

# Reframing Evaluation: Complexity, Systems, and Transformation

*Journal of MultiDisciplinary Evaluation*  
Volume 22, Issue 53, 2026

**JMDE**  
Journal of MultiDisciplinary Evaluation

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

Hayat Askar  
*Independent Consultant*

Barbara Schmidt-Abbey  
*Independent Consultant*

Liliana Olivia Lucaciu  
*Lideea Development Actions*

Anna Maria Augustyn  
*Independent Consultant*

Scott Chaplowe  
*International Evaluation  
Academy*

Neha Dhingra  
*Center for Research in  
Schemes and Policies (CRISP)*

Ian Goldman  
*International Evaluation  
Academy*

Amrita Gupta  
*Independent Consultant*

Jonathan (Jonny) Morell  
*4.669 Evaluation and  
Planning*

Nan Wehipeihana  
*Weaving Insights*

Anna Matheson  
*Victoria University of  
Wellington*

**Background:** The world is increasingly facing rapid degeneration of social and ecological systems. This requires evaluation to focus on the underlying systems rather than isolated interventions, prompting an interest in systems-informed evaluation, and evaluation that focuses on systems.

**Purpose:** This article proposes a practical conceptual framework that links widely recognized systems principles with concrete evaluation choices across the policy and evaluation lifecycle.

**Setting:** The framework is grounded in current systems thinking and complexity literature and informed by multiple traditions in evaluation, including theory-based, participatory, and learning-oriented approaches.

**Intervention:** Not applicable.

**Research Design:** A qualitative, integrative research design combining a systematic literature review, analysis of practitioner perspectives from the Three Horizons (3H)

survey, and a cross-case analysis of three real-world examples—humanitarian energy systems (RE4R), Indigenous community health systems (Healthy Families NZ), and portfolio-based municipal development (UNDP)—coauthored by practitioners and selected using systematic searching.

**Data Collection and Analysis:** Literature was identified through systematic database searches and AI-assisted screening tools; 3H survey responses were thematically analyzed with AI-supported organization and human interpretation; case examples were selected through structured search and outreach, then analyzed against the conceptual framework in collaboration with practitioner coauthors.

**Findings:** The article concludes that systems-informed evaluation is less a specialist method than a practice stance, requiring intentional choices that hold complexity, share power amongst multiple stakeholders, and support ongoing learning. An agenda emerges that may motivate collaboration among partners to advance this practice.

**Keywords:** *systems; systems thinking; evaluation; complexity; transformational change; Indigenous evaluation; portfolios; learning-oriented evaluation*

Contemporary evaluation practice has traditionally focused on assessing discrete interventions such as policies, programs, and projects against predefined outcomes. This approach is increasingly challenged by today's interconnected social and ecological crises (the polycrisis), where interventions interact with wider systems shaped by feedback loops, resource flows, uncertainty, and emergent change. Linear models struggle to capture these dynamics or to support learning across complex systems. The International Evaluation Academy (IEAc) has focused on how evaluation can address these systemic crises. As part of addressing this issue, the academy has undertaken a Three Horizons (3H) futures process on how evaluation can contribute to addressing the polycrisis, discussed below, from which the need for evaluation to focus on addressing complex adaptive systems arose strongly. Transformation thinking is being elaborated in this special edition on evaluation and the transformational imperative. This present article focuses on how evaluation can shift to a systems focus, by acknowledging that current systems thinking sits within a broader landscape of knowledge. Evaluation practice benefits from recognizing such epistemic plurality.<sup>1</sup>

Encouragingly, emerging practices show greater attention to complexity, environmental change, and adaptive learning. Approaches drawing from systems dynamics, organizational learning, and community-based practice encourage managers, funders, and evaluators to consider system-level interactions, resilience, and adaptive capacity rather than focusing solely on intervention-level outputs. Recent work argues that evaluation must adapt to such conditions by moving beyond narrow program cycles and individual interventions, toward approaches that consider interdependence, adaptation and system change (Ofir, 2021; Buckton et al., 2025; Goldman et al., 2026). These developments reposition evaluation beyond a primary accountability function, placing greater emphasis on its contribution to systemic insight, learning, improvement, and transformation.

## Methodology

The methodological design of this study is structured around two guiding questions:

1. Why must evaluation evolve to better address systems and complexity?
2. How can systems thinking be incorporated into evaluation practice in concrete and usable ways?

To answer these questions, the article adopts a qualitative, integrative research design that combines literature review, practitioner perspectives, and applied case analysis. Central to this design is the development of a practical conceptual framework that serves as an analytic heuristic for examining how systems thinking can be enacted across the evaluation life cycle. Methodologically, the article proceeds in four steps:

1. A systematic literature review on systems-informed and complexity-aware evaluation.
2. Analysis of practitioner perspectives from the Three Horizons (3H) survey.
3. Development of a conceptual framework for applying systems thinking in evaluation.
4. Testing of the framework with three illustrative case examples coauthored by practitioners, selected using systematic searching.

Combined, these elements form a triangulated approach that supports credibility, relevance, and adaptability in complex evaluation settings (Morell, 2010; Patton, 2010). Further methodological detail for each step is provided in Appendix 1.

## Developing a Heuristic Framework for Bringing Systems Thinking into Evaluation

This article centers on the need for a heuristic framework that helps practitioners to link systems principles with concrete evaluation choices across the policy and evaluation cycle, addressing the gap between a rich body of literature and the limited guidance available to practitioners and commissioners on how these

<sup>1</sup> One of the case examples included here reflects an Indigenous approach within Aotearoa New Zealand, reflecting such plurality.

ideas could inform concrete actions, timing, and responsibilities across different evaluation roles.

This section outlines the four-step synthesis through which methodological inputs were integrated into a heuristic framework. Insights from systems and complexity literature, evaluation practice, and the Three Horizons analysis were combined to identify key evaluation decision points across the evaluation life cycle.

### *Analytical Insights from Systems and Complexity Literature*

Efforts to integrate systems thinking into evaluation stem from a shared recognition that change in complex environments rarely follows linear or isolated pathways. Across evaluation, systems practice, and social-ecological research, a set of recurring concepts has been used to understand how change unfolds under conditions of interdependence and uncertainty. These include interrelationships, multiple perspectives, system boundaries, and dynamics over time (SETIG, 2018). Together, they shift evaluative attention away from discrete interventions and toward the patterns of interaction through which outcomes emerge.

These concepts have been operationalized through a range of analytical frameworks, each addressing different aspects of complexity. Some focus on sense-making under uncertainty. The Cynefin framework, for instance, distinguishes between simple, complicated, complex, and chaotic contexts, helping evaluators judge when prediction is feasible and when adaptive learning is more appropriate (Snowden & Boone, 2007). Other frameworks foreground contextual embeddedness and governance conditions. The Johannesburg five-dimensional framework brings regulatory, institutional, socioeconomic, and political dimensions into view, supporting evaluation in policy-dense and multi-actor settings (Ndhlovu et al., 2017). Socioecological frameworks, in turn, focus on coupled human–environment systems, enabling analysis of how environmental dynamics, community practices, and policy responses interact over time (Binder et al., 2013).

Across these traditions, a consistent insight emerges: No single framework can be applied uniformly across contexts (Snowden & Boone, 2007; Williams & Hummelbrunner, 2010; Binder et al., 2013). Systems-informed evaluation therefore requires analytical choices to be shaped by the characteristics of the system under study rather than by methodological preference

(SETIG, 2018; Morell, 2010). This shifts the evaluator's task from selecting a preferred method to clarifying which aspects of system behavior, interaction, and change need to be understood in a given context (Patton, 2011; Ison et al., 2021).

To support this work, a range of mapping, diagnostic, and analytical practices have been developed across systems-oriented and complexity-aware evaluation (Coffman, 2007; Reynolds et al., 2016; Latham, 2014). These include tools for identifying actors and relationships, tracing feedback loops, clarifying system boundaries, and examining configurations of conditions that contribute to change (Binder et al., 2013; Wilson-Grau & Britt, 2012; Pawson & Tilley, 1997). Other approaches focus on analyzing dynamics over time, such as system dynamics modeling, network analysis, and theory-based inquiry (Schweitzer et al., 2009; Beach & Pedersen, 2013; Befani & Mayne, 2014). While diverse in form, these practices share a common analytical purpose: making interaction, adaptation, and contribution visible within dynamic systems, rather than attributing outcomes to single causes (Patton, 2011; Befani, 2016).

