, including cross-checking and validating findings are considered extra important in the context of double complexity. Complex-multi-actor programs might entail different views and perspectives (Stern, 2017). M&E systems must incorporate multiple perspectives, capture complex collaborative dynamics among partners, while untangling complicated causal processes linking interventions and outcomes. The complex and rapidly changing environment makes it difficult to assess whether an outcome is attributable to the partnership and raises the required efforts and resources to produce credible results (Glendinning, 2002). Credible evaluations thus need to have breadth and depth. They need to capture the wide variety of intervention logics of partnerships, while going into depth to make sense of complex, multifaceted, and non-linear pathways of change. M&E systems that credibly capture double complexity are therefore likely to be complex and extensive themselves, requiring numerous resources (Douthwaite et al., 2017). An acceptable M&E system from the perspective of credibility would thus require an extensive and robust set-up.

The second logic, operating simultaneously, is that of feasibility. For an M&E system to be acceptable from the perspective of internal stakeholders, restraint in spending of resources and complexity in design are required. Different types of stakeholders (e.g. program staff and donors) might be critical of spending a large percentage of the budget on evaluation, because resources spent on evaluation cannot be invested in program implementation. Feasibility furthermore relates to the manageability and comprehensibility of the M&E system. M&E systems—especially when operating in double complexity where participation from partners is required—are dependent on the benevolence and engagement of stakeholders. Partners involved must accept the system and be able to identify an interest and logic in the M&E system (Bayer and Waters-Bayer, 2002). This is particularly relevant considering M&E as facilitating learning processes among stakeholders (Kusters et al., 2018). Feasible M&E systems thus need to function as a lean and understandable funnel structure to select and organize relevant data in an efficient manner, and present these in manageable proportions to various audiences, including donors.

Acceptance vis-à-vis different systems thus leads to potential tensions. The seemingly contradictory logics of credibility and feasibility operate simultaneously and place ostensibly opposite demands on M&E systems. While strategies to deliver credible M&E results require extensive resources and a complex set-up, M&E strategies simultaneously need to be feasible, understandable, and cheap. If viewed as mutually exclusive, the combination of these logics may lead to tensions concerning the size of the M&E system: Maintaining credibility would require an extensive, resource-intensive, and complex M&E system, whereas the logic of feasibility would call for a lean, economic, and manageable system.

To accommodate the paradox of acceptance, the 2SCALE M&E system has been designed to be both lean and credible by making smart use of limited resources. A telling example of
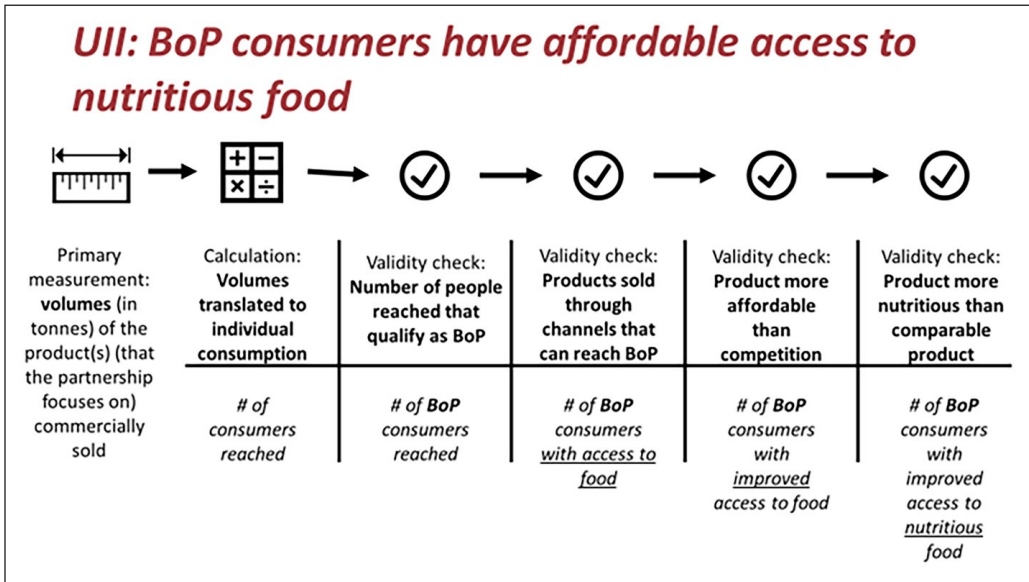
Figure 2. Calculation and validity checks to arrive at Universal Impact Indicator.
Note: UII = Universal Impact Indicators; BoP = base-of-the-pyramid.

