PRACTITIONERS' GUIDANCE TO ASSESSING SYSTEMS CHANGE

Drawing from the practical experience of MEL managers working on market systems development (MSD) programs around the world for a diverse set of donors, this guide aims to help other MEL managers find what works best for their program and teams in assessing systems change. The guidance is aided by the honest exchange from peers working in diverse contexts, team capacities, and mandates.
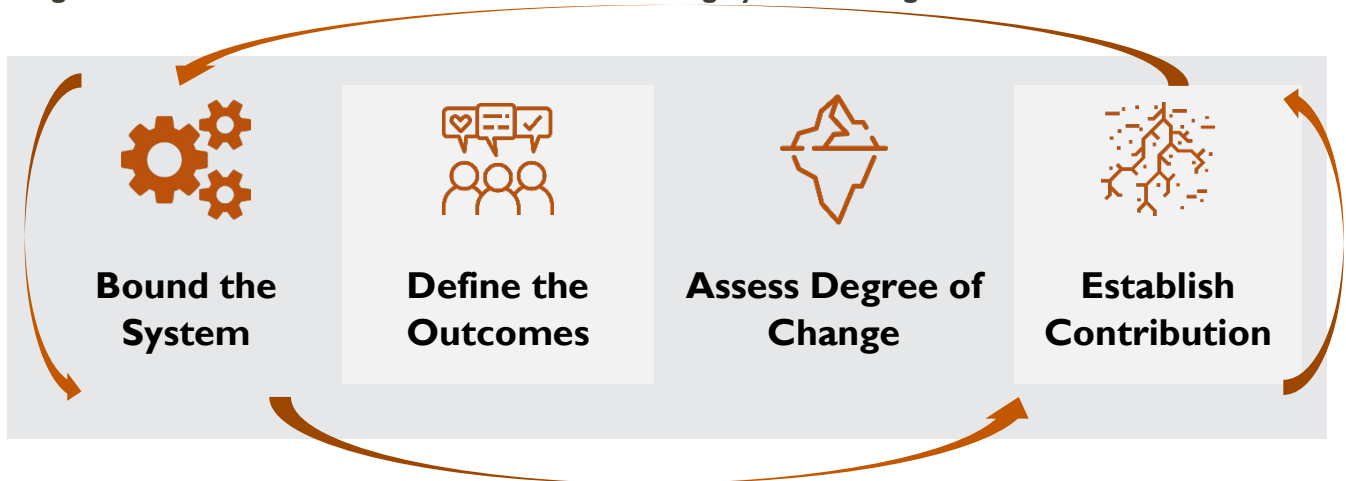
Rather than viewing systems change assessments as a backward-looking exercise to be conducted at the end of a program to justify impact, this guide asks, "*How might we assess systems change more frequently and effectively, generating feedback our teams need to better understand and catalyze change? Can this help us leave a more impactful legacy?*"

No one knows how to change complex systems. We must discover how to do it. Assessing systems change is the process of generating insights into what works to change systems.

How To Read This Guidance: This guidance is divided into four parts; Figure 1 below and continuing onto the following page, previews the content for each part. For each part, we list key points and common challenges faced by MEL managers. Then we provide a frank assessment of the pros and cons of common approaches used by practitioners. Finally, we make some recommendations and suggest practical actions for MEL managers to consider, including exercises, survey questions, analysis frameworks, and sense-making tools that be customized. The guide also uses a single case of an agricultural input system in a relatively thin market in sub-Saharan Africa to illustrate what application in real-life might look like. *This case example unfolds step-by-step in italics throughout the guide.*

The figure below and on the following page provide an overview of the four parts of the guidance and practical tips and resources to look out for.

Figure 1: Guide Overview: The Four Parts to Assessing Systems Change

Bound the System

Define the Outcomes

Assess Degree of Change

Establish Contribution

| Part	Guidance Highlights	What do you need?
Bound the System *What is the system that is changing?*	Be clear on the impact you want to achieve and then define the system in terms of how it might function differently to generate that impact. Consider multiple perspectives for why diverse actors engage in the system when defining function. Include in your boundaries, the actors (who) and factors (what) that enable or prevent the system from achieving this function or purpose. Boundaries should change as you engage and learn more about the system.	*Tips to define systems beyond sectors by function. p. 7* *Considerations for including actors in the system. p. 8* *3 diagnostic tests to find high-leverage factors in a system. p. 9*
Define the Outcomes *What parts of the system have changed?*	Articulate outcomes in well-defined, evaluable statements. Start with interventions and theorize the resulting outcomes, but also harvest outcomes from direct observation and team reflections. While behavior changes are important, changes in relationships, voice/agency, power dynamics, rules, diversity in composition, etc., are often more important. Outcomes should represent a certain scale of change to be systemic. Merge quantitative data with qualitative evidence in an effective storytelling format to effectively communicate systems change.	*8 types of outcomes to look for beyond behaviors. p. 12* *How to write well-defined outcome statements. p. 13* *7 common indicators and evidence used to quantify change. p. 15* *8 storytelling elements to communicate a change narrative. p. 16*
Assess Degree of Change *How significant are these changes?*	Assess outcomes by degrees of change across three criteria: scale, sustainability, and impact value. Establish rubrics to monitor and track progress against the expected degrees of change over time. Engage participants in the assessment of outcomes, whenever possible, to give credibility to the assessment and ensure the impacts of changes are meaningful to them. Reflect on each of the outcomes collectively and look for evidence that other parts of the system are changing, and the functionality of the system has improved to assess whole-of-system change.	*3 criteria to assess the degree of systems change. p. 19* *Indicators, survey questions, and rubrics to assess scale p. 21 and sustainability. p. 23* *Questions to gather participant feedback on change ('impact value'). p. 25* *A 3-test framework to pull it all together (whole-of-system change). p. 26*
Understand Contribution *What impact did the program have?*	Establish the causal links between the observed outcomes and the specific program interventions and include evidence to substantiate that theory with a degree of certainty. Make a clear and concise contribution claim and validate this case with key informants. The question we are asking is, *"Did the program help make the change happen; if so, to what extent?"* Exploring these causes behind outcomes and the program's contribution is critical to better understanding how to adapt our approach to bringing about systems change.	*3 tips to explain the context of change. p. 29* *How to weigh uncertainty. p. 32* *Template to develop contribution claims. p. 32* *Facilitation prompts for 2 team Pause and Reflect sessions. p. 34 and 35*

When should this guidance be used? This guidance focuses on assessing systemic change as an ongoing action done in implementation, as part of a regular monitoring and evaluation system, and not just at the end of the program. The four parts of this guidance are not intended as four steps to walk through linearly, one time each, over the life of the program. Rather, the more we cycle through these parts (or steps), the better we can track progress, identify areas for improvement, and increase our impact. However, we recognize that there are key moments in each program where there are opportunities to act. We offer some suggestions for integration entry points below.

Figure 2. Assessment as a Regular Cycle
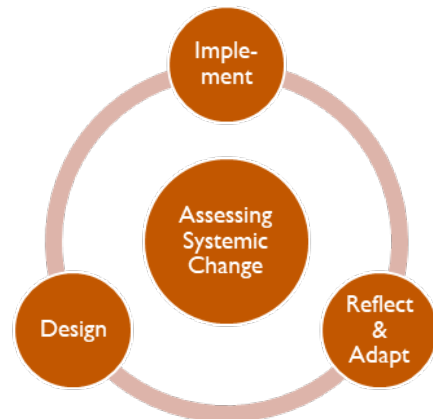
- Inception/MEL Plan Development: prioritizing the definition of the system boundaries (Part 1. Bounding the System) and the outcomes we want to achieve (Part 2. Defining the Outcomes) as part of our systems change strategy at the inception of a program, noting these will change over time.