Much of this work responds directly to the limitations of conventional evaluation approaches. Models anchored in prespecified outcomes and fixed time frames often produce routinized information that obscures feedback effects, path dependence, and delayed or indirect change (Leeuw & Furubo, 2008). Systems-oriented evaluation, by contrast, places greater emphasis on reflective learning, attention to emergent outcomes, and iterative adaptation as interventions interact with changing system conditions (Patton, 2011; Darnhofer et al., 2017; Ison et al., 2021; Geels & Schot, 2007).

Alongside conceptual frameworks, practice-oriented tools have been developed to respond to the dynamic, nonlinear, and uneven pace at which systems change over time. Rapid assessment toolkits, climate evaluation frameworks, and resilience-focused approaches embed assumptions about emergence, adaptability, and long-term viability (van den Berg & McCollum, 2021). These tools are intended not only to generate evidence, but to support ongoing sense-making by helping evaluators and stakeholders identify leverage points and adjust strategies as conditions evolve.

In the literature, we found many approaches that respond to complexity using principles of “systems evaluation.” Outcome harvesting, realist evaluation, developmental evaluation,

process tracing, and Bayesian approaches to causal inference were all developed to address nonlinear causality, contribution, and contextual variation (Pawson & Tilley, 1997; Patton, 2011; Wilson-Grau & Britt, 2012; Beach & Pedersen, 2013; Befani, 2016). Although not always framed as systems approaches, they integrate systems logic and have been applied in complex, multi-actor interventions.

The intent of this section is therefore not to catalogue frameworks or methods, but to extract analytical insights that inform how evaluators make systems visible, reason about causality under uncertainty, work with emergence and adaptation, and align analytical choices with system characteristics. Table 1 synthesizes key systems principles and the analytical insights they offer into how systems behavior and change are understood.

Table 1. Systems Principles Shaping Analytical Understanding of Systems Behavior and Change

| Systems principle                                            | Analytical insight for evaluation                                                                                                                                                                                                                                                                  |
|--------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Interrelationships                                           | Understanding how interactions among actors, institutions, and contextual factors shape system behavior and change.                                                                                                                                                                                |
| Multiple perspectives                                        | Recognizing that different actors experience and interpret system dynamics differently. Recognizing that these perspectives are necessary to understand how change occurs within complex systems.                                                                                                  |
| System boundaries                                            | Clarifying what is included, excluded, or marginal in the system under study, and how boundary choices shape what can be observed, analyzed, and claimed in evaluation.                                                                                                                            |
| System dynamics and nonlinearity<br>Emergence and adaptation | Examining how change emerges through feedback loops, delays, and nonlinear patterns rather than through linear cause–effect relationships. Identifying unplanned or unexpected outcomes and understanding how systems adapt over time in response to internal interactions and external pressures. |
| Reframed causality (contribution rather than attribution)    | Analyzing how outcomes arise from interacting mechanisms and contextual conditions, focusing on contribution to change rather than isolating single causal effects.                                                                                                                                |
| Contextual embeddedness and holistic system framing          | Situating interventions within wider governance, socioeconomic, and environmental systems, and analyzing how context conditions shape system behavior and change pathways.                                                                                                                         |

### *Paradigm Shift in Evaluation Practice*

Building on the analytical insights distilled from the literature in the previous section, this section examines how evaluation scholarship describes a broader shift in evaluation practice under conditions of complexity. The literature consistently argues that systems thinking should be embedded from early intervention planning and sustained across the full policy cycle, rather than applied only retrospectively (Morell, 2010; SETIG, 2018; Defra, 2020). Linking evaluation with planning and programming is not a new concern; it has long been a central preoccupation

of evaluation practice. What is increasingly evident, however, is a broader shift in how evaluation is understood and practiced when engaging with complex systems. Traditional evaluation models struggle to capture feedback effects, adaptive responses, and system-level change that unfold over longer periods and across multiple actors. As a result, evaluations often generate technically robust findings while remaining strategically limited in their ability to reflect how change occurs within complex systems.

A systems-informed approach reframes evaluation with interventions treated as

components within wider socioeconomic, institutional, and ecological systems, often spanning sectors, scales, and time frames. Evaluation, in this view, contributes to understanding how these systems behave, adapt, and respond to intervention efforts (Garcia & Uitto, 2025). This different approach affects how the traditional evaluation tools are used. Theories of change, logic models, and results chains remain relevant, but they require adaptation to reflect nonlinearity, feedback, and multiple pathways of change, and to support learning about how systems evolve in practice (Morell, 2025; Douthwaite & Hoffecker, 2017; Patton, 2011). In system-level and transformational contexts, this adaptation also entails greater attention to how evaluation engages with local realities and diverse knowledge systems. Effective transformation evaluation therefore requires methodologies that foreground localization and cultural responsiveness in order to capture systemic change at the local level (Shipurkar et al., 2026).

At the same time, ethical and sustainability considerations—particularly in coupled human–natural systems—are increasingly integral to this paradigm, with evaluations expected to assess resilience, justice, and long-term viability alongside effectiveness (Garcia & Uitto, 2025; Bellamy et al., 2001; Zaveri et al., 2026).

Methodologically, the shift is defined by how evaluative inquiry is configured. Systems-informed evaluation typically draws on combinations of approaches, selecting and recombining methods in response to system characteristics and evolving questions. The literature on multimethodology and methodological bricolage highlights the importance of this flexibility, emphasizing that evaluators adapt inquiry designs to the problem (Mingers & Gill, 1997; Reynolds et al., 2016; Aston & Apgar, 2022).

Where appropriate data and capacity exist, evaluators may draw on analytical and modeling approaches such as qualitative comparative analysis, social network analysis, or system dynamics to explore patterns of interaction and change. Visual tools, including system maps and causal loop diagrams, support shared sense-making by making assumptions, boundaries, and feedback visible to diverse stakeholders (Coffman, 2007; Latham, 2014). These approaches are rarely used in isolation; their value lies in how they are combined and situated within a broader evaluative process.

### *Views from Prominent Evaluation Thinkers and Activists on the Role of Systems Thinking in the Future of Evaluation*

As mentioned previously, the IEAc has been undertaking futures work on how evaluation can contribute to addressing the polycrisis. Practitioner perspectives from the Three Horizons survey were analyzed to test how systems thinking is currently understood, experienced, and imagined within deep thinkers and activists on the future of evaluation. More details on the Three Horizons methodology and sampling characteristics are covered in another article within this special edition (Ofir et al., 2026).

While systems thinking was not explicitly asked about in the survey, the analysis showed that almost all respondents used system language across the survey, indicating its strong conceptual uptake. However, this apparent consensus masked substantial variation in meaning. While some respondents referred to concrete organizational or governance systems, others invoked broader socioecological systems or more abstract notions of interconnectedness. The most common pattern was “mixed” usage, where different systems concepts were combined without clear boundaries. This conceptual plurality reflects both the richness of systems thinking in evaluation and a source of ambiguity that complicates its application in practice.

Applying the Three Horizons framework to these responses revealed how this plurality plays out over time. In Horizon 1, responses were dominated by challenges (75 mentions), reflecting present-day structural constraints and a situation not fit for purpose. Practitioners described rigid commissioning processes, accountability pressures, limited time and capacity, and institutional expectations that favor linear planning and reporting. These constraints shape how evaluation is currently practiced and help explain why systems thinking is often acknowledged albeit unevenly enacted.

Horizon 3 is a visioning space. Systems thinking was widely imagined as central to the future of evaluation that could contribute to addressing the polycrisis, linked to participatory approaches, technological innovation, and a stronger focus on transformation rather than program-level performance alone. The future vision was more coherent than present practice, suggesting that practitioners are clearer about where evaluation needs to go and that substantial change is needed.

Horizon 2 captured a transition space characterized by experimentation. Here, respondents described attempts to adapt evaluation approaches, pilot new tools, or introduce more participatory and learning-oriented practices. At the same time, these efforts remained strongly shaped by legacy practices and governance arrangements, some of which need to be retained or even amplified. The near balance between challenges (45 mentions) and ideas (41 mentions) in this horizon reflects a field in motion, not yet consolidated around shared approaches.

Across all three horizons, a consistent pattern emerges in the relationship between conceptual uptake and methodological development. Systems thinking is widely recognized and frequently referenced, and practitioners demonstrate a strong ability to articulate systems-related challenges and dynamics.

However, concrete approaches for evaluating systems themselves, beyond individual interventions, were relatively limited, with few established or widely shared methodologies suggested to support practice. This imbalance underscores the observation in Section: Paradigm Shift in Evaluation Practice that the current state of systems-informed evaluation reflects not a lack of interest or conceptual engagement, but a need for more practical, integrative guidance to support implementation.