this approach is the way in which the aggregated impact of the program is assessed through the UIIs. Instead of relying on complex and resource-intensive approaches to ensure credibility of indicators, the system relies on proxy indicators, smart calculations, and validations based on available data to arrive at the numbers captured by UIIs. To further enhance the reliability of this approach, it has been codeveloped and peer-reviewed by a team of international experts. One of the UIIs focused on enhanced BoP consumers' access to affordable and nutritious food. Directly measuring this for over 60 partnerships in more than eight different countries would require resource-intensive data collection methods such as prolonged and detailed observation combined with in-depth interviews of BoP consumers (Wolfe and Frongillo, 2001). Given the scope and resources of the program this is not feasible. Therefore, the M&E system relies on data that is already available, due to its correspondence to the daily reality and routines of partners. For instance, to measure BoP consumers' improved access to nutritious foods, data are collected on a proxy indicator of volumes of food product commercially sold by the business partner specified for different market channels. These data are usually available in the business partner's administration. To arrive at a credible estimate of BoP consumers with access to nutritious foods, the data on volumes are subjected to a series of calculations and validity checks (Figure 2). First it is calculated how the reported volumes correspond to numbers of consumers reached. Second, additional data inform estimates of the share of BoP consumers out of total consumers reached, and the affordability and nutritional value of the product compared to other (similar) products in the market. These calculations and validity checks are conducted in close collaboration between M&E-team, partners, and the program's thematic experts, to arrive at reliable figures.

   An important assumption underlying this approach related to the feasibility of the M&E system is that business partners have sales figures readily available, and that the interest in a

regular update regarding BoP consumers reached from the M&E system would motivate them to provide these figures. However, in several cases this has not been the case, making it difficult and time-consuming for 2SCALE's partnership facilitators to upload data onto the M&E system at reliable intervals to end up with credible projections of program-level impact. Therefore, in these cases, the M&E-team needs to take over data collection to still ensure reliability. Moreover, while we assumed partnerships facilitators would accept the system based on the ease of use, some of them actually distrusted the calculations and validation checks, as they were afraid the method would underestimate the actual impact made by their partnerships. This shows the continual pursuit to arrive at a system that is able to provide credible insights from different vantage points, while still being feasible to implement.

Conclusion

This article presents a managerial perspective on how to approach the multiple and competing demands placed on M&E. The article reports on a process of reflective practice, and identifies a set of paradoxes encountered in the endeavor to design, implement, and manage an M&E system to navigate the challenges of double-complexity programs. These paradoxes are not detached from one another, but are inherently connected, placing even more demands on M&E systems and teams. The five paradoxes identified in the article become more salient in programs with complex attributes, which operate in complex issue environments. With the signaling of the five paradoxes—the paradox of purpose, the paradox of position, the paradox of permeability, the paradox of method, and the paradox of acceptance—we aim to provide a language to facilitate a conversation about the accommodation of contradicting logics and demands placed on M&E. By pointing out the paradoxes, we offer a heuristic to discuss, navigate, and accommodate competing demands. A paradox approach suggests that principals should look not so much at the technique of an M&E system, but at the composition and ability of M&E-teams and M&E systems to cope with double complexity through a paradox approach.

In our experience with designing and implementing the M&E of 2SCALE, the paradoxes have offered us a lever to find a balance between competing logics and to harness tensions resulting from the contradicting demands. Identifying and accommodating the paradoxes was an integral part of setting up and managing the M&E system of 2SCALE. It has guided us in making well-considered design choices and in thinking through their consequences. Accepting the persistent nature of paradoxes in M&E encouraged us to embrace them and incorporate them into the system. The intractable nature of paradoxes means they cannot be permanently resolved. Their accommodation may raise new challenges and paradoxes therefore merit a continuous acknowledging and accommodating. This is also relevant considering a possible temporal dimension of the contradicting logics within each paradox. This helps to understand the full nature and characteristics of the paradoxes, which facilitates coming up with creative solutions to accommodate seemingly contradictory logics. As such, a paradox approach to the multiple contradictory logics is essential for M&E to be successful.