- Implementation/Monitoring: getting teams into the routine of collecting evidence and updating outcome statements (Part 2) regularly. As we learn more about the system, we recommend defining the parameters for what we would expect to change (Part 3. Assessing Degree of Change).

- Annual Results/Evaluation: using annual (or equivalent) results surveys to gather evidence to make an updated assessment of the degree of systems change (Part 3). Following this assessment, we recommend identifying the program's role in change (Part 4. Establishing Contribution).

- Adaptation/Activity Design: holding annual reflection exercises with teams on systems change, drawing from guidance in the Conclusions section. We recommend integrating this learning and feedback into intervention design for the subsequent years.

- Repeat: repeating the process each year.

BACKGROUND

There is a growing recognition of the importance of systems approaches to advance more sustainable, scalable, and impactful change. While systems approaches initially emerged from the field of ecology, they are now recognized as a powerful framework for tackling pressing challenges facing society in a wide range of fields, including public health, education, and business management.

Systems change refers to changes that scale beyond a few individual parts to the entire system and includes profound shifts in how the whole system functions, including less-observable changes that endure across time horizons extending from years to decades.

As part of USAID-funded initiatives (see Figure 3) in 2022, six MEL practitioners came together to discuss experiences assessing systems change. This paper arose from these discussions.

Figure 3: The MSD in MEL Clinics

As part of its M&E and CLA learning strategy, the Feed the Future Market Systems and Partnerships (MSP) Activity hosted a series of virtual clinics to strengthen peer-learning networks between senior MEL staff leading innovative MEL systems on MSD programs. Thirteen specialists from 12 countries, participated in a series of interactive discussions organized into three thematic areas: (i) monitoring systemic change, (ii) MEL's role in building an adaptive culture, and (iii) facilitating learning in the private sector. Participant-authored briefs captured key insights, available here. The group associated with this paper represents full-time, program-based MEL leads and advisors working on MSD programs in Honduras, Mozambique, Uganda, Rwanda, DRC, and Indonesia, funded by USAID. DFAT. and Sida.

We acknowledge there is already a wealth of guidance on how to assess systems change, and many resources are referenced in this paper and in Annex 1. These include:
- Evaluating Complexity: Propositions for Improving Practice. FSG. [1]
- Systemic Evaluation Design: A Workbook. Bob Williams. [2]
- Guidelines for Monitoring, Evaluation, and Learning in Market Systems Development. USAID. [3]
- Guidelines to the DCED Standard for Results Measurement. DCED.[4]
- Making Markets Work for the Poor (M4P) Operational Guide. Springfield Centre.[5]
- Disrupting System Dynamics: A Framework for Understanding System Changes. MarketShare Associates for USAID[6]

Our experience is that we, as MEL managers, encounter similar challenges in assessing systems change. The contexts facing each program tend to be unique and require a set of decisions and problem-solving by MEL managers to adapt guidance to the diverse realities of implementation. Furthermore, it is evident from new initiatives, such as the UNDP Strategic Innovation Unit's M&E Sandbox and Beyond,[7] that the state of M&E practice is advancing quickly. As silos among often distinct fields break down, we can learn from what others are doing to build on and improve on previous guidance and application. This supportive and collaborative spirit of problem-solving brought us together to write this guidance.

---

[1] Preskill, H., Gopal, S., et al. (2015). Evaluating Complexity: Propositions for Improving Practice. FSG. Retrieved from https://www.fsg.org/resource/evaluating-complexity/.
[2] Williams, B. (n.d.). Systemic Evaluation Design: A Workbook. Retrieved from https://bobwilliams.gumroad.com/l/evaldesign.
[3] USAID LEO Brief #5. (2016). Guidelines for Monitoring, Evaluation and Learning in Market Systems Development. Retrieved from https://www.marketlinks.org/sites/default/files/resource/files/Report20No.205120-20Guidelines20for20MEL20in20MSD.pdf
[4] Donor Committee for Enterprise Development (DCED). (2018). Guidelines to the DCED Standard for Results Measurement. Retrieved from https://www.enterprise-development.org/wp-content/uploads/2_Implementation_Guidelines_Defining_Indicators.pdf
[5] Springfield Centre. (2014). Making markets work for the poor (M4P) operational guide. Retrieved from http://www.springfieldcentre.com/wp-content/uploads/2014/09/2014-09-M4P-Operational-Guide-with-watermark1.pdf
[6] MarketShare Associates for USAID (2016). Disrupting System Dynamics: A Framework for Understanding System Changes. Retrieved from https://www.marketlinks.org/resources/disrupting-system-dynamics-framework-understanding-systemic-changes
[7] United Nations Development Programme (UNDP) Innovation. 2023. Innovative M&E: From the Sandbox and Beyond. Medium. Retrieved from https://medium.com/@undp.innovation/innovative-m-e-from-the-sandbox-and-beyond-9234d0977796

GUIDING PRINCIPLES

We emphasize the importance of the unique context of each program and the need for flexibility and adaptability when assessing systems change. Principles help provide direction and clarity to navigate this ambiguity and uncertainty. At least three principles have proven helpful for us as MEL managers:

PRINCIPLE 1: WE MUST ENGAGE DIVERSE PERSPECTIVES AT EACH STAGE.

Our experiences and beliefs shape our understanding of systems change. Relying on a few perspectives leads to blind spots and biases in our change assessments. For example, a finance expert that looks at a complex system tends to see changes related to credit and banking. An inclusion expert that looks at a system tends to see changes related to gendered social norms and exclusion.

These perceptions of change, however accurate, are often incomplete. This limited understanding is why engaging diverse perspectives is vital to understanding and assessing the system as a whole more effectively. Moreover, engaging diverse perspectives improves the legitimacy and credibility of our assessments. *Who gets to decide if the change is systemic? Are they the right people?* [8]

This inclusive engagement builds trust, promotes accountability to our local partners, and ensures that the voices and needs of those impacted by the system are heard and considered part of the assessment.

PRINCIPLE 2: WE MUST ADAPT OUR ASSESSMENTS AS WE LEARN MORE ABOUT THE SYSTEM.

Our understanding of systems will improve as we learn by intervening. Teams should actively discover these (often hidden) conditions for systems change. [9] These discoveries and insights will and should shift our program focus and lead us to change our interventions in significant ways. Accordingly, our systems change assessments must also adapt to this learning.

Programs must strike a balance between flexibility and accountability in the assessment process. It is essential to ensure changes to outcomes and measures are made with care and intentionality and that they do not undermine the credibility or validity of the assessment. At the same time, the need to adapt is continuous. If we remain stuck with outdated MEL frameworks, MEL is no longer functional.

It is, therefore, critical to have clear and transparent processes for adaptive management with funders and partners for making changes to outcomes and ensuring that these changes are based on sound evidence, feedback from stakeholders, and a clearer understanding of the needs of the system.

PRINCIPLE 3: OUR ASSESSMENT MUST REGULARLY INFORM DECISION-MAKING.

There are only so many opportunities within a program lifecycle to get it right regarding transformational interventions. As much as we want to understand everything up front, even the most rigorous formative assessments will only partially grasp the problem and opportunities.

While there is a need for more ex-post assessments, adaptive management decisions still have to be made regularly, in the present, to answer critical questions like, "*Are we doing the right things to catalyze systems change?*" We strongly advocate for more regular self-assessments of systems change which are integrated into M&E systems and used in everyday adaptive decision-making. We advocate for teams to err on the side of more frequent but potentially less precise assessments and conduct these to inform decision-making while avoiding overwhelming teams, undermining local capacity, and taxing clients' patience.

---

[8] Williams, B. (n.d.). Systemic Evaluation Design: A Workbook. Retrieved from https://bobwilliams.gumroad.com/l/evaldesign.
[9] Kramer, M. R., Kania, J., & Senge, P. (2018). The Water of Systems Change. FSG.