The views of these prominent thinkers and activists suggest that evaluation is at a critical juncture. Systems thinking is widely recognized and increasingly embedded in aspirations for the field's future. Nevertheless, current practice faces constraints from the broader evaluation ecosystem and institutional structures and limited methodological scaffolding. The concentration of challenges in Horizon 1 and ideas in Horizon 3 indicates a field that sees many problems in the present and envisages a different future while struggling to consolidate solutions in the present.

These findings carry several implications that directly informed the design of the conceptual framework. First, the observed conceptual plurality underscores the need for a framework that provides shared reference points without enforcing a single and constraining definition of "the system." Second, the gap between aspiration and practice highlights the importance of translating systems thinking into concrete evaluation decision points that can be acted upon by different roles within the evaluation process. Third, the transitional nature of Horizon 2 suggests that change in evaluation practice is

likely to reflect multiple emergent approaches gradually coalescing into field-wide change and coordinated change across multiple actors.

In this way, the Three Horizons analysis completes the conceptual grounding provided by Sections: Analytical Insights from Systems and Complexity Literature and Paradigm Shift in Evaluation Practice. While the literature clarifies how systems behavior and change can be understood and the paradigm-shift literature explains why evaluation practice must change, deep practitioner perspectives illuminate where the field currently stands and what kinds of guidance are most needed. These combined insights informed the development of the systems-informed evaluation conceptual framework presented in the following section.

### *From Principles to Practice: A Systems-Informed Evaluation Conceptual Framework*

Building on the analytical principles extracted from the literature (Table 1), the discussion of the paradigm shift in evaluation practice (Section: Paradigm Shift in Evaluation Practice), and practitioner perspectives from the Three Horizons analysis (Section: Views from Prominent Evaluation Thinkers and Activists on the Role of Systems Thinking in the Future of Evaluation), this section introduces a conceptual framework for applying systems thinking across the evaluation life cycle.

The framework in Table 2 integrates key features of systems-informed practice into 11 evaluative elements. These elements reflect current practice in complexity-aware evaluation and provide a structured way to incorporate systems thinking from commissioning and design through interpretation, use, and learning.

The 11 elements sit across the evaluation process in different ways. Some elements shape work before an evaluation begins, such as understanding systems and complexity, stakeholder ownership and governance, and terms of reference and question formulation, as well as theory of change development and use. Others guide what happens during the evaluation, including methodological design, procurement and evaluator selection, fieldwork and data collection, and analysis and interpretation, alongside validation and review. A final set becomes most visible after the evaluation concludes, where follow-up and use and communication and knowledge-sharing influence how learning continues, deepens, or loses momentum.

The framework serves two purposes. First, it offers evaluators, commissioners, and practitioners a set of guiding questions for designing and undertaking system-oriented evaluations in different contexts. Second, it illustrates how systems features may be applied in practice across diverse contexts. In each case, selected elements of the framework are visible in practice; however, these are not applied uniformly, reflecting the contextual nature of systems-informed evaluation and the varied

maturity of practice across settings. The heuristic nature of the framework is intended to accommodate different assumptions about what can be known in advance in complex systems, including both theory-guided and emergence-led approaches to inquiry. The framework is therefore presented as a heuristic to support reflective adaptation and learning rather than fixed compliance.

Table 2. Conceptual Framework Linking Systems Principles and Methods to the Evaluation Process

| Element                                              | Systems-informed evaluation focus                                                                                                                                                     |
|------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Pre-evaluation                                       |                                                                                                                                                                                       |
| 1. Understanding systems and complexity              | Clarifies whether the focus is on systems or interventions; uses system framing to guide tool selection and analysis.                                                                 |
| 2. Stakeholder ownership and governance              | Identifies who participates in governance and decision-making, and how power dynamics are managed.                                                                                    |
| 3. Terms of reference (ToR) and question formulation | Ensures questions reflect system behaviors, boundaries, and feedback, and are codeveloped where possible.                                                                             |
| 4. Theory of change (ToC) development and use        | Uses theory of change where appropriate to explore plausible pathways and feedback, while holding it lightly and remaining open to iterative emergence-led learning.                  |
| 5. Methodological design                             | Uses multimethod, adaptive designs aligned with system uncertainty; combines theory-based, participatory, and emergence-led approaches, including sensing and iterative sense-making. |
| 6. Procurement and evaluator selection               | Prioritizes systems competencies, cultural responsiveness, facilitation, and reflexive capabilities alongside technical skills.                                                       |
| During the evaluation (implementation)               |                                                                                                                                                                                       |
| 7. Fieldwork and data collection                     | Aligns data strategies with system understanding; may integrate systems mapping, lived experience, network or remote sensing data.                                                    |
| 8. Analysis and interpretation                       | Examines interdependencies, feedback, boundaries, and emergent change; encourages shared interpretation and sense-making.                                                             |
| 9. Validation and review                             | Treats validation as an inclusive and ongoing process involving users, affected groups, and those with lived experience.                                                              |
| Post-evaluation                                      |                                                                                                                                                                                       |
| 10. Follow-up and use                                | Supports system learning, adaptation, and scaling; embeds evaluation use into strategic decisions.                                                                                    |

|                                         |                                                                                                                            |
|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| 11. Communication and knowledge sharing | Tailors communication to diverse audiences; uses two-way learning processes, visualization, and culturally grounded forms. |
|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------------|

*Note.* A more detailed table with guiding questions is included in Appendix 2.

## Applying the Framework on Real-World Cases

The previous sections provided the conceptual foundation and rationale for the inquiry. We now present three illustrative case studies that offer practical snapshots of how elements of systems-oriented evaluation are emerging in diverse contexts.

### *Evaluating Transformational Change in Humanitarian Energy Systems: Lessons from the RE4R Systems Impact Assessment*

The Renewable Energy for Refugees (RE4R) initiative, funded by the IKEA Foundation and implemented by Practical Action and UNHCR between 2017 and 2022, sought to shift global humanitarian response to adopt market-based renewable energy models demonstrated in three refugee camps and host communities in Rwanda, and in the city of Irbid in Jordan. In addition to two country-level evaluations, a dedicated systems impact assessment was commissioned in 2021 to explore RE4R's contribution to global humanitarian system change. This case reflects on how that evaluation applied systems thinking to assess transformational intent, feasibility, and contribution (Chaplowe & Gupta, 2022).

A key early step was to assess whether the evaluation's ambition was matched by its framing. The independent evaluation team began by reviewing the terms of reference, which articulated clear systemic ambitions. The evaluation's transformational intent and fidelity were then validated through the recruitment interview and early discussions with Practical Action, which managed the evaluation, and

further informed by a review of the project's theory of change and relevant secondary data sources.

Together, these early exchanges functioned as an informal but important evaluability assessment. They tested whether the evaluation's design, scope, and stakeholder commitment were genuinely aligned with the aims of transformational change, and whether the approach was appropriately calibrated to the inherent complexity—and methodological limitations—of assessing such change. They also helped ensure that findings would be credible, relevant, and usable for strategic learning and influence, while confirming expectations, clarifying constraints, and gauging stakeholders' readiness to engage with potentially uncomfortable findings should systemic barriers be identified.

To establish a shared understanding of what constitutes system change, the team adopted the Transformational Change Learning Partnership (TCLP) definition used by the Climate Investment Funds (CIF): “a deep and fundamental change in a system's form, function, or process” (CIF, 2021). This conceptualization was sufficiently flexible to be understood and accepted by multiple stakeholders across humanitarian, energy, and market systems.

The CIF–TCLP framework was then adapted to generate evaluation criteria tailored to the RE4R context. Table 3 summarizes the five resulting criteria. Two notable adaptations emerged: (1) the TCLP dimension of speed was incorporated into the systems change criterion; and (2) Transformational coherence was added as a standalone criterion to examine how RE4R's efforts aligned with other transformational initiatives.