We conclude that a paradox approach is vital to understand the nature of contradicting demands in M&E, understand the dynamics of competing interests and demands, and find ways of coping that embrace apparent contradictions. In doing so, the approach recognizes the inherently complex nature of M&E and the dynamic approach required to adopt an M&E system that meets multiple demands of the various stakeholders involved. Whereas learning

may be dominant during program implementation, accountability may become relevant toward program closure. However, we must be careful about separating contradicting logics too lightly since they are always present to a greater or lesser extent. We therefore cannot claim to provide a textbook example of a well-balanced M&E system that correctly and uniformly accommodates conflicting demands for each paradox. Follow-up research could explore in more detail the conditions for and ways in which the contradictory demands could be approached. This would support not only to offer a language to discuss paradoxes, but provide an action perspective regarding how to navigate the contradicting demands inherent to the paradoxes.

Author's note

Greetje Schouten is also affiliated to KIT Royal Tropical Institute, the Netherlands.

Declaration of conflicting interests

Funding

ORCID iD

Marijn Faling https://orcid.org/0000-0002-8104-2602

References

2SCALE (n.d.) 2SCALE. Available at: https://www.2scale.org/ (accessed 30 September 2021).

Archibald T, Sharrock G, Buckley J, et al. (2018) Every practitioner a "knowledge worker": Promoting evaluative thinking to enhance learning and adaptive management in international development. *New Directions for Evaluation* 158: 73–91.

Banerjee AV, Duflo E and Kremer M (2016) The influence of randomized controlled trials on development economics research and on development policy. In: Basu K, Rosenblaat D and Sepulveda C (eds) *The State of Economics, The State of the World*. Cambridge: MIT Press, 482–8.

Bayer W and Waters-Bayer A (2002) Participatory monitoring and evaluation (PM&E) with pastoralists: A review of experiences and annotated bibliography. Report, GTZ and ETC Ecoculture, Eschborn, Germany.

Bovens M (2010) Two concepts of accountability: Accountability as a virtue and as a mechanism. *West European Politics* 33(5): 946–67.

Braskamp LA, Brandenburg DC and Ory JC (1987) Lessons about clients' expectations. In: Nowakowski J (ed.) *The Client Perspective on Evaluation: New Directions for Program Evaluation*. San Francisco, CA: Jossey-Bass, 63–74.

Casley DJ and Kumar K (1987) *Project Monitoring and Evaluation in Agriculture*. Baltimore, MD: John Hopkins University Press.

Chen HT (2015) *Practical Program Evaluation: Theory-Driven Evaluation and the Integrated Evaluation Perspective*. Thousand Oaks, CA: SAGE.

Conley-Tyler M (2005) A fundamental choice: Internal or external evaluation? *Evaluation Journal of Australasia* 4(1/2): 3.

Dentoni D, Bitzer V and Schouten G (2018) Harnessing wicked problems in multi-stakeholder partnerships. *Journal of Business Ethics* 150(2): 333–56.

Douthwaite B, Mayne J, McDougall C, et al. (2017) Evaluating complex interventions: A theory-driven realist-informed approach. *Evaluation* 23(3): 294–311.

Glendinning C (2002) Partnerships between health and social services: Developing a framework for evaluation. *Policy & Politics* 30(1): 115–27.

Guenther J and Falk I (2007) The roles of the evaluator as objective observer and active participant. Are they mutually exclusive? In: *AES international conference*, Melbourne, VIC, Australia, 3–7 September. Available at: https://www.aes.asn.au/images/images-old/stories/files/conferences/2007/Papers/John%20Guenther.pdf (accessed 17 November 2023).

Guijt I (2011) Accountability and learning: Exploding the myth of incompatibility between accountability and learning. In: Ubels J, Acquaye-Baddoo NA and Fowler A (eds) *Capacity Development in Practice*. London: Earthscan, 277–92.

Head BW and Alford J (2013) Wicked problems: Implications for public policy and management. *Administration & Society* 47(6): 711–39.

Hirschmann D (2003) Aid dependence, sustainability and technical assistance. Designing a monitoring and evaluation system in Tanzania. *Public Management Review* 5(2): 225–44.

Huey RB, Dunham AE, Overall KL, et al. (1990) Variation in locomotor performance in demographically known populations of the lizard Sceloporus merriami. *Physiological Zoology* 62: 845–72.

