Journal of MultiDisciplinary Evaluation
Volume 21, Issue 49, 2025

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
http://www.jmde.com

Complexity-Aware Evaluation for Learning: A Case Study of a Developmental Approach

Eureta Rosenberg
Rhodes University

Karen Kotschy
Association for Water and Rural Development (AWARD)

Sharon Pollard
Association for Water and Rural Development (AWARD)

Background: Ongoing learning is vital for steering programs in complex social-ecological systems, because it enables implementers to change course when needed, and ideally contributes to system-wide insights as well. While there are many calls and guidelines for complexity-aware and learning-centered monitoring and evaluation (M&E), there are fewer case examples of attempts to implement it in practice.

Purpose: This paper reflects on a multi-year case of an attempt to design and implement a complexity-aware, developmental and learning-centered M&E framework in a complex social-ecological landscape, highlighting challenges, adaptations, and evidence of learning.

Setting: Olifants River Basin, northeastern South Africa and southern Mozambique.

Intervention: Resilience in the Limpopo (Olifants Basin) program (RESILIM-O).

Research Design: Participatory action-reflection case study.

Data Collection and Analysis: Documentation of the monitoring and evaluation framework development and implementation, as well as questionnaires and focus groups probing implementers' learning, generated a case record that was analysed qualitatively.

Findings: Learning among program implementers was facilitated by working in new ways with standard M&E elements, including indicator-based targets and theories of change, and through multiple opportunities for shared reflection, including novel reporting templates, reflection days, and collaborative case studies. Challenges notwithstanding, the participatory, developmental M&E approach built understanding of and competence in M&E in the organization. The hybrid framework seemed to successfully combine the need for accountability with the desire for organizational learning, and associated features were subsequently adopted elsewhere.

Keywords: *complexity-aware evaluation; developmental evaluation; hybrid design; organizational learning; reflective practice.*

Features of complex systems are well documented (Preiser et al., 2018) and include high levels of interconnectedness, "radical" openness, profound contextual influences, dynamism, adaptive capacity, emergence, complex causality, and nonlinear pathways of change. These features have implications for the way in which programmatic interventions in these systems are evaluated. As a result, much has been written about the need for monitoring and evaluation (M&E) designs to be sensitive to the complexity of program contexts and of nonlinear pathways to change (e.g., Bellamy et al., 2001; Patton, 2008, 2010; Funnell & Rogers, 2011; Pringle, 2011; Douthwaite & Hoffecker, 2017; Hertz et al., 2021; USAID, 2021).

This paper focuses on one particular implication, which is the need for learning. Ongoing learning is vital to steer programs in complex systems, because it enables implementers to progressively set a course of action and to change course when needed (Woodhill, 2007; Pollard et al., 2011; Roux et al., 2017). In complex contexts even the best program designs cannot simply be "rolled out," but must be experimented with, reflected on, and—where necessary—adapted (Patton, 2008, 2010; Funnell & Rogers, 2011). M&E is an obvious mechanism for supporting—as well as capturing and sharing—the associated learning.

Several authors have published guidelines for learning-centered or learning-centric M&E (e.g., Bakewell, 2003; Woodhill 2007; Pringle et al., 2011; Villaneuva, 2011). Authors across different fields have also highlighted challenges with implementing this type of M&E (e.g., Wongtschowski et al., 2016; Chapman et al., 2016; Khan et al., 2018; Chirau & Blaser-Mapitsa, 2020).

This paper shares a case example of learning-centered M&E with a view to further illuminate challenges and also ways in which those challenges could be addressed. The case is based on a critical analysis of a complexity-aware program evaluation design that has been implemented and refined over a 6-year period, well documented, and subsequently taken up in other contexts. It presents an opportunity to explore implementer responses to known and emerging challenges, and probes the extent to which evaluation design features can support learning in organizational contexts. A body of such critical and exploratory implementation case analyses puts the field in a good position to deepen and extend evaluation theory and practice for the increasingly complex nature and contexts of program interventions, where ongoing learning is vital.

The RESILIM-O Program as an Evaluation Case Study

The RESILIM-O program was implemented in the Olifants River Basin in northeastern South Africa and southern Mozambique from 2013 to 2020, by the Association for Water and Rural Development (AWARD). AWARD is a South African nonprofit organization involved in social-ecological research and development practice. The RESILIM-O program presented as a valuable case study, because it was explicitly responding to complexity and was implemented over several years and at a significant scale, consisting of 23 complementary projects addressing a range of development and environmental issues. Its aim was to reduce rural communities' vulnerability to climate change through improved transboundary water and biodiversity governance and management of the Olifants Basin, while enhancing the resilience of people and ecosystems through systemic and social learning approaches (AWARD, 2020).

AWARD approached the Olifants Basin as a complex social-ecological landscape featuring livelihoods in agriculture, mining, and wildlife tourism; layers of overlapping and often faltering government; both pristine and degraded biodiversity; and rural communities affected by climate-related water shortages. Program interventions were guided by complex systems theory and resilience and social learning principles (Walker & Salt, 2006; Pollard et al., 2014). As a case study of complexity-aware evaluation, RESILIM-O's program theory (Pawson & Tilley, 1997; Weiss, 1997) was also significant, because it involved multiple, nonlinear pathways for strengthening capacity, agency, and governance among institutional role-players and intended beneficiaries (AWARD, 2020).

The funder also expressed a willingness to experiment with complexity-aware evaluation design. This, together with AWARD's interest in developing an M&E framework that was aligned with RESILIM-O's complexity-aware program design, meant that an explicit experiment was included in the program design, and evaluation-related learning data was collected throughout. AWARD intentionally set out to implement and refine a hybrid monitoring, evaluation, reporting and learning (MERL) framework able to meet the need for both accountability and learning in complex contexts and programs.