PART 1. BOUNDING THE SYSTEM

KEY POINTS
- ✓ Bounding is defining who and what we consider to be part of the system.
- ✓ Later, when we assess how much the system has changed ([Part 3](#)), we need to engage multiple perspectives. Participation in this will require a degree of shared framing of the system.
- ✓ These boundaries *will* and *should* change over time as we learn more about the system.
- ✓ Teams can also better direct resources toward areas with significant potential for change by carefully considering who to partner with and what issues to focus on.

THE CHALLENGE

Defining boundaries for systems is a complex task. Take an agricultural input system. [10] It may be clear that input distributors and retailers are part of the system. There may be questions about whether farmers, regulators, transporters, and financial institutions are part of it as well.

These decisions around boundaries have significant implications for how we assess systems change. Take the assessment of scale. To understand scale, we would want to understand the proportion of a group that experiences a change in relation to a total population. Definitions change assessments.

In sum, moving away from ambiguous statements about system change is important to improve assessments. While there may not be a single, correct answer to boundaries, it is crucial to be explicit about our assumptions. This allows us to make more precise and valid assessments of change.

WHAT APPROACHES ARE PRACTITIONERS USING?

The key difference we see in approaches to bounding the system is whether to define the system by the value chain or sector of interest or bind the system by its function or purpose.

	Bound by Sector or Area	Bound by Function or Purpose
What	Bounding the system by sector or area is to define the system by the actors directly or indirectly involved in producing, processing, and delivering a specific product or service. This approach includes actors like service providers involved in supporting functions and actors like regulators involved in setting the rules.	Bounding the system by function or purpose allows us to define the system to include all actors (and factors) that contribute to the overall goal or purpose of the system. This approach considers all actors (or factors), regardless of sector or activity, so long as they interact as part of system function.
Pros	Many systems change programs are awarded to intervene in certain sectors or areas. This approach to boundaries aligns the assessment to those funding mandates. Also, system actors often self-organize by sector or area, making it easier to establish clear boundaries, providing a more manageable scope for assessment.	Bounding by function or purpose helps to ensure that relevant conditions for systems change are considered, even if they are not part of the same sector or area. This approach helps to broaden our perspective to have a more integrated and comprehensive understanding of systems change.
Cons	Sectoral boundaries may exclude important parts of systems. This can limit our ability to	Systems change programs may face practical constraints such as limited resources, time,

---

[10] This case is loosely based on the Feed the Future Inova Activity in Mozambique, but certain changes have been made to illustrate points.

	Bound by Sector or Area	Bound by Function or Purpose
	address issues or opportunities that may arise from interactions between different systems. For example, a maize system might not capture important changes needed in the broader financial system, such as the capacity of banks to reach underserved markets, including maize.	or political will, which make it difficult to take a more holistic approach. Bounding systems by function also requires higher levels of interdisciplinary collaboration and coordination, which is challenging to achieve in practice.
	Use caution when bounding by a specific sector or area. Do not mistake a change in one sector as a change that affects the entire system. Be wary of binding constraints and issues outside the sector or area of interest that limit impacts.	☑ **Recommendation** This approach provides a more complete understanding of the system and helps avoid overlooking essential parts of the system necessary for change.

RECOMMENDATIONS

Below are seven recommendations we found helpful to bounding the system.

1. Link the prioritized system function to desired impact.
2. Consider the alternative functions of the system.
3. Identify the actors who are considered part of the system.
4. Include the actors across the multiple expected levels of impact.
5. Discover the factors which limit the system from changing.
6. Prioritize the critical factors we will address directly.
7. Specify the geographies in which the change is expected to happen.

1.1 Link the prioritized system function to desired impact.

It's important to start with a clear understanding of the impact we want to achieve. The impact must be relevant to the needs of the target population group, for instance, job opportunities for youth not in school. Then, define the function or purpose of the system that will generate that impact for the target population group.

In the case of an agricultural input system, if we want to improve the productivity of smallholder farmers as desired impact, we might define the purpose or function of the input system to provide farmers with better access to quality inputs such as seeds and fertilizers, along with localized technical support.

1.2 Consider the alternative functions of the system.

Take time to define the function of a system considering multiple perspectives. People may view the system's function differently and engage in the system for different reasons, often with conflicting goals and interests. This can affect how they perceive change to the status quo and impact the effectiveness of systems change efforts. [11]

In the case of the agricultural input system, an input distributor and an environmental group advocating for organic production may have differing viewpoints. Both groups hold valid views regarding what constitutes a functional input system and would assess systems change in distinct ways. Whose perspective do we consider? This decision will significantly impact how we assess systems change in the input system.

---

[11] Ibid.

1.3 Identify the actors who are considered part of the system.

We suggest the following rule: If an actor is removed from the system and its function is significantly affected, or the system doesn't function without that actor, then we should consider that actor as part of the system. Identifying actors' roles in supporting system functioning is also vital to identifying levers and barriers to change.

It is clear that distributors and retailers are critical actors in the input system. Do farmers, regulators, transporters, and banks play a role in the functioning of the input system? Is it possible for the input system to work without a regulator, for example? Could regulators have a more impactful role in improving the system?

1.4 Include the actors across the multiple expected levels of impact.

Consider the actors we will engage directly and those we hope to influence indirectly. These different groups of actors may be referred to as primary, secondary, or even tertiary-level participants. As we understand interconnections and interdependencies between actors, we can find influential partners to work with to change systems.

In the input system case, a program may partner with an input distributor (primary) to improve distribution models. This can help input retailers (secondary) better stock inputs and give farmers easier access to seeds (secondary). Transport companies (tertiary) can also benefit indirectly from spillover effects.

1.5 Discover the factors which limit the system from changing.

Discover the factors considered binding constraints or preventing the system from achieving its purpose or goal. The factors, or *what* in the system needs to change, are often discovered or revealed through analysis and implementation. We need to adjust the system boundaries over time as we discover new factors.

In an input system case, we need to consider a range of factors that can impact the input system's ability to function as intended. These may include more explicit factors like retail strategies and inventory financing. Other factors that are not easily noticeable, such as the level of trust among actors, can be just as important.

1.6 Prioritize the critical factors we will address directly.

Programs are typically unable to address all factors that affect a system. Instead, they tend to focus on influencing a few strategic factors known as root causes or leverage areas. Focusing on these constraining factors, we can focus our efforts and resources on the areas where we can have the greatest impact in changing the system as a whole.

When assessing the agricultural input system, some critical factors may include the long distances farmers need to travel and the quality standards of input products. We may go deeper to understand the underlying causes, such as farmer demand for inputs or lack of competition. We will have to consider which factors to address directly and which we expect to change as a result of changes in other factors.

For some quick reference, here are three diagnostic tests that can be useful to help prioritize factors drawing from structural analysis[12] and growth diagnostics,[13] among other methods. In addition, methods including root cause analysis, cross-impact analysis, and causal loop diagramming can also be useful.

---

[12] Godet, M. (1994). From anticipation to action: A handbook of strategic prospective; Future-oriented studies. UNESCO Publishing. ISBN: 978-92-3-102832-8.

[13] Ricardo Hausmann & Bailey Klinger & Rodrigo Wagner, 2008. "Doing Growth Diagnostics in Practice: A 'Mindbook,'" CID Working Papers 177, Center for International Development at Harvard University.

Figure 4: Example Diagnostic Tests

Test 1: Actors in the system should be behaving in ways to overcome or bypass constraining factors. Understanding why actors are behaving in different ways, we can gain insights into which factors are driving underlying behaviors and how they are affecting system performance.

Test 2: Actors less intensive in a factor should be more likely to survive and thrive, and vice versa. Also called the camels and hippos in the Sahara test in Inclusive Growth Diagnostics, this diagnostic test looks for evidence of parts of the system are thriving to explain why other parts are not.