Table 3. Transformational Evaluation Criteria Adapted from CIF–TCLP

| Criterion                  | Purpose and scope                                                                                                                                                               |
|----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Transformational relevance | Assesses what is being done to affect systems change, including systemic assessment, project design and implementation, monitoring and communication, and capacity development. |

|                            |                                                                                                                                                                       |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Systems change             | Assesses extent to which systems change has occurred, including structural, attitudinal, and transactional change, as well as the speed at which change has occurred. |
| Scalability                | Assesses extent to which system change can increase at a broader scale, including vertical, horizontal, functional scaling.                                           |
| Transformational coherence | Assesses extent to which systems change aligns and integrates with other transformational efforts within the wider system.                                            |
| Adaptable sustainability   | Assesses extent to which system change is likely to persist and adapt over time under evolving conditions.                                                            |

*Note.* From “Transformational Change Concepts: Working Definition, Dimensions, Questions and Signals,” by Climate Investment Funds (CIF), 2021, retrieved from <https://www.cif.org/knowledge-documents/transformational-change-concepts-working-definition-dimensions-questions-and>

To operationalize these criteria, the team developed detailed rubrics to guide data collection, analysis, and interpretation. Each rubric used a four-level rating scale—*harmful*, *insufficient*, *good enough*, and *exemplary*—to

avoid neutral “middle-ground” judgments and encourage evaluators to weigh evidence toward either positive or critical interpretation. Table 4 illustrates the rubric applied to the systems change criterion.

Table 4. Example Rubric for Assessing Systems Change Contribution

| Rating level | Interpretation for systems change contribution                                                                                                          |
|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| Harmful      | Conditions created by the intervention hinder or undermine the uptake of renewable energy in the global humanitarian system.                            |
| Insufficient | Conditions created by the intervention partially support the uptake of renewable energy within the global humanitarian system but remain inadequate.    |
| Good enough  | Conditions created by the intervention support the uptake of renewable energy within the global humanitarian system, though gaps or limitations remain. |
| Exemplary    | Conditions created by the intervention significantly support the uptake of renewable energy within the global humanitarian system.                      |

*Note.* The rubric was tailored for the RE4R evaluation, with explicit descriptors and evidence markers for each criterion.

The team found that a structured yet flexible framework, grounded in criteria and a rubric codeveloped with stakeholders, created clarity for stakeholders and discipline for evaluators working in a fast-paced humanitarian context. The collaborative definitional work at the outset was critical: Without early agreement on what transformational change means, evidence risks being interpreted in inconsistent or overly optimistic ways. The case therefore demonstrates how systems-oriented evaluation

benefits from negotiated expectations, shared conceptual grounding, and methodological transparency, particularly when navigating emergent, politically sensitive, and highly complex change.

*Indigenous Evaluation: Applying Systems Thinking in the Evaluation of Healthy Families NZ*

Healthy Families NZ (HFNZ), launched in 2014, is a community-driven initiative designed to strengthen prevention systems and improve health and well-being by tackling the broader causes of chronic disease. Grounded in the places where people live, learn, work, and play, HFNZ is active in 10 geographically and culturally diverse locations across Aotearoa. Each site is supported by a local lead provider organization and an embedded team, with a strong emphasis on equity, community leadership, and improved Māori health outcomes. The initiative is underpinned by local knowledge, mātauranga Māori, and co-design, positioning systems change as a community-led process.

The HFNZ evaluation (2014–2026; Matheson et al., 2020) embraces systems thinking and complexity as core to its approach. Its purpose is to understand if and how prevention systems are shifting, how those shifts are strengthened by local leadership, and whether change makes healthier lives more achievable for communities. The evaluation therefore focuses not on discrete activities but on how systems themselves are being reshaped across time and place.

Collaboration has shaped the evaluation from the outset. Terms of reference and evaluation questions were codeveloped by the Ministry of Health and the evaluation team, with regular input from national and local HFNZ teams. This iterative process ensures alignment with HFNZ's values—equity, systems change, and responsiveness to whānau and community priorities—and protects the relevance of the evaluation as the initiative evolves. Procurement reinforced this intent by prioritizing capability in systems evaluation, kaupapa Māori approaches, economic evaluation, and mixed methods. The resulting team blends expertise in public health, Indigenous evaluation, biostatistics, and systems thinking, and continues to adapt its practice alongside the initiative.

Power-sharing is central to both HFNZ and its evaluation. Rather than positioning communities as data sources, the evaluation honors HFNZ's kaupapa by using relational methods such as outcome narratives, participatory interviews, and storytelling. Insights are generated through a “community-up” stance, where local experience and wisdom shape what is interpreted, valued, and acted upon. This approach moves away from extractive methodologies and recognizes that power is exercised not only through participation but

through how voices influence meaning and decision-making.

Each HFNZ location has a strategic leadership group (SLG) comprising community members, local leaders, and sector representatives. These groups guide implementation and evolve in response to community dynamics. At the national level, the Ministry works alongside local teams as strategic partners, rather than contract managers, reinforcing a shared leadership approach.

Methodologically, the evaluation uses a longitudinal comparative case study design. It incorporates tools such as qualitative comparative analysis (QCA), cost-consequences analysis, and Te Pae Māhutonga (a Māori health framework). In its second phase (2019–2022), the prevention action framework was introduced to assess the direction and quality of system strengthening in each location. This framework is grounded in kaupapa Māori and mātauranga Māori and informed by systems thinking and cumulative learning from earlier evaluation phases.

Evaluation practice draws on a broad toolkit, including system mapping, co-design workshops, the Donella Meadows Six Conditions of Systems Change framework, and kaupapa Māori methods. Storytelling is used purposefully to elevate community knowledge and shift thinking from linear interventions to more holistic system change.

Stakeholder engagement is intentionally relational, continuous, and non-extractive. Communities are not only respondents; they are coanalysts and sense-makers. Evaluation teams emphasize humility (māhaki), stepping back so that leadership rests with communities. This shared leadership is vital to surfacing what matters in each place and ensuring that the evaluation strengthens, rather than merely observes, community aspirations.

### *UNDP Portfolio Approach*

This case illustrates a situation in which both the initiative and its evaluation adopt systems-oriented approaches, resulting in an intentionally porous boundary between implementation and evaluation. Participatory monitoring, evaluation, and learning (MEL) and adaptive management are embedded within ongoing practice, supporting iterative learning rather than treating evaluation as a separate, retrospective exercise. The case is based on the publicly available

documents and does not reflect the official views of the organizations involved.

The United Nations Development Program (UNDP) has increasingly incorporated systems thinking into policy formulation, program implementation, and evaluation. This shift recognizes that complex development challenges, such as poverty, governance, or climate adaptation, are embedded within adaptive systems, where linear evaluation models risk missing how relationships, feedback loops, and institutional dynamics influence change. Systems approaches enable the mapping of interdependencies, identification of leverage points, and exploration of how change unfolds across multiple levels.

One expression of this shift is the portfolio approach, developed across several UNDP country contexts and designed to work with interconnected sets of interventions rather than isolated projects (UNDP, 2022). Although, at the time of the rollout, the official UNDP Evaluation Guidelines (UNDP, 2021) did not prescribe distinct methods for evaluating portfolios, they have opened space for methodological experimentation. Central to this is the Systems Learning Framework, which emphasizes adaptive management, feedback loops, and distributed learning.

An illustrative application is, for instance, the evaluation of the Mayors for Economic Growth (M4EG) initiative, implemented across Armenia, Azerbaijan, Georgia, Moldova, and Ukraine with funding from the European Commission (UNDP, 2024). M4EG supports municipalities to navigate complex local development challenges within a region affected by conflict, instability, and unequal access to resources (European Commission / UNDP, n.d.). Rather than assessing individual projects, the systems lens foregrounds locally embedded solutions and integrated approaches to urban development. The evaluation combined systemic inquiry with Monitoring, Evaluation and Learning (MEL) practices, drawing on both UNDP standards and OECD-DAC criteria.

The evaluation aimed to understand the added value of the portfolio approach while also examining contextual constraints. It took a learning-focused approach, drawing on single-, double-, and triple-loop learning, and used narrative analysis to understand how participants made sense of change. The findings indicate that the portfolio approach helped municipalities test systems-based methods alongside existing work, supported learning at

multiple levels, and strengthened peer-to-peer exchange.

The systems approach also prompted a rethinking of conventional planning tools, such as feasibility studies, transition plans, and SWOT analyses, by placing greater emphasis on local resources, iterative learning, and adaptive management. Sense-making workshops and social listening practices were identified as particularly useful for helping cities and local facilitators recognize systemic drivers of their challenges. This approach challenged the traditional intervention logic / ToC not only for municipalities but also within UNDP operations, creating space for more participatory and multidimensional planning practices. [The M4EG Systems Learning Framework (UNDP, 2024, p. 45) provides a visual illustration.]

However, this type of systems approach can also encounter significant challenges. In the reported case, the shift required time and investment in capacity development. Success depended on iterative cycles of reflection, discussion, and coaching, often facilitated in person. These processes were essential for strengthening evaluative thinking and reflexive capacity. The evaluation revealed that first- and second-order learning were frequently evident, while third-order learning, i.e., the ability to question underlying assumptions about one's own learning processes, remained limited.