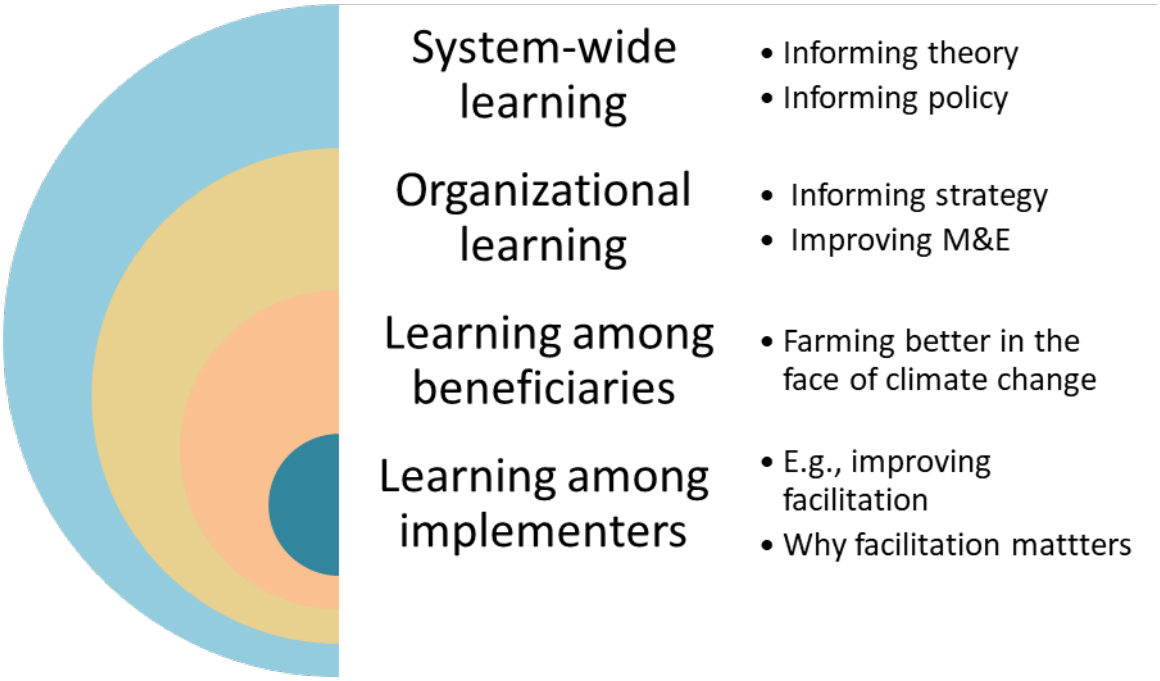
In summary, RESILIM-O presents a good case study because …

- It was a multi-year program, implemented at significant scale.
- The program was overtly approached as a complex intervention in a complex context, addressing a complex set of social-ecological challenges.
- The program implementer was interested in evaluation innovation.
- The funder was willing to experiment with a hybrid evaluation design that met both accountability and learning needs, and as a result,
- Data on the M&E framework and its implementation was collected from inception, and data on learning during the period 2017–2019.

Focus on Learning

While the concept of learning in relation to complex adaptive systems has been explored elsewhere in some depth (e.g., Roux et al., 2017), in this paper it will only be noted that such learning involves more than information sharing or sense-making. The kind of learning required for sustainable rural development and resilience-building is potentially transformative and at least generative of new action, either in theory or in practice (Engeström & Sannino, 2012; Lotz-Sisitka et al., 2016; Chikunda, 2017). In Figure 1 we outline four dimensions of learning relevant to complexity-aware M&E.

Figure 1. Connected Dimensions of Learning Relevant to Evaluation of Social-Ecological Resilience-Building Programs



Firstly, and most relevant to this particular paper, consider the *learning of the program implementers*. In complex contexts there are no blueprints for action, and implementers need to learn all the time in relation to the context and the outcomes of their actions, through action learning and reflective practice (Schön, 1983; Scharmer, 2009). Many competencies required for sustainability work, such as facilitation, are learned on the job and in teams, building on formal training (Rosenberg et al., 2017).

A second dimension of learning involves the learning among intended *program beneficiaries*. Although a vital intended program outcome, this is not addressed in this paper.

The third dimension of learning, *organizational learning*, is closely connected to the first, as it is an aggregate of the learning of individuals, that becomes embedded in social units (Reed et al., 2010). Arising from the interactions between individuals, organizational resources, processes and information, across various

boundaries, organizational learning results in new understandings and practices (Boreham & Morgan, 2004; Engeström, 2016). The concept of strategic adaptive management (Pollard et al., 2011) requires learning by individuals and organizations alike (Cundill & Fabricius, 2009; Williams & Brown, 2018).

The fourth dimension is *system-wide learning* across program role-players, across programs and organizations, and across sociogeographical contexts. This form of social learning and knowledge-building advances theory and policy on regional or international scales (Ison et al., 2021).

These dimensions of learning are all desirable when programs are implemented in complex contexts. However, they are not always supported by M&E practices. Monitoring and evaluation need to both *support* and *capture* learning in programs implemented in complex systems. M&E may fail in this regard for a number of reasons.

Why M&E Practices Can Fail to Optimally Support Learning in Complex Systems

What follows is a range of challenges associated with M&E frameworks and practices, as identified in the literature.

Theories of Change That Ignore Complexity Do Not Prioritize Learning. Mismatches between program logic models and program design and context often inhibit learning. In complexity, the pathways to success need to be worked out during and through action-taking and reflection. If the program's theory of change (ToC) or logic model does not include learning feedback loops or the possibility of emergent outcomes, it is unlikely that resources will be allocated toward facilitating and understanding these (Woodhill, 2007; Villaneuva et al., 2011).

The Accountability Agenda Crowds Out Learning. A heavy donor accountability agenda can overshadow, inhibit, or fail to support learning. Donor monitoring and reporting requirements are often extensive and consume precious resources, which, in smaller implementing agencies in particular, can leave little space for evaluation or learning-focused activities (Bornstein, 2006; Taylor & Soal, 2011; Mueller-Hirth, 2012; Kachur et al., 2016). Performance-based systems may support accountability and transparency, but also provide incentives to hide failures and overstate successes (Wongtschowski et al., 2016; Mushwana Mudau, 2020). This stifles reflection on disruptions, dissonances, and contradictions, which is vital for learning (Wals, 2007; Schulz, 2010).

Evaluation Is Separated from Monitoring. Conventionally, monitoring (the routine collection of data) is the task of program implementers, while evaluation (sense-making based on the monitoring data) is undertaken by external experts, midway through and at the end of a program. Woodhill (2007) argued that this tends to exclude the implementers from the sense-making that could precede learning and improved practice. For a variety of reasons, practitioners do not benefit optimally from reading someone else's evaluation report, and if reports are only produced at program closure, they are not helpful for adaptive management.

Quantitative Indicators Are Over-Emphasized. Another standard M&E practice is to report almost exclusively on quantitative indicators. While this enables the aggregation of outcome data on national, regional, and global scales, when M&E is *exclusively* about progress against quantitative indicators, the likelihood of learning is greatly reduced. Programs that are the most transformational are often the least easily measured with quantitative indicators (Natsios, 2010).

M&E Is Seen as a Purely Technical Function. M&E is often seen as a technical function related to monitoring and reporting systems, indicators, and data storage. Wongtschowski et al. (2016) reasoned that such a technical approach fails to harness the power of M&E for building and supporting meaningful partnerships, promoting learning, and building capacity. Furthermore, M&E work, particularly in complex contexts, requires strategic leadership and advocacy and not merely technical management (Rosenberg & Kotschy, 2020; Patton, 2021).