Test 3: Higher leverage factors should have a high degree of influence on other factors, such that when the factor changes, other factors change. Conversely, higher leverage factors should have a low degree of dependence on other factors, in that the factor doesn't change when other factors change.

1.7 Specify the geographies in which the change is expected to happen.

Geography can be local if we assess change within specific communities of interest. Geography can also be regional, national, or international, depending on how we expect change to scale in the system. For example, some programs will be piloted in specific geographies with the expectation that they will be scaled to broader geographies.

For example, in the input system, the program may work with the wholesaler to pilot new distribution strategies with a group of retailers in a specific department. The theory is that if this pilot succeeds, the distributor can expand these new distribution strategies nationally and internationally. It is only necessary to define geography as part of the assessment so the bounds of assessing that change are commonly understood.

PART 2. DEFINING THE OUTCOMES

KEY POINTS

✓ To say a system has changed, we think it is useful to start with the following: "What parts (or outcomes) have changed? [14] Are these outcomes sufficient to achieve the desired impacts?"

✓ We will tend to discover new outcomes beyond our initial theories as we observe changes in the system, requiring us to modify and update our MEL plans to ensure they are relevant.

✓ Be flexible and adaptable in (re)defining outcome descriptions to explain the story of change by merging qualitative and quantitative evidence that we collect over time.

✓ Later, when we assess *how much* change has happened in the system (Part 3), these definitions of outcomes will be critically important in assessing the degree of change.

THE CHALLENGE

A change in one part of the system is rarely sufficient for the whole system to change. Instead, multiple parts of the system must change to achieve the desired impacts.

Often, we are unsure what outcomes we need to achieve to change the system as a whole. Instead, we must discover them as we implement, learn, and adapt through fast iterations (e.g., piloting, probing, staging interventions, etc.) that allow us to identify what is working and what is not and to adjust along the way. This form of navigation by wayfaring requires flexibility and adaptability in defining outcomes. As we find our way, we will discover new outcomes necessary to achieve, observe unexpected positive or negative outcomes that ripple from interventions, and adjust plans to respond to changing contexts.

Wicked Problems
Some hallmarks of wicked problems are:
- There is significant disagreement about the solution to the problem (or what the problem is in the first place)
- When we try and intervene, there are often unintended consequences, positive and negative, that impact our success
- Even when we progress a little, the context changes, and so what worked before doesn't work the same again

WHAT APPROACHES ARE PRACTITIONERS USING?

The key difference we see is whether to start with the theory of change to identify outcomes or to begin with observation of the system and reflect on outcomes that result.

	Identify Outcomes by Theory	**Identify Outcomes by Observation**
What	Theory-based approaches start with a theory of change, which outlines the logic of how the intervention is expected to lead to desired outcomes and impact. Outcomes are identified by theorizing the causal mechanisms that link program interventions and desired impact.	Observation-based approaches gather data and evidence from the context to identify outcomes that have occurred. This approach triangulates different sources of information (data, perceptions, feedback) to determine the outcomes and the underlying factors contributing to or hindering those outcomes.

---

[14] In the systems idea of an outcome, you control the products and services that your program generates, directly or through partnerships (outputs). Through these outputs, usually through local actors, you influence changes in the system (outcomes). And over time, your outcomes will contribute to significant changes in how the system function to improve the conditions of people's lives, etc. (impact).

	Identify Outcomes by Theory	Identify Outcomes by Observation
Pros	Theory-based approaches are grounded in intervention logic. This approach helps to close gaps in monitoring and evaluation, ties in the program contribution to change, and aligns with standards for results-based measurement. Proponents argue this approach increases the likelihood of achieving systems change.	Observation-based approaches are often more grounded in the context and aligned with the experiences and perspectives of actors in the system. This approach ensures outcomes are relevant and meaningful to the needs and priorities of stakeholders, making them more likely to be the 'right' outcomes.
Cons	Theories tend to require a higher degree of specificity in how change happens. Since theories are often established early, they can be based on an incomplete understanding of the system and incorrect assumptions. In practice, theories can become codified in program governance, creating too narrow and rigid interpretations of change, meaning we may miss or become blind to essential changes in the system.	Observations tend to happen after the fact, meaning we can miss out on baseline references and opportunities to collect data. It can be challenging to orient teams around observation-based approaches, especially as articulating a theory is helpful for teams to define and explore their understanding and assumptions about how systems change.
☑	**Recommendation—Do a Bit of Both!** We recommend finding the right balance between theory and observation. There is no one-size-fits-all recommendation, but in general, we would emphasize the importance of observation at the outset as part of *systems practice* to listen and engage with the system to appreciate how it currently operates.[15] As we discover how the system works and the effects of early interventions, we would want to shift the emphasis toward articulating more nuanced theories of change. These theories can be incredibly useful in aligning teams and partners around a shared vision for collective change. In practice, what this means for MEL managers is that we will need to modify and update our MEL plans more regularly to ensure they remain relevant and effective in capturing impact.

RECOMMENDATIONS

Below are the key recommendations and steps we found helpful in defining outcomes.

1. Start with the theory of change to identify outcomes
2. Identify new outcomes through observation and retrospection
3. Consider different types of outcomes needed for the system to change.
4. Define outcome statements as clearly as possible based on current understanding.
5. Consider how to aggregate outcomes at a higher level.
6. Determine how and if we will quantify the outcomes.
7. Gather qualitative evidence on the outcomes, intended or unintended.
8. Merge the quantitative and qualitative into a single outcome description.

---

[15] USAID. (2017). The 5Rs of Program Monitoring, Evaluation, and Learning: A Technical Note. Retrieved from
https://usaidlearninglab.org/sites/default/files/resource/files/5rs_techncial_note_ver_2_1_final.pdf

2.1 Start with the theory of change to identify outcomes.

The theory-based approach is often called the inwards-out[16] or intervention lens[17] to identify outcomes. By making the program theory of change explicit, teams can identify the outcomes they plan to achieve and track progress toward them over time. This step may involve tools such as result chains or outcome maps that help to visualize the expected outcomes. At the outset of programs, the theory of change tends to be very generalized and based on broad assumptions and limited information. Over time, teams should update these theories with more detail based on discoveries and evidence.

In the input system case, we might partner with a distributor to increase input distribution to retailers (outcome 1). That distributor might focus on the impacts it brings to its sales (impact 1). However, as part of the discussion on shared value, we might highlight the potential it has to increase smallholder adoption of inputs (outcome 2) which would increase farm-level yields (impact 2).

2.2 Identify new outcomes through observation and retrospection.

The second approach is often called the outward-in or helicopter lens[18] approach. This approach involves observing broader changes in the system, whether or not they are directly related to program interventions or observations as part of the assessment of systems changes. Structured methods, such as outcome harvesting and most significant change, can be valuable for collecting and analyzing data from observations. However, the regular discipline by teams of observation and collective reflection is perhaps the gold standard for systems change practice.

In the input system case, the program technical lead notices that farmers are also cultivating a wider variety of crops (new outcome 1), which looks to be affecting the farm household's dietary diversity (new impact 1). However, a gender advisor notes that the input retailers are overwhelmingly men. Women are not allowed to obtain inputs from men due to local gender norms, prohibiting them from purchasing goods without their husbands (new outcome 2). Gender inequality in the region is deteriorating as a result (new impact 2).

Note: Part 4 will discuss the challenges and recommendations for establishing causality. The complexity of development contexts can make it difficult to be empirically rigorous. Theory-based methods can estimate the influence of intervention when clear counterfactuals cannot be determined. While we emphasize observation in this step, it is essential to note the importance of theory later.