Hummelbrunner R and Jones H (2013) A guide for planning and strategy development in the face of complexity. ODI Background Note, March. Available at: https://cdn.odi.org/media/documents/8287.pdf (accessed 19 September 2022).

INTRAC (2017) Participatory evaluation. Available at: https://www.intrac.org/wpcms/wp-content/uploads/2017/01/Participatory-evaluation.pdf (accessed 25 October 2021).

INTRAC (2019a) Complex M&E systems. Available at: https://www.intrac.org/wpcms/wp-content/uploads/2019/03/Praxis-Series-6.-Complex-ME-Systems.pdf (accessed 25 October 2021).

INTRAC (2019b) The supporting environment for M&E. Available at: https://www.intrac.org/wpcms/wp-content/uploads/2019/09/The-Supporting-Environment-for-ME.pdf (accessed 25 October 2021).

Jarzabkowski P, Le JK and Van de Ven AH (2013) Responding to competing strategic demands: How organizing, belonging, and performing paradoxes coevolve. *Strategic Organization* 11(3): 245–80.

Johnson RB (1998) Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning* 21: 93–110.

King J, McKegg K, Oakden J, et al. (2013) Evaluative rubrics: A method for surfacing values and improving the credibility of evaluation. *Journal of Multidisciplinary Evaluation* 9(21): 11–20.

Kniker T (2011) Evaluation survivor: How to outwit, outplay, and outlast as an internal government evaluator. *New Directions for Evaluation* 132: 57–72.

Korstjens I and Moser A (2018) Series; practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice* 24(1): 120–4.

Kusters K, Buck L, De Graaf M, et al. (2018) Participatory planning, monitoring and evaluation of multi-stakeholder platforms in integrated landscape initiatives. *Environmental Management* 62: 170–81.

Lewis MW (2000) Exploring paradox: Toward a more comprehensive guide. *Academy of Management Review* 25: 760–76.

Lewis MW and Smith WK (2014) Paradox as metatheoretical perspective: Sharpening the focus and widening the scope. *The Journal of Applied Behavioral Science* 50(2): 127–49.

Lincoln YS and Guba EG (1985) *Naturalistic Inquiry*. Thousand Oaks, CA: SAGE.

Ling T (2012) Evaluating complex and unfolding interventions in real time. *Evaluation* 18(1): 79–91.

McNiff J and Whitehead J (2002) *Action Research: Principles and Practice*, 2nd edn. London: Routledge Falmer.

Mapitsa CB and Chirau TJ (2019) Institutionalizing the evaluation function: A South African study of impartiality, use and cost. *Evaluation and Program Planning* 75: 38–42.

Margolis JD and Walsh J (2003) Misery loves company: Rethinking social initiatives by business. *Administrative Science Quarterly* 48: 268–305.

Mayne J and Stern E (2013) *Impact evaluation of natural resource research programs: A broader view*. ACIAR Impact Assessment Series Report No. 84. Canberra, ACT, Australia: Australian Centre for International Agriculture Research (ACIAR). Available at: http://aciar.gov.au/publication/ias084 (accessed 19 September 2022).

Mills AJ, Durepos G and Wiebe E (2010) *Encyclopedia of Case Study Research*. London: SAGE.

Patton MQ (2008) *Utilization-Focused Evaluation*, 4th edn. London: SAGE.

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

Picciotto R (2018) Accountability and learning in development evaluation: A commentary on Lauren Cogen's thesis. *Evaluation* 24(3): 363–71.

Raimondo E (2018) The power and dysfunctions of evaluation systems in international organizations. *Evaluation* 24(1): 26–41.

Ramalingam B (2013) *Aid on the Edge of Chaos: Rethinking International Cooperation in a Complex World*. Oxford: Oxford University Press.

Ramalingam B and Jones H (2008) *Exploring the science of complexity. Ideas and implications for development and humanitarian efforts*. Working Paper No. 285. London: Overseas Development Institute.

Reinecke J and Ansari S (2015) When times collide: Temporal brokerage at the intersection of markets and developments. *Academy of Management Journal* 58(2): 618–48.

Reinertsen H, Bjørkdahl K and McNeill D (2022) Accountability versus learning in aid evaluation: A practice-oriented exploration of persistent dilemmas. *Evaluation* 28(3): 356–78.

Ripley M and Jaccard S (2016) *The Science in Adaptive Management*. Geneva: The Lab, International Labour Organization.