Learning Is Treated as Equivalent to Knowledge Transfer. Learning is often equated to transfer of knowledge during training or through communications products (Woodhill, 2007), being focused primarily on "capturing lessons" rather than on the *process* of learning. Capturing lessons is often left until the end of an intervention and seen as a once-off communication task, with the product being a brochure, guideline, or one-time learning event. As noted by Patton (2021), this is insufficient for tracking, documenting, and interpreting innovations as they unfold in complex, dynamic contexts.

M&E Is Not Functionally Integrated with Planning and Decision-Making Processes. While many development interventions appear to have sufficient monitoring to manage the operational basics of implementation and financial management, strategic adaptive management requires ongoing connections between M&E functions and strategic planning processes (Pollard et al., 2011). If managers perceive M&E as simply number counting and bland reporting, they may not engage closely with it—creating a self-fulfilling prophecy where the lack of attention renders M&E ineffective in terms of decision-making (Woodhill, 2007).

This case study aims to investigate how the RESILIM-O implementers and their M&E team aimed to overcome these common limitations of standard M&E practices; the challenges they encountered; and the extent to which the MERL framework developed has supported learning, both in AWARD and in other programs, organizations, and contexts.

Case Study Research Method

The research on which this paper is based was praxiological, participatory, and reflexive, and addressed the following guiding research questions:

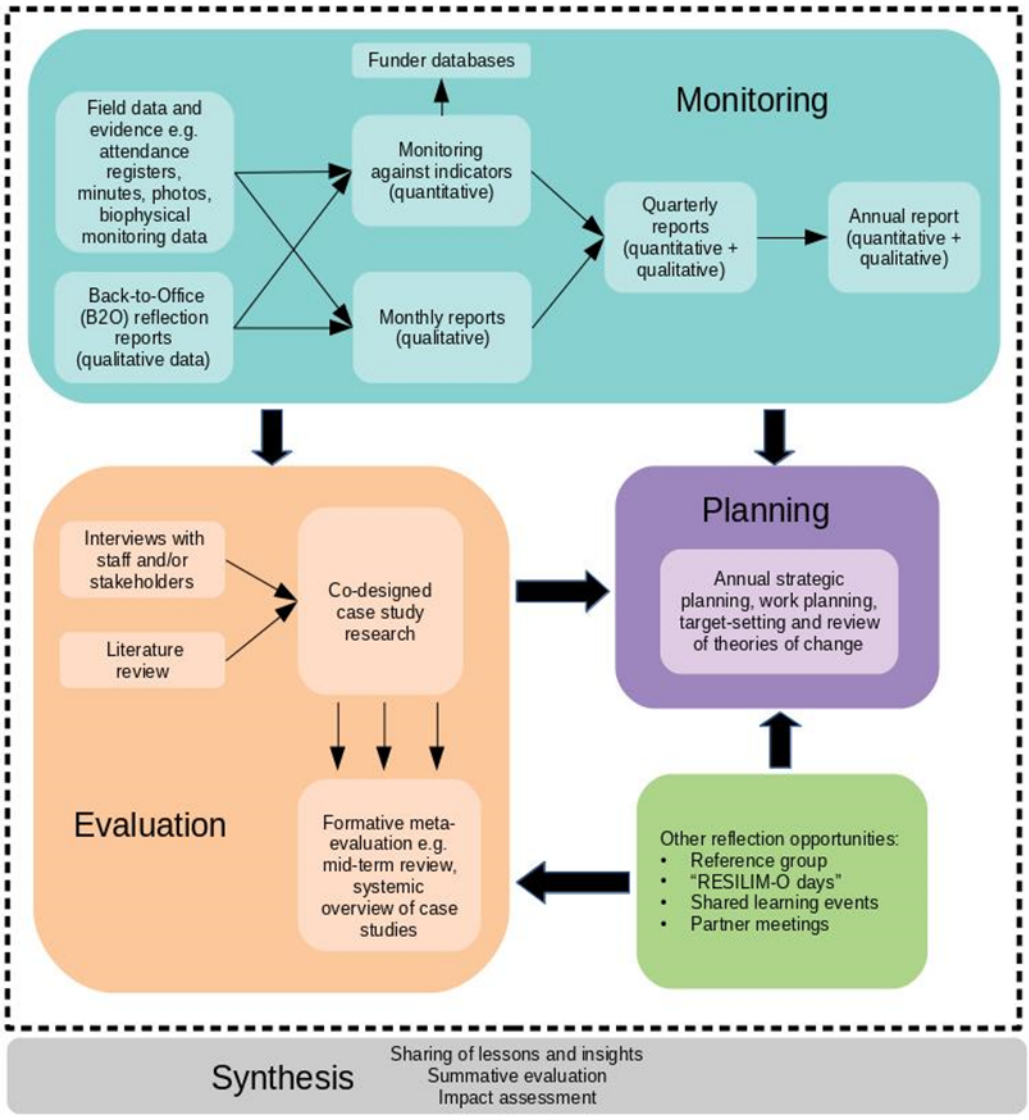
1. What are the design features of the RESILIM-O MERL framework that are explicitly intended to support learning, and how do these evolve over the life of the program?
2. What learning, if any, do implementers associate with these MERL features?

The research involved a review of the data on the evolving M&E design and implementation which was documented over the lifespan of the program, and data regarding learning collected from implementers in the latter half of program implementation.

Features of the RESILIM-O MERL System

An overview of the MERL system as it existed toward the end of implementation is provided in Figure 1, with the system elements described further in Rosenberg et al. (2017). This system was developed and refined over time, and the development process itself was documented.

Figure 2. Elements of the RESILIM-O MERL Framework



Review of the Elements of the MERL System

The elements (or tools) of the MERL system are listed in Column 1 of Table 1, with Figure 1 indicating how these elements relate to each other. The middle column of Table 1 shows the number of documents produced, as an indication of the scale of the MERL system and the body of work that was available for this research. The third column shows the number of revisions made to elements, an indication of necessary refinement of the system over time.