2.3 Consider different types of outcomes needed for the system to change.

There is a tendency for systems approaches to be viewed as synonymous with behavior change. Shifts in behaviors and practices are indeed part of systems change, but there are almost always other types of necessary outcomes, including changes in power dynamics, relationships, and social norms, as the text below highlights. Teams should consider the multiple outcomes that will influence the system's performance and achievement of program goals. There are various theories[19] on types of outcomes, but you should define your own, working with stakeholders to identify key drivers of change.

---

[16] BEAM Exchange. (n.d.). Module 1: Systemic Change Monitoring Framework. Retrieved from https://beamexchange.org/guidance/monitoring-overview/systemic-change-monitoring-framework/elements-monitoring-framework/.
[17] Posthumus, H., Kessler, A., & et al. (2021). A Pragmatic Approach to Assessing System Change. DCED. Retrieved from https://beamexchange.org/resources/1334/.
[18] Ibid.
[19] Sources include Kramer, M. R., Kania, J., & Senge, P. (2018). The Water of Systems Change. FSG.. Downing, J., Field, M., Ripley, M., & Sebstad, J. (2018). Market Systems Resilience: A Framework for Measurement. USAID. and USAID. (2016). Technical Note: The 5Rs Framework in the Program Cycle. MarketShare Associates for USAID. (2016). Disrupting System Dynamics: A Framework for Understanding System Changes.

Figure 5: Example Outcome Areas Important in Systems Change
- Changes in behaviors and practices of actors, i.e., who does what differently,
- Changes in relationships between actors, i.e., how actors coordinate and collaborate,
- Changes in how power is distributed and exercised to coerce or reward others,
- Changes in voice and agency to express opinions, make decisions and act on them,
- Changes in the rules, including formal laws or informal norms that govern behaviors,
- Changes in diversity and composition of the actors or elements in a system,
- Changes in access to resources by actors within the system,
- Changes in mental models or subconscious values and beliefs of actors in the system.

In the input system case, increasing the number of distributors in the market (resulting in a shift in the diversity and composition of the actors) may put pressure on the level of competition. As a result, distributors may have less ability to mark up the price of inputs they sell to retailers (a shift in power dynamics). Distributors may be more inclined to develop loyalty programs (a change in behaviors and practices) to develop lasting market relationships with retailers to protect their margins (a shift in relationships between actors).

Comparing Guidance: The Pragmatic Approach prioritizes a broader systems perspective when assessing systems change, often referred to as a helicopter lens, also mentioned in the DCED and M4P Guidance documents. In this guide, we also suggest that taking a broader systems perspective should be standard practice. It's critical to regularly reflect on the conditions for systems change, whether due to shifts in the context or as a result of interventions.

2.4 Define outcome statements as clearly as possible based on current understanding.

Articulate outcomes in well-defined, evaluable outcome statements describing who or what changed the system, when, and where. [20] Well-defined, evaluable outcome statements are important because they provide clear and specific definitions of what changed, allowing MEL teams to identify the evidence needed to track and assess change. One helpful way to create outcome descriptions is to merge the actors and factors.

In the input system case, an outcome statement that addresses counterfeit inputs (as the factor) and distributors (as the actor) could read, "Input distributors adopt mobile verification systems for certified seeds and inputs." Not all outcome statements, however, need a single actor or agent for change. For example, we could reformulate an outcome statement: "New regulations protect the input supply chain from counterfeits."

Note that there is flexibility in defining outcome statements. This flexibility involves balancing precision and clarity with deliberate ambiguity and vagueness. When describing an outcome early in its inception, the characteristics of the change (i.e., the who or what) are often unknown or uncertain. We want to leave space in the statement for systems change to evolve in different ways. [21] Outcome statements should become clearer and more exact as the systems change develops and matures over time.

---

[20] Wilson-Grau, R. (2018). Outcome Harvesting: Principles, Steps and Evaluation Applications.
[21] For a superb discussion on the trouble with outcomes, Smaldino, P., & Nguyen, C. T. (Hosts). (2018, October 31). Paul Smaldino & C. Thi Nguyen on Problems with Value Metrics & Governance at Scale (EPE 06) [Audio podcast episode]. In Complexity Podcast. Simplecast.

2.5 Consider how to aggregate outcomes at a higher level

Systems change can occur at different scales, ranging from the individual level of households and organizations to the meta-level of societal values and beliefs (See Figure 6 below). However, we tend not to consider individual-level outcomes significant unless change happens at a particular scale. Framing outcome definitions in a way that allows for individual-level outcomes to be mapped or aggregated at scale is important for assessing change.[22]

In the input system case, if five input retailers change their stocking strategy with the new distribution model, we wouldn't necessarily want five different (individual-level) outcomes for each retailer. Instead, we would define an outcome statement that is aggregable of these five cases (e.g., a network-level outcome) so that later we assess the significance of that scale.

Figure 6: Further Considerations for scale of Change Adapted from the <u>Socio-Ecological Model</u>

- The individual level: This level includes households, organizations, and individuals. Changes at this level can involve shifts in behavior, practices, and attitudes. Typically, we would not consider this as systems change unless a large number or proportion of individuals are making the change. For this reason, framing individual-level outcomes in ways that can be aggregated is important for assessment.

- The network level: This level includes communities, networks, and supply chains. Changes at this level can involve shifts in relationships, access to resources, and composition of networks. We are more likely to consider this as systems change because of the *influence* of networks on individuals. However, we would similarly consider the extent of participation as a measure of scale of change.

- The institutional level: This level includes government, market, and cultural institutions. Changes at this level can involve shifts in voice and agency, rules, and how power is exercised. We are more likely to consider this as systems change because of the *influence* of institutions on networks of individuals. However, we would consider the extent of that influence as a measure of scale of change.

- The meta level: This level includes the collective values, beliefs, and worldviews of society. Changes at this level can involve shifts in collective identities and mental models. These changes are often the strongest indicators of systems change given their influence at all scales. This level is also the slowest to change requiring sustained effort and collaboration by a multitude of actors over time.

2.6 Determine how and if you will quantify the outcomes.

Develop valid quantitative indicators to measure the most critical outcomes necessary for achieving the systems' goals. These should be responsive to change over time. When quantitative indicators are not possible, use qualitative evidence of systems change (See 2.7 below).
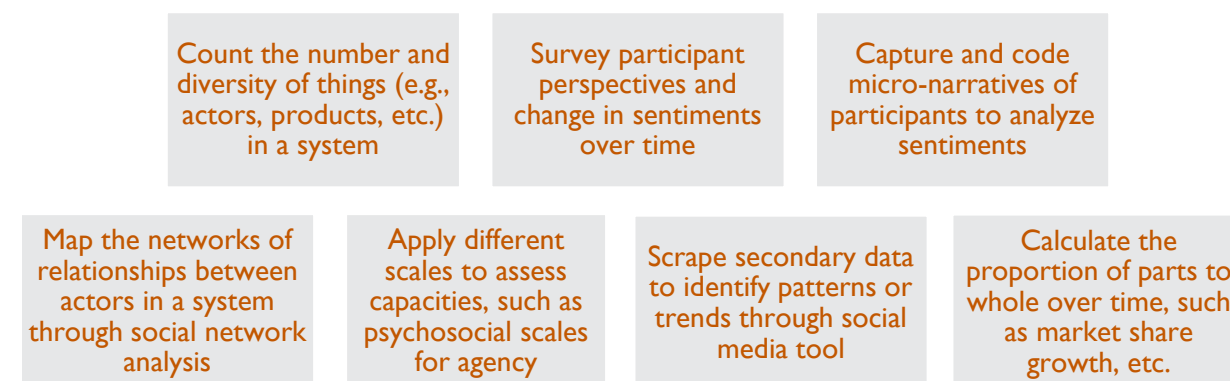
---

[22] The Socio-Ecological Model (SEM) is often attributed to Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Harvard University Press.