The challenges in applying systems approaches in localized context were also related to language barriers, which affected accessibility. Systems terminology introduced in English was not always paired with specialist translation, slowing comprehension in some settings. As a result, translation and localization of technical language were identified as key enablers of inclusive systems learning. Moreover, motivation and participation depended on the balance between learning activities and visible infrastructure investments, revealing a tension between experimenting with systems approaches and delivering familiar development outputs.

Overall, the portfolio approach shows how a systems perspective can reorient development and evaluation practice through systemic sense-making, adaptive action, and community-level experimentation, while also demonstrating that such approaches depend on time, translation, capacity building, and sustained relational work.

## How the Cases Applied Systems Thinking Across the Evaluation Cycle

This section shows how the framework elements were addressed across the three case examples organized around the stages of the evaluation life cycle: before the evaluation begins (pre-evaluation), during the evaluation, and after the evaluation concludes (follow-up and use).

### *Before an Evaluation Begins (Pre-Evaluation)*

The elements below relate to the Before an Evaluation Begins stage, showing how the cases approached framing and design.

Table 5. Pre-Evaluation Phase Elements Compared Across Cases

| Element                                              | RE4R                                                                                                                    | Healthy Families NZ                                                                                   | UNDP portfolio approach                                                                                                                             |
|------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Understanding systems and complexity              | Systems mapping predated the evaluation and was further developed during it.                                            | Embraces systems thinking and kaupapa Māori to capture systemic health determinants.                  | Applied iterative systems tools (e.g., sense-making) to enable adaptive management.                                                                 |
| 2. Stakeholder ownership and governance              | Collaborative engagement with Practical Action during design, but less so during analysis due to humanitarian workload. | Non-extractive practice with communities as coanalysts and sense-makers; power shared “community-up.” | Broad involvement and continuous learning mechanisms that supported ownership, e.g., sense-making.                                                  |
| 3. Terms of reference (ToR) and question formulation | ToR largely predefined but aligned with strong transformational intent.                                                 | ToR and questions codeveloped with iterative revisions through participatory processes.               | Guided by UNDP evaluation standards embedding adaptability and scalability.                                                                         |
| 4. Methodology                                       | CIF’s transformational framework adapted to context with stakeholder input.                                             | Blended system mapping, co-design, Māori frameworks, Meadows’ systems conditions.                     | Learning-oriented, portfolio-level inquiry combining participatory MEL, narrative analysis, and sense-making within an adaptive learning structure. |
| 5. Theory of change (ToC) development and use        | ToC reviewed for evaluability; not reframed but complemented with mapping and rubrics.                                  | Framework captures direction and quality of change, grounded in Indigenous knowledge.                 | Conventional ToC challenged and replaced by portfolio logic reflecting dynamic systems.                                                             |
| 6. Procurement and evaluator selection               | Independent selection prioritized complex systems expertise.                                                            | Open tender prioritized systems thinking, kaupapa Māori, mixed methods, and sector knowledge.         | Systems knowledge required alongside thematic and regional expertise.                                                                               |

Cross-case patterns across the pre-evaluation phase reveal that cases consistently began with some form of systems framing, whether through

systems mapping, grounding analysis in local knowledge systems (e.g., kaupapa Māori in New

Zealand), or using UNDP's portfolio-based learning tools to support adaptive management.

Stakeholder involvement was similarly central, with efforts to embed participation, power-sharing, and local insight through co-design, collaborative governance, and inclusive question formulation. Participation included user voices and affected groups, although depth varied. Governance involved steering committees, collaborative leadership, and use of participatory tools such as rubrics and evolving ToCs. In Healthy Families NZ, power-sharing was deliberate.

A systems lens shaped question framing and ToR design, though with different levels of explicitness. Where ToRs were codeveloped by commissioning bodies and evaluation teams, they embedded equity, values, and systems orientation from the start. Systems mapping was foundational, providing an entry point to complexity. In practice, the ways in which complexity was understood varied: Systems mapping was deepened over time in some cases; others grounded analysis in local knowledge systems (e.g., kaupapa Māori in New Zealand); and UNDP applied portfolio-based learning tools to support adaptive management. Methodologies in all three cases blended theory-based and adaptive approaches, often combining realist

Table 6. Evaluation Implementation Elements Compared Across Cases

| Element                          | RE4R                                                                                                                                                                                                             | Healthy Families NZ                                                              | UNDP portfolio approach                                                                          |
|----------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| 7. Fieldwork and data collection | Remote data collection focused on interviews with implementing agencies, donors, and technical partners, complemented by validation workshops to test emerging system insights.                                  | Māori frameworks and community engagement shaped mixed field and remote methods. | Field and remote data integrated through regular MEL within portfolio learning.                  |
| 8. Analysis and interpretation   | Analysis applied the CIF Transformational Change Learning Partnership framework and structured rubrics, triangulating interview data, document review, and systems mapping to assess patterns of systems change. | Layered analysis combining expert assessment and use of Māori frameworks.        | Systemic inquiry combined with MEL practices, OECD-DAC criteria, and UNDP evaluation guidelines. |

evaluation, contribution analysis, kaupapa Māori, participatory action research, outcome harvesting, and co-design. Some incorporated system-level data sources such as remote sensing and geospatial data.

Across the cases, ToCs were mobilized differently to reflect systemic pathways. Some relied on conventional ToCs complemented by mapping and rubrics; others used Indigenous or portfolio frameworks capable of capturing direction and quality of systems change. In all three cases procurement and evaluator selection emphasized the need for systems expertise, though emphasis varied across cases. Some tenders foregrounded competency in both systems practice and technical domains (e.g., economic evaluation, kaupapa Māori), while others emphasized mission-oriented innovation and regional experience.

### *During the Evaluation*

During implementation, the evaluations combined remote and in-person data collection, structured and culturally grounded analytical approaches, and ongoing validation with stakeholders.

|                          |                                                                                                                                                                                              |                                                             |                                                                           |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|---------------------------------------------------------------------------|
| 9. Validation and review | Findings and rubric-based judgments were reviewed and refined through validation workshops with Practical Action and key stakeholders, testing interpretations and system-level conclusions. | Communities as coanalysts; narrative shapes interpretation. | Regular consultations and validation workshops with diverse stakeholders. |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|---------------------------------------------------------------------------|

Fieldwork methods reflected different balances between depth and practicality. Some relied heavily on remote data because of operational constraints, while others wove regular monitoring, place-based fieldwork, and Indigenous knowledge practices into system analysis. Structured frameworks and reflective processes were central to analysis—the use of rubrics and transformational frameworks in the case of RE4R, and the use of multi-layered approaches in HFNZ. UNDP blended systemic inquiry and OECD-DAC criteria and official UNDP evaluation guidelines (UNDP, 2021).

Validation was central across cases and took place through ongoing consultation, coanalysis, and narrative sense-making.

### *After the Evaluation Concludes (Follow-Up and Use)*

Across the three evaluations, follow-up processes advanced learning in different ways: RE4R shaped leadership decisions and reused methods; Healthy Families NZ deepened system learning through community-led leadership; and the UNDP portfolio approach applied findings to guide follow-up, inform adaptation, and support scaling across portfolios. Even where constraints limited design ambition, collaborative sense-making and validation helped embed learning and shift power.

Table 7. Post-Evaluation Elements Compared Across Cases

| Element                                 | RE4R                                                                                                                                                                                                                                                                                                        | Healthy Families NZ                                                      | UNDP portfolio approach                                                    |
|-----------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------|----------------------------------------------------------------------------|
| 10. Follow-up and use                   | Findings informed leadership strategy and learning, with the learning brief disseminated to other humanitarian actors. In addition, the Practical Action evaluation manager confirmed that capacity built through the rubric-based approach has been applied in other evaluations and assessment processes. | Shared leadership strengthens community aspirations and system learning. | Findings used for learning, follow-up, and scaling across portfolios.      |
| 11. Communication and knowledge sharing | Rubrics supported clearer communication and learning, with a sector-targeted learning brief                                                                                                                                                                                                                 | Storytelling shifted thinking from linear to systemic and relational.    | Two-way communication across formats and languages throughout the process. |

| Element | RE4R                                                                                                                                                  | Healthy Families NZ | UNDP portfolio approach |
|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|-------------------------|
|         | delivered as part of the project's original theory of change and communication plan to effect systems change by contributing to sector-wide learning. |                     |                         |

Communication played a crucial role in embedding learning and reinforcing systemic understanding. Rubrics provided clarity in some cases; Indigenous storytelling and narrative repositioned power in others; in multilingual contexts, translation and adapted formats ensured accessibility. Two-way communication before, during, and after evaluation was key to uptake and shared understanding.