Table 1. Elements of the RESILIM-O MERL System with Revisions Between 2014 and 2020

MERL system element	No. produced	No. of revisions
MERL framework document: Describes the MERL design to guide the implementers, funder, and MERL team. Used for making decisions about monitoring, evaluation, reporting, and learning. Reviewed annually and updated when necessary.	3	3, but core principles remained unchanged
Back-to-Office (B2O) reports: A monitoring tool used as a first layer of reporting and reflecting on field activities.	657	6 (template)
Monthly reports: Reporting and reflection by staff. Main source of narrative data. Compiled by each project within the program. Revisions aimed to enhance reflection and learning in a nonstandard manner.	330	3 (template)
Quarterly reports: Standard donor-required reports. Compiled from monthly reports and quantitative data pertinent to chosen indicators, as well as reflections by program implementers and MERL team.	20	3 (template)
Annual reports: Similar to quarterly reports, with greater focus on progress and achievements over the year. Featured both narrative and indicator-based content; informed annual strategic planning.	8	1 (structure and style)
Presentations on MERL guidelines and implementation requirements: Shared with implementers and their external partners. Required MERL team to record and periodically consolidate design decisions and template versions.	4	4
MERL team meeting minutes: Provide a record of the history of the MERL system, issues that arose, and how these influenced the design decisions taken.	66	NA
Case study and meta evaluation reports and presentations: Provided formative evaluation of particular projects or aspects of program implementation, with meta-evaluation reports reflecting on the implications and value of the case studies for the program as a whole.	15	1 (methodological, to strengthen analysis)
Logic models and theories of change: RESILIM-O had an overarching ToC, and each project compiled its own (sub) ToC, with reference to the high-level program theory.	24	At least once annually at project level
Indicators and indicator protocols: Set of indicators chosen for reporting (program level) and project management (project level), with protocols for defining and measuring progress. Indicators selected from a prescribed but extensive list provided by USAID.	2	2 at program level; annually at project level
Databases: Indicator related data recorded and entered monthly into USAID's global databases; reported quarterly against project targets.	2	2 (template)
Reference group meetings: The reference group, which met twice a year, comprised external specialists and AWARD executives. A space to reflect on praxis and global experiences relevant to RESILIM-O at a 'meta-level'.	7	NA
Steering Committee/partners meetings, RESILIM-O days, and shared learning events: Meetings for sharing work aspects and insights internally and externally and gaining feedback.	70	NA

Questionnaires and Focus Group

To assess the extent to which the MERL system supported learning, the MERL team administered two questionnaires via email: one in 2017 with AWARD's internal implementer staff and one in 2018 with external implementing partners. The MERL team then held a focus group discussion in 2018 with internal implementers (approximately 30 staff). Implementers were asked to reflect on the

features of the MERL framework and whether these supported individual and/or organizational learning. A clear distinction between these two forms of learning was not made.

Collective Review of This Information and Joint Preparation of the Paper

To prepare this paper, the MERL team conducted a detailed historical overview of the evolution of the MERL system. To understand which insights would be most useful for other program and M&E designers, the interim findings were presented at two international and two national gatherings.

Positionality and Validity Considerations

We, the authors, were MERL specialist, MERL manager, and program director for this program, and at various stages we also led action-reflection research activities. We shared interim findings with implementers for member checking purposes, and presented findings annually to RESILIM-O's Reference Group, who were not involved in implementation, and whose explicit role was to provide critical feedback. This—together with implementer feedback, which was often unsolicited, and frequently critical, and reported by the MERL officer—served to guard against an over-subjective attachment by the MERL team to the evolving MERL process. The funder's representative served on the Reference Group but was not involved in the research processes.

Results

The results of the research into the RESILIM-O MERL system, and what it enabled, are presented in three parts:

- Standard M&E elements used in nonstandard ways
- Nonstandard MERL elements developed in the program
- Learning reported by program implementers

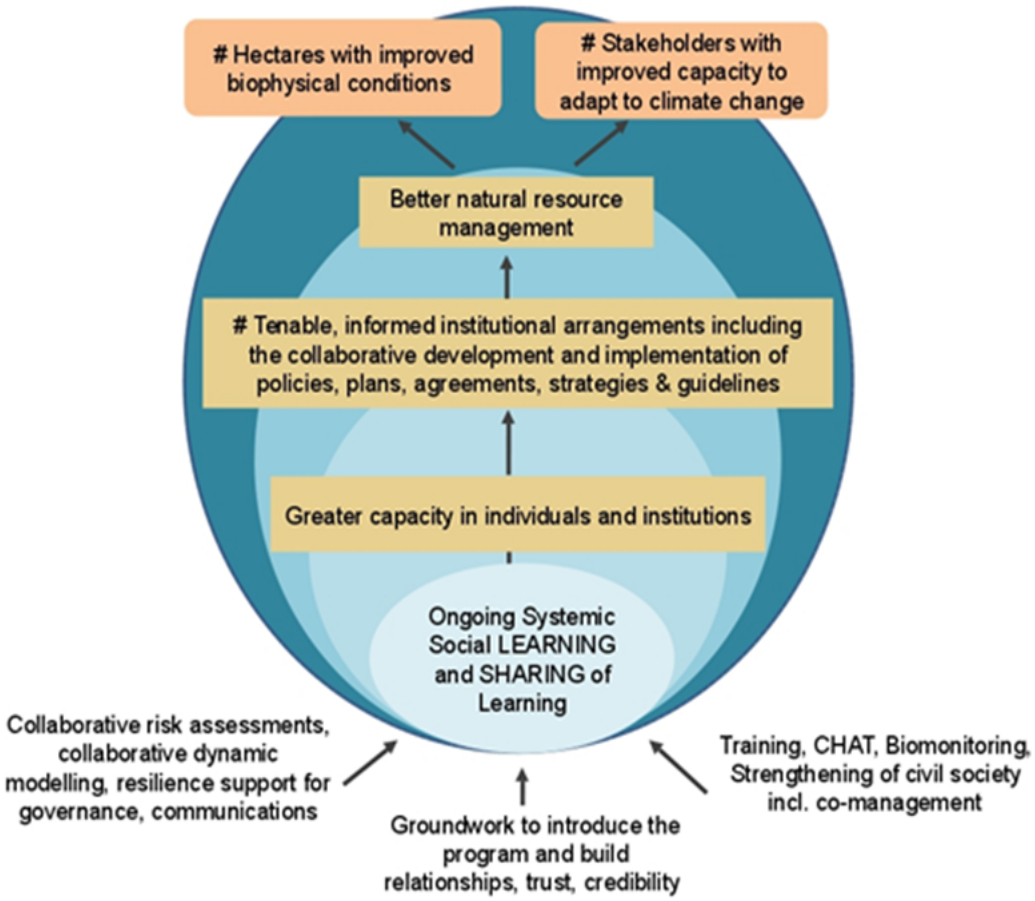
Standard M&E Elements Used in Nonstandard Ways

Quantitative indicators of progress or success, targets against these indicators, and baselines from which to measure progress or impact, are standard elements of program M&E. AWARD selected indicators from the funder's extensive indicator lists that best matched the RESILIM-O program's intended impact. Examples were "Number of stakeholders with increased capacity to adapt to the impacts of climate change" and "Number of hectares of biologically significant areas showing improved biophysical conditions."