In the input system case, we may define the goal of the inputs system to increase smallholder yields by providing more individualized support and availability of high-quality inputs, such as seeds and fertilizer. Smallholder yields in the zones of influence and the change in retailer sales could serve as impact indicators. If the quality of relationships between distributors and retailers also mattered, rather than try and proxy this change with a less reliable indicator, we could focus on gathering qualitative evidence to substantiate the outcome.

We could use many indicators and methodologies to quantify systems change outcomes. [23] Some are more difficult and costly than others. Below are some common ones.

Figure 7: Indicators and Methods to Capture Quantitative Degrees of Systems Change

Count the number and diversity of things (e.g., actors, products, etc.) in a system	Survey participant perspectives and change in sentiments over time	Capture and code micro-narratives of participants to analyze sentiments
Map the networks of relationships between actors in a system through social network analysis	Apply different scales to assess capacities, such as psychosocial scales for agency	Scrape secondary data to identify patterns or trends through social media tool	Calculate the proportion of parts to whole over time, such as market share growth, etc.

Comparing Guidance: *Both* DCED *and* M4P *guidance suggest including relevant indicators for each change in the logic model or result chain. We caution against over-quantifying systems change and the overuse of indicators. Some changes are impossible to quantify, and arbitrary targets can lead to ineffective decision-making. Too many indicators also create compliance burdens and take time away from learning and adaptation processes. Instead, we propose teams shift the focus to evidence and assessing the quality of evidence of changes and causal links of interventions.*

2.7 Gather qualitative evidence on intended and unintended outcomes.

Many changes and aspects of change are not quantifiable and require qualitative evidence. Qualitative evidence should be collected for all intended and unintended outcomes to provide a more comprehensive and holistic understanding of systems change. Teams can choose from qualitative methods such as interviews, focus groups, case studies, expert opinions, open-ended surveys, and direct observation. [24] It is important to follow sound research methods to collect and analyze the qualitative data to help to minimize bias and to improve the validity and reliability of the findings.

Tip: We advise using a narrative format to communicate qualitative evidence. Storytelling through this format can help convey the complexity and depth of qualitative evidence of change in an interesting and understandable way. When telling the story, consider the audience's familiarity with the topic, the most important messages you want to convey about the change, and how you will support those messages

---

[23] Several stocktaking papers include: Colville, J. (Ed). United Nations Development Programme, Bureau for Development Policy. (2013). Innovations in monitoring & evaluating results, Report #41: Testing Tools For Assessing Systemic Change. (2016) USAID, among others.
[24] We are not going to recommend one source. Rather, you can simple search "qualitative research methods" and you will find a wealth of resources on-line from credible sources.

with context, viewpoints, and examples. The storytelling elements in Figure 8 can help to inform this format.

> **Figure 8: Eight Elements for Storytelling Systems Change Evidence**
>
> 1. **Place**: The change happens in a specific context.
> 2. **People**: Multiple actors are involved in the change.
> 3. **Key Events**: The change occurred over some time.
> 4. **Story Driver**: Mechanisms that moved the change forward.
> 5. **Context**: How external context influenced the story.
> 6. **Before/After**: How change evolved (for better and for worse)
> 7. **Purpose/Significance**: The meaning of the change for some populations.
> 8. **Contribution**: The program role in the change is specified.

2.8 Merge the quantitative and qualitative into a single outcome description.

Many results are difficult—even impossible—to capture with quantitative data and methods alone,
and thus, a principle to adhere to when evaluating systems change efforts is
"No numbers without narrative, no narratives without numbers."[25] *— Mark Cabaj*

The objective is to formulate as clear and measurable an outcome description as reasonable, explaining the story of change and merging quantitative and qualitative evidence to substantiate key messages. These statements should be updated periodically as the outcome develops and further evidence is collected by MEL teams.

In the input system, an outcome statement might read as follows:

Farmers in the X department have difficulty accessing inputs due to prohibitively high prices. Farmers are distributed across the zone and must travel long distances to retailers. Because there are few retail locations, high mark-ups are frequent, making farm inputs unaffordable for most farmers. The inability of farmers to access quality, affordable inputs poses a considerable barrier to agricultural productivity. The Y program[26] promoted using agents as retail points for input distributors in rural areas. More than 45 new retailer agents have opened in rural areas, distributing $20 million worth of fertilizers and pesticides, representing a 40% increase in input sales. A survey indicates 8,000 farmers have adopted new inputs through this model, increasing yields by 20%.

Evidence shows that the number and variety of input products retail agents sell have increased. Agents have also started incorporating demonstration plots and farmer field schools to help teach farmers how to apply the inputs correctly. Importantly, interviews with distributors showed they were confident in their relationships with retailers and had begun work with financial institutions to help their agents access credit.

---

[25] Hallie Preskill, & Joelle Cook. (2019). How Do You Evaluate Systems Change? A Place to Start. FSG. Retrieved from https://www.fsg.org/blog/how-do-you-evaluate-systems-change-place-start/.
[26] This case is loosely based on the Feed the Future Inova Activity in Mozambique, but certain changes have been made to illustrate points as a strawman example for this report.

PART 3. ASSESSING DEGREE OF CHANGE

KEY POINTS
- ✓ Assess how much the system-level outcomes (identified in Part 2) have changed.
- ✓ Use objective criteria and set rubrics for degree of expected change. Gather diverse perspectives to make sense and give credibility in the assessment of these systems changes.
- ✓ Focus the assessment on decisions and adaptations it will inform.

THE CHALLENGE

One of the main challenges in assessing systems change is that it's difficult to determine whether the system has truly changed until years later. This points in part to the importance of conducting ex-post assessments. At the same time, waiting for years to receive feedback on whether our efforts are effective at changing the system would make us ineffective at facilitating systems change.

The challenge is detecting early indications of systems change can be quite subjective. Our assessments are prone to inaccuracies, blind spots, and other limitations. The more common approach of using before and after comparisons with baselines may not always be appropriate since complex systems are inherently dynamic and constantly evolving (i.e., correlation is not the same as causation). [27]

The evolution of a system is impacted by its past events, also known as path dependence. Some underlying issues or factors tend to steer the system along a general path. At the same time, there is uncertainty about what issues or factors need (or can) to be changed to shift that course. [28] Navigating this uncertainty requires us to regularly assess the system's progress and evolution and feed this information back into our decision-making process to discover how to influence the system effectively.

WHAT APPROACHES ARE PRACTITIONERS USING?

Programs use multiple methods to assess the significance and degree of change. Below, we have outlined three common methods. Each approach has its own set of variations, exceptions, and combinations used by different programs in different contexts.

APPROACH 1. ASSESS THE CONSISTENCY OF EVIDENCE WITH A LOGIC MODEL

	Consistency of Evidence with a Logic Model
What?	A popular method is to employ a logic model (also known as a result chain[29]) with indicators to assess the degree of change at each stage of the logic model. The logic models articulate explicit theories for how program interventions are expected to bring about systems change through different causal pathways. If the program meets these target benchmarks, we can consider the logic valid and conclude that the system has indeed changed.
Example	Samarth-NMDP used a result chain format to capture "waves" (e.g., second-order effects or levels of impact) that resulted from the program interventions. [30]

[27] Jenal, M. (2019). Measuring Systemic Change in Market Systems Development - A Stock Taking. USAID. Retrieved from https://beamexchange.org/resources/1415/

[28] Kramer, M. R., Kania, J., & Senge, P. (2018). The Water of Systems Change. FSG..