Rodriguez S and Acree J (2020) Biopolitical power and paradoxes in evaluation research with transnational migrant youth. *Evaluation* 26(4): 456–73.

Schoenefeld JJ and Jordan AJ (2019) Environmental policy evaluation in the EU: Between learning, accountability, and political opportunities? *Environmental Politics* 28(2): 365–84.

Schön D (1983) *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.

Schouten G and Vellema S (2019) Partnering for inclusive business in food provisioning. *Current Opinion in Environmental Sustainability* 41: 38–42.

Shapiro JZ and Blackwell DL (1987) Large-scale evaluation on a limited budget: The partnership experience. *New Directions for Evaluation* 36: 53–62.

Simister N (2019) *Complex M&E systems: Raising standards, lowering the bar*. Praxis Series Paper No. 6. Oxford: INTRAC.

Slawinski N and Bansal P (2015) Short on time: Intertemporal tensions in business sustainability. *Organizational Science* 26(2): 311–631.

Smith WK and Lewis M (2011) Toward a theory of paradox: A dynamic equilibrium model of organizing. *The Academy of Management Review* 36(2): 381–403.

Smith WK and Tracey P (2016) Institutional complexity and paradox theory: Complementarities of competing demands. *Strategic Organization* 14(4): 455–66.

Stadtler L and Van Wassenhove LN (2016) Coopetition as a paradox: Integrative approaches in a multi-company, cross-sector partnership. *Organization Studies* 37(5): 655–85.

Stern E (2017) Evaluating partnerships. In: Liebenthal A, Feinstein ON and Ingram GK (eds) *Evaluation & Development: The Partnership Dimension* (World Bank Series on Evaluation and Development). London: Routledge, pp. 29–42.

Stern E, Stame N, Mayne J, et al. (2012) *Broadening the range of designs and methods for impact evaluations*. DFID Working Paper 38, April. Available at: https://assets.publishing.service.gov.uk/media/57a08a6740f0b6497400059e/DFIDWorkingPaper38.pdf

Termeer CJAM and Dewulf A (2018) A small wins framework to overcome the evaluation paradox of wicked problems. *Policy and Society* 38(2): 298–314.

Tracy SJ (2010) Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry* 16: 837–51.

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: 315–32.

Vangen S (2017) Developing practice-oriented theory on collaboration: A paradox lens. *Public Administration Review* 77(2): 263–72.

Vellema S, Schouten G and Faling M (2022) Monitoring systemic change in inclusive agribusiness. *IDS Bulletin* 53(1): 103–122.

Vellema S, Schouten G and Van Tulder R (2020) Partnering capacities for inclusive development in food provisioning. *Development Policy Review* 38(6): 710–27.

Waldman DA, Putnam LP, Miron-Spektor E, et al. (2019) The role of paradox theory in decision making and management research. *Organizational Behavior and Human Decision Processes* 155: 1–6.

Weiss CH (1972) *Evaluation Research: Methods for Assessing Program Effectiveness*. Englewood Cliffs, NJ: Prentice Hall.

Wilson-Grau R and Britt H (2012) *Outcome Harvesting*. Cairo, Egypt: Ford Foundation.

Wolfe WS and Frongillo EA (2001) Building household food-security measurement tools from the ground up. *Food and Nutrition Bulletin* 22(1): 5–12.

Wongtschowski M, Oonk L and Mur R (2016) *Monitoring and evaluation for accountability and learning*. KIT Working Papers No. 3. Amsterdam: Royal Tropical Institute.

World Bank Group (2019) *World Bank Group Evaluation Principles*. Washington, DC: International Bank for Reconstruction and Development/The World Bank. Available at: https://ieg.worldbankgroup.org/sites/default/files/Data/reports/WorldBankEvaluationPrinciples.pdf (accessed 25 October 2021).

Marijn Faling is Assistant Professor at the International Institute of Social Studies (ISS), Erasmus University Rotterdam. She researches collaborative change processes regarding food security, climate change, and inclusive development.

Greetje Schouten PhD is a Senior Advisor inclusive value chains and responsible business conduct at KIT Royal Tropical Institute. She has a part-time research affiliation at Wageningen University.

Sietze Vellema is Associate Professor at the Knowledge, Technology and Innovation group, Wageningen University, Netherlands. His evaluation practice and action research centers on food security and inclusive agribusiness in Africa.