## Insights to Carry Forward: Lessons for Future Systems-Oriented Evaluation

Drawing on the cross-case analysis, this section highlights practical lessons that readers can take forward in their own evaluation work. Across contexts, systems approaches are gaining traction, yet their depth and consistency—and the extent to which that traction is sustained—is informed by the institutional and operational context in which evaluated interventions are delivered, including constraints, time, capacity, and willingness to share power. Applied to the case studies, this analytic heuristic helps tease out practices that strengthen systemic inquiry and reveals where further development is needed to embed systems thinking more meaningfully.

### *Understanding Systems as Complex Adaptive Systems*

Evaluations were most insightful when systems were understood as relational and dynamic rather than static. Iterative collaboration and system mapping surfaced feedback loops, contextual influences, and emerging shifts. Investing early in shared system understanding helps ground the evaluation in what actually drives change, not in what is easiest to measure.

This was most visible in the RE4R and Healthy Families NZ cases, where explicit system framing shaped both evaluation design and interpretation, while in the UNDP portfolio case systems understanding evolved iteratively but

unevenly, constrained by capacity and translation challenges.

### *Ownership and Stakeholder Governance*

Broad participation—especially when it included those with lived or local knowledge as in the Maori case—strengthened legitimacy and helped direct attention to the conditions that support or hinder change. Where power-sharing was intentional, systems change was understood from the perspectives of those who experience it, not only those who manage programs.

Deep ownership was most evident in the Healthy Families NZ evaluation, where communities acted as coanalysts throughout. In contrast, RE4R demonstrated strong collaboration during design but more limited engagement during analysis, reflecting operational pressures common in humanitarian settings.

### *Conceptualizing ToR and Evaluation Questions*

Framing change as multilevel supported questions that better reflected how systems evolve. Designing ToRs around transformational relevance, adaptive learning, and feedback loops helped manage expectations about what evaluation can responsibly claim and how change might unfold beyond the intervention cycle. While codeveloped ToRs enabled adaptive questioning in Healthy Families NZ, RE4R illustrates that even predefined ToRs can support systemic inquiry when transformational intent is made explicit.

### *Theory of Change Development and Use*

ToCs become instruments for learning only when they are revisited, questioned, and adapted as systems shift. Across all three cases, ToCs were strongest as sense-making devices at the design

stage; fewer examples showed sustained, adaptive use during implementation, pointing to an area where systems-informed practice remains underdeveloped.

### *Developing the Methodology*

Blending theory-based approaches with participatory and adaptive tools produced more nuanced understanding of nonlinear change. Remote engagement widened reach, but face-to-face interaction remained indispensable for building relational insight and strengthening validity in complex contexts.

Blended methodologies were evident across all cases, though their depth varied—from rubric-based synthesis in RE4R, to mixed-method, Indigenous-informed designs in Healthy Families NZ, to learning-oriented sense-making in the UNDP portfolio.

### *Procurement and Evaluator Selection*

Systems-oriented evaluation demands specific capabilities—mapping complexity, working with emergence, coproducing evidence with communities, and applying theory-based causal reasoning. Procurement processes that recognize and resource these competencies are pivotal. Several cases suggest that capacity development is needed for both commissioners and evaluators. From the case examples in this article, Healthy Families NZ demonstrates what becomes possible when systems and Indigenous competencies are explicitly specified in procurement, while other cases reveal how commissioning practices can limit systemic ambition even where intent exists.

Due to the influencing and commissioning power of evaluation commissioners, creating a mind shift within commissioning practices is an important capacity development task. Evaluators may positively contribute toward this.

### *System-Informed Analysis and Interpretation*

Structured systemic lenses (e.g., using the transformational change criteria in Table 3) enabled more coherent synthesis without oversimplifying complexity. Regular sense-making with stakeholders reduced the risk of fragmented findings and helped evaluations capture direction, quality, and durability of change.

While structured lenses supported coherence across cases, only limited evidence was found of third-order learning, particularly in the UNDP portfolio case, highlighting the difficulty of sustaining methodological reflexivity.

### *Validation Processes*

Validation was most meaningful when embedded throughout the evaluation cycle rather than staged at the end. Continuous involvement of affected actors reinforced credibility, supported co-owned interpretation, and enabled early course correction. Validation was most influential where it functioned as continuous co-interpretation, as in Healthy Families NZ, rather than as a discrete stage, as seen more clearly in RE4R.

### *Follow-Up and Use*

Follow-up was strongest where evaluations were explicitly designed for learning rather than accountability, as illustrated by RE4R's sector-facing learning brief and the UNDP portfolio's scaling processes.

Goldman and Pabari (2020) point to the need to build the motivation, capability, and opportunity for decision makers to use evidence. Evaluations that explicitly aimed at supporting learning and adaptive management had greater influence. Building capacity through "learning by doing" helped ensure that systemic practices persisted beyond the life of the evaluation itself.

### *Communication*

Communication proved most powerful when treated as an ongoing relational practice rather than a final deliverable. Plain-language insights, visual storytelling, local translation, and interactive formats supported wider uptake and helped embed evaluation knowledge across systems.

While diverse communication formats supported uptake, the UNDP portfolio case illustrates how challenges to translate unfamiliar systems terminology can constrain inclusion and learning.

These lessons invite evaluators, commissioners, and practitioners to see systemic evaluation as a way of working that distributes power, foregrounds learning, and nurtures change over time.

## Conclusion

This article argues that integrating systems thinking enhances conventional evaluation by foregrounding interrelationships, multiple perspectives, boundaries, feedback, adaptation, and context (Horizon 1, current reality). At the same time, the field is still evolving and is even at a crossroads. Capacity limitations, compressed timelines, commissioner focus on isolated interventions, procurement practices that overlook systems competencies, and inconsistent inclusion of Indigenous and local knowledge constrain what evaluation can surface. These are design conditions to work with, not reasons to retreat.

We see from cutting-edge thinkers and activists how much they are thinking of systems approaches but also that a great deal is needed to translate this into practice. The cases provide examples that shifts are possible in practice (Horizon 2, transitional experiments): shared ownership and power-sharing, theory-based approaches for complex causality, adaptive indicators, structured sense-making (e.g., rubrics), and diversified communication tailored to different users. Where these shifts may occur, evaluations could generate more credible and usable learning.

The framework developed in the article (Table 2) translates this into actionable elements across the evaluation life cycle—serving as a heuristic checklist to analyze how systems principles are pursued prior to, during, or after evaluation; across evaluation design, commissioning, methodology, analysis, validation, communication, and follow-up.<sup>2</sup>

This article thus proposes Horizon 3's practical transformational agenda:

- Start upstream, during planning and commissioning.
- Reframe systems dimensions, including boundaries, perspectives, interrelationships, etc., and causality through theory-based and participatory sense-making, notably involving users and those affected by interventions.
- Build capacities for commissioners, evaluators, and users through learning by doing.

- Specify systems competencies and signal openness to innovation in terms of reference and procurement.
- Use adaptive indicators and staged reflection throughout the evaluation cycle.
- Resource ongoing communication, validation, and follow-up as core activities, especially with users and those affected and especially marginalized voices. Communicate through diverse and accessible formats in appropriate languages.

In addition to the implications for evaluation practice, a research agenda is also emerging, and some efforts are taking place by the evaluation community. For example, the European Evaluation Society, IEAc, and other partners are collaborating to build on the lessons discussed in this article. This is likely to be part of work that a global partnership is taking forward including these partners, EvalPartners, International Organization for Collaboration in Evaluation (IOCE), the regional evaluation associations, IDEAS and the EvalforEarth Community of practice.

Ultimately, evaluation faces a transformational imperative: to strengthen learning and sense-making amid the rapid degeneration of ecological systems, rising inequality, authoritarianism, and current and future conflict. Evaluative practice needs to understand these complex systems, the causes of decline, and how we can contribute to the regeneration needed. Embedding systems thinking throughout the intervention cycle positions evaluation to contribute to decisions that are more transformative, adaptive, inclusive, and durable (Horizon 3, the aspirational future). This requires humility: recognizing partial perspectives, valuing diverse ways of knowing, and working with rather than against complexity. In doing so, evaluation moves beyond judging performance and becomes a shared practice of emergent learning and collective stewardship in changing systems toward health and life.