Implementers reported that establishing baselines and targets was initially difficult. They used 2 years to research the Limpopo-Olifants Basin and engage stakeholders before setting targets, and even then these were regarded as provisional.

Quantitative indicators taken out of context can be reductionist as they do not readily allow the tracking of multiple and emergent (unexpected) pathways to change. The RESILIM-O MERL system responded to these limitations in two ways. The first was to connect up the selected high-level indicators using a ToC, so as to reflect the program's systemic approach (Figure 3). The ToC positioned systemic, social learning as foundational, both a mechanism for change and an outcome. The concentric circles depict the nonlinear way in which system-wide social learning was expected to spiral out from all program activities.

Figure 3. RESILIM-O's high-level theory of change showing links between indicators and nonlinear learning (blue circles).



These MERL processes comprised nonstandard ways of working with standard indicators, targets, and ToCs, in that they were designed to facilitate ongoing learning and refinement, in addition to being accountable to the funder.

Quarterly and annual reporting is another standard feature of M&E systems. Donors often allow (or require) reports to combine quantitative indicators with a complementary narrative component. The RESILIM-O MERL system was intended to include a strong narrative element to complement and illuminate quantitative data. This meant that RESILIM-O also had to gather narrative, reflective, and evaluative data. These were the less standard features of the MERL framework, outlined next.

Nonstandard M&E Elements Developed in the Program

One of the ways in which AWARD experimented with its MERL system was exploring how to gather and process the narrative or qualitative data in ways that were practical, feasible, and learning-centered. In the first year of implementation, a process narrator was employed to collect narrative data. Focusing this work (what to record and what not), however, proved difficult in the early stages of the program, and implementers struggled to make time for it, particularly as it was still unfocused. Over time, the MERL team, with the support of the program director, developed and introduced several ways of gathering narrative, reflective, and evaluative data. These are described next.

Innovations in Reporting. Project implementation leads were asked to set aside half a day, once a month, to produce monthly reports. This is a

common enough practice, and was also resisted by busy implementers, but over time the evaluative dimension of these reports became more successfully guided through a template that encouraged team-based reflection on the activities, progress, and challenges of the month. While this was not at first a popular addition to the workload, implementers started to find these reports valuable (see the Learning Reported by Implementers section below).

Another reporting innovation was the introduction of a back-to-office (B2O) report. This was a derivative of a simple field report and captured monitoring data for implementation-related events such as field visits, workshops, and beneficiary training. The reports were to be completed soon after the event, in teams wherever possible (often during the return journey). The template required implementers to reflect on the event and document observations and any learning.

These reporting innovations required several revisions of the templates (Table 1) and support from the program executives. The director noted, "It took time and conscious effort to build a culture and habit of reflection" to assist staff to move from a description of the activities and outcomes to also reflecting on the *meaning* of these (i.e., sense-making for learning). Teaming up more experienced practitioners with less experienced team members helped to, over time, deepen reflections on what happened in the field.

Other Opportunities for Reflection. The director introduced RESILIM-O days—monthly events where at least some time was set aside to "pause and reflect." Whilst monthly meetings are common in large implementation teams, they are often limited to descriptions of progress, where people listen passively until it is their turn to report. At RESILIM-O days the purpose of sharing and learning was emphasized, and implementers were encouraged to be critically reflective about their own work and engage with each other's. Conscious efforts were needed to shift reflections beyond operational matters.

In 2018 shared learning events were introduced as another reflective space. They extended participation beyond program staff to external implementation partners and beneficiaries. The focus group referred to in the Case Study Research Method section above took place during such a shared learning event dedicated to reflecting on the role of MERL tools and processes in learning.

Another internal learning opportunity introduced toward the end of the program was a Month in Pictures section in the monthly meeting, where project implementers presented some thoughtfully selected photographs of their work in the previous month. This developed into an effective way of making meetings more engaging, and served multiple purposes: a chance to update fellow implementers, to showcase and feel good about teams' work, and also to raise issues and ask for feedback.

Collaborative, Evaluative Case Studies. Evaluative case studies were undertaken throughout the implementation phase of the program. Project implementers were involved in defining the focus and scope of these studies, articulating their project's ToC, and identifying potential informants. The (external) case study evaluators shared their findings with implementers in interactive sessions, where they were invited to assist with sense-making. Disagreements over interpretation sometimes arose, and were approached as a focus for learning. The case studies provided a means to look more deeply into specific project mechanisms and contexts, and to deepen evaluative thinking in the program. As we note below, they were not at first particularly helpful in supporting learning.

Learning Reported by Implementers

The survey and focus group results reflect both individual and organizational learning—and learning at multiple levels, from "Did we achieve our targets?" to "How should we measure and set our targets?" to "Are we measuring the right things?"

Several implementers gave examples of ways in which elements of the MERL system helped them to gain program-related insights.

By 2017 staff had come to find the field and monthly reporting a learning experience, as the following quotations illustrate:

- "Going through the process of the monthly report forces me to really sit down and think [about] what we have achieved as a team this month … and to what extent we have made progress. It helps to give a bit of a snapshot of the bigger picture, drawing us out of the nitty-gritty details of daily to-do-lists."
- "I am finding the monthly report in its current format to be very useful for reflective learning. It is much more streamlined than the previous version.… It takes me, on average, 4 hours to complete, but it is 4 hours well spent because I would use this time to think about activities for the following month and get a collective view on the two projects."

- "The B2O is very productive, especially if you do it as a team and straight away, because you carry the energy of the meeting into it. Whether it was a good meeting or a bad one."

Respondents emphasized the collective aspect of reporting and reflection. For example: "Working on the B2O with others is useful because you learn from what others have seen that you didn't pick up." Staff described the reporting template itself as significant: "The questions are carefully designed to promote reflection" and "it forces you to not simply repeat what happened or what was said in the meeting—you also, for example, need to look at what was not said."

Case studies were not always optimal opportunities for learning. Sometimes the timing was "off" and results were not ready to inform annual planning; sometimes the analysis was too limited, resulting in implementers gaining "no new insights." Reports were, at times, not read or acted on. Over time, the MERL team addressed shortcomings by presenting shorter reports at strategic moments toward year-end. One instance was reported where implementers were surprised by the findings of a case study and made an implementation change as a result. Staff also reported that being interviewed for the case studies provided a valuable opportunity to reflect in the company of another.

RESILIM-O days were valued by staff as "a time to ask questions and enable interactive learning," "a space to reflect on what we are doing and to push ourselves to see how we can apply what we learn," "a chance for staff who don't go into the field to learn about what is happening in the program," and an opportunity to give feedback on tools developed in the program, such as the MERL reporting templates.