[29] Donor Committee for Enterprise Development (DCED). (2018). Guidelines to the DCED Standard for Results Measurement. Retrieved from https://www.enterprise-development.org/wp-content/uploads/2_Implementation_Guidelines_Defining_Indicators.pdf

[30] Ripley, M., & Nippard, D. (2014). Making Sense of Messiness: Monitoring and Measuring Change in Market Systems: A Practitioner's Perspective. Samarth-NMDP and The Springfield Centre.

Consistency of Evidence with a Logic Model	
Pros	This approach helps ensure interventions are designed and implemented more intentionally and strategically to facilitate systems change. It also helps link the systems change assessment to be practical and relevant to the program activities.
Cons	This approach is vulnerable to mistaken assumptions and blind spots in causal logic. There may also be inherent arbitrariness around indicator selection and target setting, which may not accurately correspond to actual systems change.

APPROACH 2. ASSESS CHARACTERISTICS OF THE SYSTEM AS A WHOLE

Assess Characteristics of the System as a Whole	
What?	An emerging set of methods seeks to define and monitor the overall characteristics of a system over time, related to one or more objective functions (resilience, etc.). If evidence shows that system characteristics have changed and that measures of system functionality have improved, then an assessment can be made that the system has changed.
Examples	IDE developed a *Market Systems Resilience Index* to track the characteristics of the resilience of a market system using a mix of evidence and a participatory scoring system.[31] The USAID/Honduras TMS Activity uses structural analysis to prioritize determinants of systems change and track changes through a set of systems indicators. [32]
Pros	This approach provides a more holistic understanding of the systems change process and its impact beyond program interventions and logic. It can contribute to a relatively more objective assessment of systems change if done methodologically.
Cons	This approach can be complex and resource-intensive since the assessment must consider many factors and issues, making it less manageable in specific contexts. Moreover, system-level changes can be slow to happen and hard to observe in the broader system.

APPROACH 3. ASSESS OUTCOMES BY CRITERIA AND RUBRICS

Assess Outcomes by Criteria and Rubrics	
What?	Another method is to assess outcomes using a set of criteria and rubrics to assess the degrees of scale and sustainability of impact. These outcomes are considered systemic if they have met a certain level of significance based on these parameters of change.
Example	MarketShare Associates applies the Disrupting System Dynamics framework[33] to assess outcomes by depth and strength of change according to specific indicators and rubrics.

---

[31] Ambrosino, C., Oudwater, N., Singh, N., Chiew, Y. L., Oosterom, M., & van Empel, L. (2018). Introducing and operationalizing the Market System Resilience Index (MSRI). Resilience Measurement, Evidence & Learning Conference. Retrieved from https://www.marketlinks.org/sites/default/files/media/file/2020-09/RMEL_Conference-R_MSRI_FINAL_0.pdf

[32] USAID/Honduras Transforming Market Systems (TMS) Activity. (2021). Honduras Market Systems Diagnostic. Retrieved from https://www.acdivoca.org/wp-content/uploads/2021/06/TMS-Market-Systems-Diagnostic.pdf

[33] USAID. (n.d.). Disrupting System Dynamics: A Framework for Understanding System Changes. Retrieved from *https://www.marketlinks.org/resources/disrupting-system-dynamics-framework-understanding-systemic-changes*

Assess Outcomes by Criteria and Rubrics	
Pros	By using multiple parameters, this approach captures the multi-dimensionality of systems change. This method also emphasizes participatory assessments, including with local actors, which can provide a more comprehensive understanding of the change and its impacts.
Cons	This approach relies on inherent assumptions about the relationship between outcome changes and broader systems-level change. Therefore, its reliability often depends on who participates in that assessment and the level of agreement with the conclusions.
☑	**Recommendation** We think that Approach 3, assessing the degree of change of outcomes based on criteria and rubrics, will be the best option for most teams. This method finds an excellent middle ground between rigor and practicality. Approach 1 may not be helpful for teams to try and articulate every change pathway with indicators in all contexts attached. Similarly, Approach 2 may not be within the scope or resources of a program to analyze broader system characteristics. Approach 3 can feed system insights back into programmatic decision-making—the ultimate goal.

In Part 4, we will discuss the utility of logic models (or result chains) in establishing program contribution to systems change.

RECOMMENDATIONS

Below are three recommendations we found helpful to assess the degree of change.

> 1. Define the criteria by which you will assess systems change.
> 2. Establish rubrics to assess the criteria of change (scale, sustainability, and impact value).
> 3. Assess change at the level of the system as a whole.

3.1 Define the criteria by which you will assess systems change.

Assessing systems change can be subjective and hard to define, often falling into the "*I know it when I see it*" category. Our suggestion is to establish a set of criteria that can be used to evaluate the extent that a system has changed. We have found scale and sustainability to be the two most common criteria for assessing systems change.[34] [35] Our third recommended criterion is impact value.

---

[34] Loveridge, D. (2022). Systems Change Frameworks. Medium. Retrieved from https://medium.com/@donna_loveridge/systems-change-frameworks-5b4d8584c7ad

[35] There are strong arguments made to consider depth of change in reference to whether changes occur at the structural and behavioral levels of the system. See USAID. (n.d.). Disrupting System Dynamics: A Framework for Understanding System Changes. Retrieved from *https://www.marketlinks.org/resources/disrupting-system-dynamics-framework-understanding-systemic-changes*

Figure 9: Three Criteria to Assess Degree of Systems Change

SCALE

The number and proportion of target population(s) in the system affected by the change

SUSTAINABILITY

The capacity of the change to endure in the long-term

IMPACT VALUE

The degree to which change brings value to the target group(s)
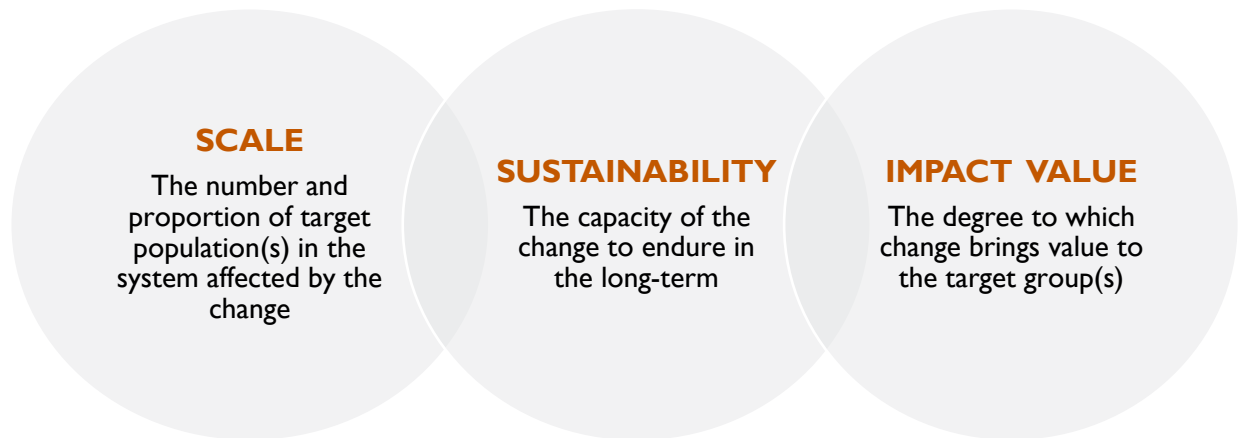
In the input system case, we could measure scale by looking at the percentage of retailers adopting new distribution strategies and the number of farmers accessing inputs and services. For sustainability, we could assess the profitability of distribution and the environmental impact of inputs and technologies. For impact value, we could ask about farmers' points of view on changes and to what degree changes address their basic needs.