## References

- Aston, P., & Apgar, M. (2022). Building confidence in evaluative judgements: Methodological bricolage and contribution analysis in complex contexts. *Evaluation*, 28(4), 481–499.

<sup>2</sup> To assist practical implementation, Appendix 2 provides a version of the framework with detailed questions that can be applied at each stage.

- Beach, D., & Pedersen, R. B. (2013). *Process-tracing methods: Foundations and guidelines*. University of Michigan Press.
- Befani, B., & Mayne, J. (2014). Process tracing and contribution analysis: A combined approach to generative causal inference for impact evaluation. *IDS Bulletin*, 45(6), 17–36. <https://doi.org/10.1111/1759-5436.12110>
- Befani, B. (2016). Testing contribution claims with Bayesian updating. *CECAN Evaluation and Policy Practice Note No. 2.1*. Centre for the Evaluation of Complexity Across the Nexus. <https://www.cecan.ac.uk/sites/default/files/2018-01/BARBARA%20v2.5.pdf>
- 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(4), 407–423. <https://doi.org/10.1006/jema.2001.0493>
- Binder, C. R., Hinkel, J., Bots, P. W. G., & Pahl-Wostl, C. (2013). Comparison of frameworks for analyzing social-ecological systems. *Ecology and Society*, 18(4). <https://doi.org/10.5751/es-05551-180426>
- Buckton, S. J., Fazey, I., Ball, P., Ofir, Z., Colvin, J., Darby, M., Hejnowicz, A. P., Leicester, G., Newman, R., Page, G. G., Parsons, K., & van Mierlo, B. (2025). Twelve principles for transformation-focused evaluation. *PLOS Sustainability and Transformation*, 4(4), e0000164. <https://doi.org/10.1371/journal.pstr.0000164>
- Chaplowe, S., & Gupta, T. (2022). *Final report: Independent evaluation (systems impact assessment) of the Renewable Energy for Refugees project*. Practical Action and UNHCR.
- Climate Investment Funds (CIF). (2021). *Transformational change concepts: Working definition, dimensions, questions and signals*. Retrieved from <https://www.cif.org/knowledge-documents/transformational-change-concepts-working-definition-dimensions-questions-and>
- Coffman, J. (2007). *A framework for evaluating systems initiatives*. Harvard Family Research Project.
- Darnhofer, I., Gibbon, D., & Dedieu, B. (2017). Towards a complexity-aware theory of change for participatory research programs. *Agricultural Systems*, 153, 36–45.
- Department for Environment, Food and Rural Affairs (Defra). (2020). *Complexity evaluation framework: Recognising complexity and key considerations for complexity-appropriate evaluation* (Version 2.0, updated by Steer Economic Development). <https://www.betterevaluation.org/tools-resources/complexity-evaluation-framework-recognising-complexity-key-considerations-for-complexity-appropriate>
- Douthwaite, B., & Hoffercker, E. (2017). Towards a complexity-aware theory of change for participatory research programs working within agricultural innovation systems. *Agricultural Systems*, 155, 88–98. <https://doi.org/10.1016/j.agry.2017.04.002>
- European Commission / UNDP. (n.d.). *Mayors for Economic Growth (M4EG)*. <https://eum4eg.com>
- Garcia, J. R., & Uitto, J. I. (2025). Using a systems lens in evaluations at the nexus of environment and development. *Evaluation and Program Planning*, 112, 102560. <https://doi.org/10.1016/j.evalprogplan.2025.102560>
- Geels, F. W., & Schot, J. (2007). Typology of sociotechnical transition pathways. *Research Policy*, 36(3), 399–417.
- Goldman, I., Cairney, P., Carden, F., Dias, T., Dillon, B., Ferrari, R., & Morgan, G. (2026). Futures, Feedback, and Flexibility: Reimagining Public Sector Management and Evaluation in the Time of Polycrisis. *Journal of MultiDisciplinary Evaluation*, 22(53), 233–254.
- Goldman, I., & Pabari, M. (2020). Lessons for using evidence in policy and practice. In I. Goldman & M. Pabari (Eds.), *Using evidence in policy and practice: Lessons from Africa*. Routledge. <https://www.taylorfrancis.com/books/e/9781003007043>
- Ison, R., Blackmore, C., & Collins, L. (2021). Fostering complexity thinking in action research for change in social-ecological systems. *Systemic Practice and Action Research*, 34(1), 1–19.
- Latham, N. (2014). Pathways to system change: The evaluation of complex systems. *Evaluation*, 20(1), 60–77.
- Leeuw, F. L., & Furubo, J.-E. (2008). Evaluation systems: What are they and why study them? *Evaluation*, 14(2), 157–169. <https://doi.org/10.1177/1356389007087537>
- Matheson, A., Gray, R., Wehipeihana, N., et al. (2020). *Evaluation of Healthy Families NZ*:

- Interim report 2020*. Te Herenga Waka – Victoria University of Wellington, Wellington.
- Mingers, J., & Gill, A. (Eds.). (1997). *Multimethodology: Theory and practice of combining management science methodologies*. John Wiley & Sons.
- Morell, J. A. (2010). *Evaluation in the face of uncertainty: Anticipating surprise and responding to the inevitable*. Guilford Press.
- Morell, J. A. (2025). Characteristics of sustainability: A complexity-based view to inform evaluation. *Evaluation and Program Planning*, 111, 102561.
- Morell, J. A., Hilscher, R., Magura, S., & Ford, J. (2010). Integrating evaluation and agent-based modeling: Rationale and an example for adopting evidence-based practices. *Journal of MultiDisciplinary Evaluation*, 6(14), 32–57. <https://doi.org/10.56645/jmde.v6i14.275>
- Ndhlovu, L., Smith, L., & Narsoo, S. (2017). The search for a method to unpack urban complexity: Case study of the City of Johannesburg. *African Evaluation Journal*, 5(1), 1–11. <https://doi.org/10.4102/aej.v5i1.195>
- Ofir, Z. (2021). Evaluation in transition: The promise and challenge of South-South cooperation. *Canadian Journal of Program Evaluation*, 36(1), 18–38. <https://doi.org/10.3138/cjpe.71630>
- Ofir, Z., Kendrick, I., Goldman, I., Ferretti, S., Norins, J., Sithole, S., Belzer, A., Chaplowe, S., Diwakar, Y., Kalugampitiya, A., Lareza, A. E., Verveld-Nguyen, L., Zaveri, S. (2026). Where Next for Evaluation? Navigating the Polycrisis Through Three Horizons. *Journal of MultiDisciplinary Evaluation*, 22(53), 12–40.
- Patton, M. Q. (2010). *Developmental evaluation applying complexity concepts to enhance innovation and use*. Guilford Press. [http://www.guilford.com/cgi-bin/cartscript.cgi?page=pr/patton.htm&dir=research/res\\_eval&cart\\_id=519053.25995](http://www.guilford.com/cgi-bin/cartscript.cgi?page=pr/patton.htm&dir=research/res_eval&cart_id=519053.25995)
- Patton, M. Q. (2011). A complexity-based plan for evaluating transformation. *Evaluation*, 17(4), 373–389.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage Publications.
- Reynolds, M., Gates, E., Hummelbrunner, R., Marra, M., & Williams, B. (2016). Towards systemic evaluation. *Systems Research and Behavioral Science*, 33(5), 662–673. <https://doi.org/10.1002/sres.2423>
- Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., White, H., & Helbing, D. (2009). Economic networks: The new challenges. *Science*, 325(5939), 422–425.
- Snowden, D. J., & Boone, M. E. (2007). A leader's framework for decision making. *Harvard Business Review*, 85(11), 68–149.
- Systems in Evaluation Topical Interest Group (SETIG). (2018). *Principles for effective use of systems thinking in evaluation practice*. American Evaluation Association. <https://www.betterevaluation.org/sites/default/files/SETIG-Principles-FINAL-DRAFT-2018-9-9.pdf>
- Shipurkar, T. J., Narayanan, P., Manjutha, C., Ponnuchamy, M., Khan, N., & Joseph, A. R. (2026). Measuring Collectivization Initiatives: Creating a Knowledge Base from Collective Voices. *Journal of MultiDisciplinary Evaluation*, 22(53), 255–279.
- UNDP. (2021). *UNDP evaluation guidelines*. [https://erc.undp.org/pdf/UNDP\\_Evaluation\\_Guidelines.pdf](https://erc.undp.org/pdf/UNDP_Evaluation_Guidelines.pdf)
- UNDP. (2022). *System change: A guidebook for adopting portfolio approaches*. <https://www.undp.org/publications/system-change-guidebook-adopting-portfolio-approaches>
- UNDP. (2024). *Mayors for Economic Growth: Final evaluation report*. <https://erc.undp.org/evaluation/documents/download/23754>
- UNDP. (n.d.). *Systems portfolio resource hub*. <https://www.undp.org/systems-portfolio>
- van den Berg, R. D., & McCollum, D. (2021). Meta-analysis of climate mitigation evaluations. *Evaluation and Program Planning*, 85, 101910.
- Williams, B., & Hummelbrunner, R. (2010). *Systems concepts in action: A practitioner's toolkit*. Stanford University Press.
- Wilson-Grau, R., & Britt, H. (2012). *Outcome harvesting*. Ford Foundation. <https://www.betterevaluation.org/tools-resources/outcome-harvesting-o>
- Zaveri, S., Goldman, I., Norins, J., Ghiano, C., Kotschy, K., & Soal, S. (in press). *How can evaluations support a socially just transition to planetary health?* In G. Batra, J. I. Uitto, & M. A. Spearman (Eds.), *Integrating for sustainability: Evaluation across environmental and socioeconomic domains* (Chapter 11). Springer.