Sometimes the learning reported was of a very practical nature—such as coming to realize that one needs "to spend more time on process design"—and several of the examples reported were simply instances of learning, over time, how to "do things better" (including MERL). Some implementers, on the other hand, reported very deep learning; for example, that their "instincts" had changed—in this case referring to processes for facilitating social learning. One staff member reported developing a fuller understanding of what he had previously only understood in theory—namely how transdisciplinarity works in practice. Another spoke of coming to "unlearn what I've known all my life," referring to assumptions about how intended development outcomes can be achieved.

Discussion and Conclusions

In this paper we described and analysed an M&E framework that aimed to meet the dual purposes of accounting to the funder and optimizing learning among implementers of a resilience-building program in Southern Africa, an intervention in a complex social-ecological system where real-time learning was of utmost importance.

This case study focused on learning by program implementers, which was found to have ranged from practical know-how, to deep personal and professional insights, to new organizational knowledge and practices (Boreham & Morgan, 2004). Using the "loop learning" framework derived from Schön (1983), we propose that the learning reported by RESILIM-O implementers ranged from single-loop learning (e.g. that one needs "to spend more time on process design") to second- and third- loop learning—for example, when one's "instincts" about processes for facilitating learning change, or when one transforms one's long-held assumptions about how to achieve development—coming to "unlearn what I've known all my life."

Some of the learning reported was strongly linked to individuals; conceptual, philosophical and practical in nature, and likely to enhance these individuals' professional capacity (although we have not collected data on how they have used their deeper learning). Other learning reported is in the organizational realm, particularly in becoming more adept as an organization at critically reflective practices like gathering relevant data from the field, making sense of it as individuals and collectives, and informing program decisions—that is, strategic adaptive management (Pollard et al., 2011).

The results of the historical overview, questionnaires, and focus group suggest that this learning was facilitated by working in new ways with standard M&E elements and introducing nonstandard features, in particular multiple opportunities for reflection and learning. After a fair amount of resistance during a difficult inception period, implementers started to value opportunities for shared reflection using MERL tools and events. The RESILIM-O MERL design and its iterative, collaborative adjustments over time managed to overcome many of the factors, outlined in the Why M&E Practices Can Fail... section above, that can inhibit learning in development programs.

Table 2. MERL Practices in RESILIM-O That Addressed Common Inhibitors to Learning

M&E practices that can inhibit learning within implementing agencies	How these practices were addressed in RESILIM-O
Program logic models fail to take complexity into account and do not prioritize learning.	Iterative, participatory development of a program ToC in the exploratory phase (first 2 years), with explicit efforts to take complexity into account and to prioritize learning.
ToC used only at the beginning of the program to lay out how things are expected to unfold; not revisited or questioned.	Regular (annual) reflection on sub-project ToCs, based on monitoring data, experience, and case studies, with modification if necessary.
Prioritizing *only* the accountability agenda crowds out learning.	Negotiation between AWARD and USAID created space and allowed sufficient resources for MERL throughout the program. AWARD's leadership prioritized learning and built it into many parts of RESILIM-O.
Reporting does not meet the needs of program implementers and is seen as a burden.	B2O reports were used to concisely capture relevant, evaluative information and promote collaborative reflections that implementers appreciated. Monthly reports drew together the reflections from the B2O reports. These were compiled collaboratively on a day dedicated to this purpose, using a template that promoted reflection on successes and challenges. Reporting to the funder (by the MERL team) combined quantitative and qualitative data.
Separation of monitoring and evaluation: M&E performed only by designated staff, with little involvement by program implementers.	Integration of monitoring and evaluation, through a developmental evaluation approach and part-internal, part-external MERL personnel. Reflection on outcomes and collaborative setting of targets by teams during annual work planning also helped to integrate monitoring and evaluation.
Evaluation done by external experts and/or only on program completion.	All implementers were involved in reporting as an activity that both provided monitoring data and promoted reflection.
Quantitative indicators are overemphasized	Both quantitative and qualitative data were collected, with qualitative data and written reflections being valued and promoted.
Narrative data is either not collected, or not in a form that is easily used for evaluation (e.g., meeting minutes or presentations that are not reflective or clearly linked to the aims of the project). It is then difficult to see how different aspects of the work fit together, to go back to the data and answer future questions, or to capture unintended outcomes or failures.	Double- and triple-loop learning were promoted through regular reflection activities: B2O reports, monthly reports, RESILIM-O days, "month in pictures," quarterly reflection on quantitative data, shared learning events, and attendance at conferences and other events (followed by reflection through B2O reports).
M&E is seen as a purely technical function.	MERL was promoted as an organizational development and strategic adaptive management function. It involved a combination of technical, managerial, and strategic advisory staff, who advocated for M&E use within and beyond the organization.

M&E practices that can inhibit learning within implementing agencies	How these practices were addressed in RESILIM-O
Learning is treated as equivalent to knowledge transfer.	Following initial conflict and through experimentation, AWARD could communicate *both* successes and learning (including failures) in RESILIM-O reporting. The developmental evaluation approach encouraged learning throughout the program and was not restricted to sharing of "lessons learnt" at the end.
M&E not integrated with program planning and decision-making.	MERL was integrated into annual work planning processes (collaborative revision of ToCs, objectives, and targets) and strategic decision-making.
Whether M&E processes are achieving their intended purpose is seldom evaluated.	The adoption, effectiveness, and feasibility of the MERL framework was reviewed informally during regular MERL team meetings and more formally through reflection events (e.g., RESILIM-O days, reference group meetings), "meta-evaluations," and conference presentations and papers.

The historical overview summarized in Table 1 shows that most elements of the RESILIM-O MERL system required adjustments over time. It is not surprising that a MERL system customized for a particular program and context cannot be fully designed at the start. With his concept of *developmental evaluation*, Patton (2010) argued that at different times in its life, a program will require different things from its M&E framework. In the case of RESILIM-O, the funder allowed an extended (2-year) "Phase 1," which created space for experimentation, innovation, and the iterative design of the M&E system, informed by a growing understanding of the context, and of implementation as well. This time and space was necessary for shaping a hybrid MERL framework that served the needs of the implementer and the funder.

This case of a developmental evaluation approach supports the proposal that since the concept was first introduced, new purposes of developmental evaluation emerged (Patton, 2021). In particular, the RESILIM-O case shows that developmental evaluation promotes individual and organizational learning and capacity development. It allowed evaluation to be embedded in the work of an organization, with the benefits also extending beyond the program for which the MERL system was developed. AWARD subsequently included aspects of MERL into other projects and started to include evaluation work as a new organizational competency.

The learning and experience were also "spun out" of the RESILIM-O program to benefit other projects, programs, and organizations (Pringle, 2011). The RESILIM-O MERL design was adopted for the design of a participatory MERL system in at least one other catchment in South Africa (Tsitsa Project, 2021). Following several invitations to share the emerging model, the learning about M&E design and practice was also incorporated into a university degree module and a short course. These spin-offs demonstrate the value that can be generated by co-design, participatory implementation, and allowing time and space for experimentation; that is, a developmental approach.

The results shared here also suggest that in addition to a responsive design, a successful MERL system requires a responsive disposition from implementers and MERL team alike. The director and project leads had to demonstrate to implementing staff and partners that learning and reflection were important. As argued elsewhere (Rosenberg & Kotschy, 2020) championing the role of MERL by working through barriers and building learning processes for adaptive management required relational and transformational skills in addition to technical skills.

The funder demonstrated their support by organizing several shared learning events themselves and inviting the MERL team to share lessons learnt in wider forums. The MERL team in turn had to be comfortable with trying out the innovation and working with program

implementers in a supporting, rather than a "policing," capacity. Interestingly, some team members who were trained and experienced in standard M&E practices found it difficult to adapt to the hybrid MERL approach. It became evident that a desire and ability to learn, despite years of experience, is part of the disposition needed to successfully develop and implement complexity-aware, developmental MERL systems.

Questioning standard M&E is not novel (nor is designing an alternative), but this analysis provides practice-based validation of some of the principles for learning-centered M&E that have been proposed elsewhere (Woodhill, 2007; Patton, 2010; Pringle et al., 2011). The case also provides actionable pointers that others may wish to explore for developing coherent, complexity-aware M&E frameworks in order to further build and deepen the practice and theory of monitoring, evaluation, and learning in complex sustainable development and resilience-building contexts. Actionable pointers include:

1. Where a context and a program are overtly complex, a *hybrid* M&E framework can *both* ensure accountability *and* facilitate learning, and combine both *standard and nonstandard* M&E elements.
2. A developmental approach, suitable for accountability and learning, requires *time* and *space* for the MERL framework to evolve with the program, as it responds to its growing understanding of a complex context.
3. Experiment with standard and nonstandard M&E elements (ranging from ToC to reporting formats and dedicated opportunities for shared reflection) and accept that they may need to be refined over time; involve implementers in such refinements.
4. Combining quantitative and complementary qualitative data in reporting is valuable, but careful attention is needed to the (participatory) collection of qualitative data, so that it is informative rather than onerous.
5. Building implementers' capacity to integrate evaluation for learning into program implementation processes requires not only buy-in, but explicit strategic commitment from the program leadership, including explicit use of evaluation findings during strategic planning, and "soft skills" from the MERL team.

Acknowledgments

The authors acknowledge Vhutshilo Mushwana Mudau, Jane Burt, Derick du Toit, Ray Ison, and Harry Biggs, who contributed to the development of the case; and USAID, for funding the RESILIM-O program, its evaluation, and the associated research. The process of producing the paper has been led by the three authors and is neither commissioned nor funded.

References

AWARD (2020). *RESILIM-Olifants: Resilience in the Limpopo Basin program final report*. AWARD, Hoedspruit. https://award.org.za/wp/wp-content/uploads/2020/12/AWARD-RESILIM-Olifants-FINAL-REPORT-PUBLIC-Oct-2020-web.pdf

Bakewell, O. (2003). *Sharpening the development process: A practical guide to monitoring and evaluation* (Praxis Guide No. 1). INTRAC.

Bellamy, J. A., Walker, D. H., McDonald, G. T., & Syme, G. J. (2001). A systems approach to the evaluation of natural resource management initiatives. *Journal of Environmental Management*, *63*, 407–423. https://doi.org/10.1006/jema.2001.0493

Boreham, N., & Morgan, C. (2004) A sociocultural analysis of organisational learning. *Oxford Review of Education, 30*(3), 307–325. https://doi.org/10.1080/0305498042000260467

Bornstein, L. (2006). Systems of accountability, webs of deceit? Monitoring and evaluation in South African NGOs. *Development, 49*(2), 52–61. https://doi.org/10.1057/palgrave.development.1100261

Chapman, S., Sullivan, C., Palm, C., Huynh, U., Diru, W., & Masira, J. (2016). Monitoring and evaluation to support adaptive co-management: Lessons learned from the Millennium Villages Project. *Journal of Environmental Management, 183,* 142–151. http://dx.doi.org/10.1016/j.jenvman.2016.08.014

Chikunda, C. (2017). Mediating sensemaking: Enriching social learning for resilience in natural resources management learning. *Proceedings of the 9th Action Learning Action Research and 13th Participatory Action Research World Congress 2015*. Action Learning Action Research Association (ALARA).

https://www.alarassociation.org/publications/other-alara-publications/world-congress-proceedings-2015

Chirau, T. J., & Blaser-Mapitsa, C. (2020). How performance management regulations shape evaluation practice in South African municipalities. *Evaluation and Program Planning, 82*, Article 101831. https://doi.org/10.1016/j.evalprogplan.2020.101831

Cundill, G., & Fabricius, C. (2009). Monitoring in adaptive co-management: Toward a learning based approach. *Journal of Environmental Management, 90*(11), 3205–3211. https://doi.org/10.1016/j.jenvman.2009.05.012

Douthwaite, B., & Hoffecker, E. (2017). Towards a complexity-aware theory of change for participatory research programs working within agricultural innovation systems. *Agricultural Systems*, *155*: 88–102. https://doi.org/10.1016/j.agsy.2017.04.002

Engeström, Y. (2016). *Studies in expansive learning: Learning what is not yet there*. Cambridge University Press. http://dx.doi.org/10.1017/CBO9781316225363

Engeström, Y., & Sannino, A. (2012). Whatever happened to process theories of learning? *Learning, Culture and Social Interaction*, *1*(1), 45–56. https://doi.org/10.1016/j.lcsi.2012.03.002

Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. John Wiley & Sons.