## Appendix 1: Methodological Implementation and Data Sources

### *Systematic Literature Review*

An extensive literature review was undertaken using advanced search and screening tools, including AI-assisted platforms (ChatGPT, Perplexity, ELICIT, PETAL) to query Scopus, Web of Science, PubMed, JSTOR and Google Scholar. Two categories structured the search strategy:

- A. Analytical frameworks for understanding systems and complex change.
- B. Methodological approaches that operationalize systems thinking within evaluation practice.

Keywords included “systems evaluation framework,” “complexity-aware evaluation,” “systems mapping,” “network analysis,” and “participatory systems evaluation.” The search identified 80 core sources, supplemented by 25 expert-recommended publications. AI tools assisted with organizing and screening the corpus. However, all interpretive decisions, citation checks, and inclusion choices were made by the researcher team. Targeted follow-up searches addressed gaps identified during screening. Key findings from this process informed Table 1.

### *Analysis of Three Horizons (3H) Practitioner Responses*

To complement the literature review, we analyzed 39 responses from a 3H survey with 13 in-depth questions, sent to over 100 carefully selected individuals working in the evaluation field, as policy advisors using evaluation evidence, and as professionals and academics working in the public management, civil society, and transformation fields, across all regions (see Ofir et al., 2026 for methodology). The 3H survey follows the 3H methodology, an approach to use when the future is very uncertain but there is a level of agency to bring about a preferred future. It is based on the current situation of evaluation and its ability to contribute to the polycrisis (Horizon 1), an aspired future (Horizon 3), and how to get there (Horizon 2). More details on the Three Horizons methodology are covered in another article within this special edition (Ofir et al., 2026).

Systems thinking was not explicitly asked about in the survey but occurred frequently in respondents’ narratives. An initial thematic analysis of the responses was undertaken by an external researcher, supported by AI as an organizational tool to segment narrative components and group recurring concepts. A follow-up round of analysis explored gaps, tensions, and possible directions for systems-oriented evaluation.

### *Development of a Framework*

We used the literature and responses from the Three Horizons survey to develop a heuristic framework to show evaluation practitioners how systems thinking could be incorporated during different stages of the evaluation cycle: conceptualizing and designing the evaluation, implementing the evaluation, and follow-up. This was then applied and tested on the case examples.

### *Exploring Application of the Framework in Case Examples*

To identify applications in evaluation practice, we undertook a structured search for evaluations applying systems approaches. Searches across Google Scholar, Scopus, Web of Science, PubMed, JSTOR, SpringerLink, ProQuest Dissertations, and Policy Archive were complemented by outreach through professional networks (including LinkedIn) and direct contact with systems evaluation practitioners. This helped identify less visible but instructive practice examples.

From this process, six potential cases were shortlisted based on:

- diversity across geographic and institutional contexts,
- alignment with the framework developed in this article, and
- willingness of case proposers to contribute as coauthors to this article, contributing their own rich accounts of their experience of the case.

Three diverse cases met these criteria and were codeveloped with the practitioners involved in the evaluations as case authors for this article. The cases examined were:

1. Transformational change in humanitarian energy systems (RE4R Systems Impact Assessment),

2. Indigenous-led systems thinking in Healthy Families NZ, and
3. A systems portfolio approach within UNDP.

We then tested out the framework on these cases.

## Appendix 2: Guiding Questions for Applying the Conceptual Framework

Below are a set of guiding questions designed to help users apply the conceptual framework presented in Table 2.

| Element                                              | How principles and approaches are applied in practice                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Understanding systems and complexity              | <p>Is the evaluation focused on systems or interventions?</p> <p>How does this distinction affect the choice of approaches and tools?</p> <p>How were the system and its complexity understood and addressed in the evaluation, and what value did this bring?</p>                                                                                                                                                                                                                                                                                                                                        |
| 2. Stakeholder ownership and governance              | <p>Which stakeholder groups were engaged (not just program staff or funders, but those with experiential knowledge—community members, line workers, users, and those affected by system cycles)?</p> <p>How were stakeholders involved in governance and ownership of the evaluation?</p> <p>How did this affect the process and outcomes?</p> <p>What governance or decision-making role did they have?</p> <p>How were power dynamics addressed to ensure meaningful participation?</p>                                                                                                                 |
| 3. Terms of Reference (ToR) and question formulation | <p>Who contributed to defining the evaluation questions?</p> <p>How was systems thinking integrated into the formulation of ToR and evaluation questions? What systems methods and tools were used (e.g., systems mapping or alternatives)?</p> <p>How did ToR reflect the ToC from a systems perspective? (e.g., conceptualizing beyond linear, cause-effect models to embrace feedback loops, emergence, and context-specific change)</p>                                                                                                                                                               |
| 4. Theory of Change (ToC) development and use        | <p>How did the evaluation's methodology and, where applicable, its Theory of Change reflect a systems perspective (including theory-based or emergence-led approaches) to understanding system change? Consideration of theory-based approaches like contribution analysis, process tracing, and realist evaluation. Inclusion of tools to capture evolving, nonlinear change and center stakeholder voices, like participatory action research (PAR), and outcome harvesting</p> <p>Analyzing large-scale, dynamic systems influences e.g. through use of geospatial and remote sensing data to map.</p> |
| 5. Methodological design                             | <p>How did the evaluation's methodology (with or without a formal ToC) reflect a systems perspective?</p> <p>Was the ToC system-focused and integrated into the overall planning process?</p> <p>How were tools and data sources chosen to capture evolving, nonlinear change?</p>                                                                                                                                                                                                                                                                                                                        |
| 6. Procurement and evaluator selection               | <p>Who gets involved in the process, what competences are looked for, including system competences?</p> <p>How were procurement and evaluator selection processes adapted to prioritize systems competencies?</p>                                                                                                                                                                                                                                                                                                                                                                                         |
| 7. Fieldwork and data collection                     | <p>How did system understanding influence the design and implementation of fieldwork?</p> <p>Did system framing influence sampling strategies, data collection methods, and timing?</p> <p>Was the evaluation approached as first-order (within the system) or second-order (reflecting on the system itself)?</p>                                                                                                                                                                                                                                                                                        |

| Element                                 | How principles and approaches are applied in practice                                                                                                                                                                                                                                                                                                               |
|-----------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 8. Analysis and interpretation          | <p>How does analysis in a systems approach differ from traditional evaluation, and what insights did this enable?</p> <p>How are feedback loops, interdependencies, and contextual dynamics examined?</p> <p>Are system-level influences explicitly analyzed alongside intervention-level effects?</p>                                                              |
| 9. Validation and review                | <p>Who is involved, and how much power do users and those affected have in this process?</p> <p>How inclusive and influential was the validation process in shaping the final findings?</p> <p>How did the process incorporate different types of knowledge (technical, lived experience, systemic)?</p>                                                            |
| 10. Follow-up and use                   | <p>Were there different forms of evaluation use than in conventional evaluations?</p> <p>How did follow-up processes support systemic learning, adaptation, and improvement?</p> <p>Were management responses and improvement plans framed with a systems perspective?</p> <p>How were mechanisms for ongoing reflection and adaptation embedded into practice?</p> |
| 11. Communication and knowledge sharing | <p>How was communication managed before, during, and after the evaluation to support shared learning and system change?</p> <p>How was communication two-way with stakeholders and those affected by the system?</p> <p>How were findings tailored to different audiences to promote understanding and action?</p>                                                  |