Hertz, T., Brattander, E., & Rose, L. (2021). Complexity-aware monitoring and evaluation. *Journal of MultiDisciplinary Evaluation*, *17*(41), 35–50. https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/679

Ison, R. L., Collins, K. B., & Iaquinto, B. L. (2021). Designing an inquiry-based learning system: Innovating in research praxis to transform science–policy–practice relations for sustainable development. *Systems Research and Behavioural Science*, *38*(5), 610–624. https://doi.org/10.1002/sres.2811

Kachur, D., Soal, S., & van Blerk, R. (2016). Stretching between learning and accountability: Experiences of South African non-governmental organisations. *African Evaluation Journal, 4*(1), Article 71. http://dx.doi.org/10.4102/aej.v4i1.71

Khan, M. S., Rothman-Ostrow, P., Spencer, J., Hasan, N., Sabirovic, M., Rahman-Shepherd, A., Shaikh, N., Heymann, D. L., & Dar, O. (2018). The growth and strategic functioning of One Health networks: A systematic analysis. *The Lancet Planetary Health*, *2*, e264–273. https://doi.org/10.1016/S2542-5196(18)30084-6

Lotz-Sisitka, H., Ali, M. B., Mphepo, G., Chaves, M., Macintyre, T., Pesanayi, T., Wals, A., Mukute, M., Kronlid, D., Tran, D. T., Joon, D., & McGarry, D. (2016). Co-designing research on transgressive learning in times of climate change. *Current Opinion in Environmental Sustainability, 20*, 50–55. http://dx.doi.org/10.1016/j.cosust.2016.04.004

Mushwana Mudau, V. (2020). *A study on reporting and learning in three natural resource management programmes in South Africa* [Unpublished master's thesis]. Rhodes University.

Mueller-Hirth, N. (2012). If you don't count, you don't count: Monitoring and evaluation in South African NGOs. *Development and Change, 43*(3), 649–670. https://doi.org/10.1111/j.1467-7660.2012.01776.x

Natsios, A. (2010, July 1). *The clash of the counter-bureaucracy and development*. Center for Global Development, July 2010. https://www.cgdev.org/publication/clash-counter-bureaucracy-and-development

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). SAGE.

Patton, M. Q. (2010). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Guilford Press.

Patton, M. Q. (2021). Emergent developmental evaluation developments. *Journal of MultiDisciplinary Evaluation*, *17*(41), 23–34. http://dx.doi.org/10.56645/jmde.v17i41.699

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. SAGE.

Pollard, S., du Toit, D., & Biggs, H. (2011). River management under transformation: The emergence of strategic adaptive management of river systems in the Kruger National Park. *Koedoe, 53*(2), 1–14. http://dx.doi.org/10.4102/koedoe.v53i2.1011

Pollard, S., Biggs, H., & du Toit, D. (2014). A systemic framework for context-based decision making in natural resource management: Reflections on an integrative assessment of water and livelihood security outcomes following policy reform in South Africa. *Ecology and Society*, *19*(2), Article 63. http://dx.doi.org/10.5751/ES-06312-190263

Preiser, R., Biggs, R., de Vos, A., & Folke, C. (2018). Social-ecological systems as complex adaptive

systems: Organizing principles for advancing research methods and approaches. *Ecology and Society, 23*(4), Article 46. https://doi.org/10.5751/ES-10558-230446

Pringle, P. (2011). *AdaptME toolkit: Adaptation monitoring and evaluation*. United Kingdom Climate Impacts Programme (UKCIP).

Reed, M. S., Evely, A. C., Cundill, G., Fazey, I., Glass, J., Laing, A., Newig, J., Parrish, B., Prell, C., Raymond, C., & Stringer, L. C. (2010). What is social learning? *Ecology and Society*, *15*(4). http://dx.doi.org/10.5751/ES-03564-1504r01

Rosenberg, E., Kotschy, K., Burt, J., Mudau, V., du Toit, D., & Pollard, S. (2017). *Monitoring, evaluation, reporting and learning for the USAID RESILIM-O program: MERL framework*. AWARD. https://award.org.za/wp/wp-content/uploads/2021/04/AWARD-TECH-REPORT-18-RESILIM-O-Monitoring-Evaluation-Reporting-Learning-MERL-Framework-2017.pdf

Rosenberg, E, & Kotschy, K. (2020). Monitoring and evaluation in a changing world: A Southern African perspective on the skills needed for a new approach. *African Evaluation Journal*, *8*(1), Article 472. https://doi.org/10.4102/aej.v8i1.472

Roux, D. J., Nel, J. L., Cundill, G., O'Farrell, P., & Fabricius, C. (2017). Transdisciplinary research for systemic change: Who to learn with, what to learn about and how to learn. *Sustainability Science*, *12*(5), 711–726. https://doi.org/10.1007/s11625-017-0446-0

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Schulz, K. (2010). *Being wrong: Adventures in the margin of error*. Portobello.

Scharmer, C. O. (2009). Theory U: Leading from future. *Carl-Auer-Systeme Verlag*.

Taylor, J., & Soal, S. (2011). Measurement in developmental practice: From the mundane to the transformational. In A. Fowler & C. Malunga (Eds.), *NGO management* (Chapter 26). Routledge. http://dx.doi.org/10.4324/9781849775427-34

Tsitsa Project, Kotschy, K., Cockburn, J., Conde-Aller, L., & Rosenberg, E. (2021). *Participatory monitoring, evaluation, reflection and learning (PMERL): Building a participatory and sustainable system for evaluating impact* (Tsitsa Project Practice and Policy Brief No. 5). Rhodes University. https://www.ru.ac.za/tsitsaproject/resourcesresearch/keyinformation/

USAID. (2021). *Complexity-aware monitoring discussion note*. United States Agency for International Development. https://usaidlearninglab.org/library/complexity-aware-monitoring-discussion-note-brief

Villanueva, P. S. (2011). *Learning to ADAPT: Monitoring and evaluation approaches in climate change adaptation and disaster risk reduction—challenges, gaps and ways forward* (Strengthening Climate Resilience Discussion Paper 9). Institute of Development Studies.

Walker, B., & Salt, D. (2006). *Resilience thinking: Sustaining people and ecosystems in a changing world*. Island Press.

Wals, A. E. (2007). Think Piece. Learning in a changing world and changing in a learning world: Reflexively fumbling towards sustainability. *Southern African Journal of Environmental Education*, *24*, 35–45.

Weiss, C. H. (1997). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation*, *76*, 41–55. https://doi.org/10.1002/ev.1086

Williams, B. K., & Brown, E. D. (2018). Double-loop learning in adaptive management: The need, the challenge, and the opportunity. *Environmental Management, 62*, 995–1006. https://doi.org/10.1007/s00267-018-1107-5

Wongtschowski, M., Oonk, L., & Mur, R. (2016). *Monitoring and evaluation for accountability and learning* (Working Paper 2016-3, Kit Sustainable Economic Development and Gender). https://www.kit.nl/wp-content/uploads/2019/10/WPS_3-2016-web.pdf

Woodhill, J. (2007). M&E as learning: Rethinking the dominant paradigm. In J. de Graaf, J. Camerson, S. Sombatpanit, C. Pieri, & J. Woodhill (Eds.), *Monitoring and evaluation of social conservation and watershed development projects* (pp. 83–107). Science Publishers.