Comparing Guidance: *Both* DCED *and* M4P *Guidance also recommend using sustainability and scale of change to measure the impact of market systems programs. We consider adding 'impact value' to be important to explicitly include participants' perspectives in ensuring that impacts are relevant and significant for target populations.*

3.2 Establish rubrics to assess criteria of change (scale, sustainability, and impact value).

Whatever criteria you select, it is important to establish some rubrics for assessing how much or to what degree. Systems change is not a binary—yes or no—response, so we need some way to determine the degrees of systems changes. In the next steps, we offer recommendations of rubrics for scale, sustainability, and impact value.

Note: A rubric enables a team to assess systems change more qualitatively than is possible using only measures, metrics, or indicators. An example of a rubric would be when a supervisor gives an employee a 3 out of 5 on the employee's ability to meet deadlines. The supervisor presumably has not kept a tracker of the submission dates of every single deliverable. The rubric, however, spells out the difference between what qualifies as a 3 vs. 4 on the scale.

In the input system case, there are several rubrics that we could use. A sustainability rubric could assess the distribution model's reported profitability. A score of 1 indicates that retailers are operating at a loss, 2 is breaking even, and 3 is profitable. Another rubric could assess how prevalent distributors and retailers perceive cheating and non-compliance on contracts. A score of 1 indicates widespread reports of cheating, 2 is a mix of cheating and compliance, and 3, aside from a few cases, has little to no cheating reported on contracts.

SCALE:

Scale is defined as the number and proportion of the target population(s) affected by the change. The key components of this definition are presented below:

- *Number:* The desire to reach large numbers reflects a priority in systems approaches toward finding scalable solutions with the potential to replicate or expand beyond direct participants.

- *Proportion:* A significant scale does not always require a large number to benefit from a change. So, we are also interested in the proportion of a population that benefits from a change. If scaling processes are happening, you also don't need to directly reach the entire population.

- *Target population:* The target population should include all individuals or system actors that are part of the target group, whether or not the program directly assists them.

- *Scaling process:* It is also necessary to define the positive feedback mechanism or process that reinforces change to explain how the change scales to impact the broader system.

In theory and in practice, the concept of scale is complex. We offer recommendations for how teams can navigate this question with the caveat that the idea of scale is relative to the systems being assessed. In terms of the degree of scale, nothing is more or less systemic about change that is local to a community or national to an economy or country. Scale is relative to the desired system function.

In the input system case, we estimated 40 retailers in the geographic zone of intervention. Out of this group, 95% have adopted new distribution practices. The major distributors have fully embraced the new distribution model and are continuing to grow and scale it to new retailers and other zones. These retailers together serve a market of 100,000 farmers out of an estimated 600,000 farmers in the zone of intervention.

The table below shows how, using the input system case, we might develop rubrics to look at scale across different population groups and the hypothesized mechanism for scale. In other words, what targets will the program aim for and assess against, as contextualized 'barometers' for scale?

Criteria	Example Rubrics	Discussion
Population: Agro-dealers (in zone of influence)	1. Very low portion of target group (0-20%) 2. Low portion of target group (21-40%) 3. Moderate portion of target group (41-60%) 4. High portion of target group (61-80%) **5. Very high portion of target group (81-100%)**	If there is a population of 40 retailers, we might assess the most significant scale of change to be 80% or more adopt the distribution model. Alternatively, if the hypothesis is that as distributors continue to invest in and drive this change, we may even set a higher target of 60-plus retailers, with the expectation that scaling should increase the number of retailers.
Population: Smallholder Farmers (in zone of influence)	1. Very low portion of target group (0-5%) 2. Low portion of target group (6-15%) 3. Moderate portion of target group (16-25%) **4. High portion of target group (26-50%)** 5. Very high portion of target group (51%+)	We might change the assessment criteria for smallholder farmers to account for the larger population. If in a zone of intervention, 100,000 out of 600,000 farmers, 16% of the total population, purchase inputs could be a significant threshold of change, especially when considering factors such as the current market outreach of retailers and the limited program timeframe.
Mechanism for Expansion (of distribution model by the wholesaler)	1. The wholesaler is maintaining or reducing the number of agro-dealers in the distribution model 2. The wholesaler is increasing the number of agro-dealers in the distribution model only in ZOI **3. The wholesaler is increasing the number of agro-dealers in new geographic zones**	In terms of the scaling mechanism, if the hypothesis is that the wholesaler investment in the new distribution model is the force scaling the change, we may develop some rubric to assess this degree of change by the wholesaler.

Here are five considerations for defining scale assessments.

1. The boundaries of who is (and is not) part of the system will impact how we assess scale in our target populations, reinforcing the importance of having well-defined boundaries (Part 1).

2. It can be hard to determine the total numbers of a target population. Programs may address this by trying to conduct census counts. Often, we rely on best-guess estimates. At the same time, the number of a population may shift over time, which may shift our criteria for assessing scale.

3. Numbers are relative to the desired outcome and function of the system. For example, twenty might seem like an insignificant number. However, if twenty refers to the number of new countries adopting net zero emissions targets, it could be a significant scale of change in climate systems.

4. Proportions are also relative to other criteria for scale. A change that reaches 15% of a population may be *significant* if it is a *tipping point* where the change will scale to the rest of the population.

5. Multiple target population groups may define how we assess scale. For example, if we are assessing systemic change in municipal tax administration, we might consider municipalities as one population group, and we might also consider local enterprises as another population group.

__Comparing Guidance:__ Adopt-Adapt-Expand-Respond (AAER) is a matrix framework often used to guide teams to monitor the progress of systems change along the four stages of A, A, E, and R. One valuable element of AAER is its ability to simplify an often-complex process, and it can be especially helpful for organizing the tracking of scale of uptake of a behavior or innovation. However, we consider using criteria and rubrics for scale, sustainability, and value of the impact of change to be more objective measures. Further, using these three criteria allows teams to consider (and pursue) multiple pathways for systems change, which do not always happen through scaling up and out.

SUSTAINABILITY:
Sustainability refers to the capacity of change to endure in the long-term. Although this definition may seem straightforward, the concept it represents is quite complex and can be interpreted in multiple ways. We suggest that teams decide on how to interpret sustainability considering local perspectives and the context they are working in.

- Sustainable change … we expect that the change to endure into the future.
- Sustainable systems … we expect the system to maintain its ability to function into the future.
- Sustainable impact… we expect the system to meet the needs of the present without compromising the needs of the future.

At each of these levels, we consider different perspectives for assessing sustainability:

How actors respond to the change

Perspective 1. How actors respond to the change.
This perspective assesses the response of actors to sustain the change. The idea is that these responses are signals or early indicators that the changes are likely to be sustainable. We use proxy indicators drawing from resources and frameworks such as Disrupting System Dynamics[36] and lean data[37] guidance to measure them.

---

[36] MarketShare Associates for USAID (2016). Disrupting System Dynamics: A Framework for Understanding System Changes. Retrieved from https://www.marketlinks.org/resources/disrupting-system-dynamics-framework-understanding-systemic-changes
[37] Acumen. (2015). Lean Data Field Guide: Tips for Collecting Customer Data to Build More Impactful Businesses. Retrieved from https://acumen.org/lean-data//

Figure 10: Proxy Indicators to Assess Sustainability of Change

Consideration	Question	Assessment
Satisfaction	*Is the population satisfied with the change?*	Uses satisfaction metrics like Net Promoter Score (NSP) and Customer Satisfaction (CSAT).
Investment	*Is the target population investing in the change on their own?*	Measures whether actors are investing in the change and intend to continue investing in the future.
Innovation	*Is the population innovating or adapting to the change*	Assesses the degree of novelty of innovation and incorporation into business models. [38]
Imitation	*Are others adopting the change without program assistance?*	Collects evidence of others not directly assisted by a program observing and imitating the change. [39]
Continued Use	*Does the target population intend to continue the change?*	Asks the likelihood that the person would continue to adopt change in the future.

In the input case, we might use investment as a proxy indicator for measuring sustainability. To determine the sustainability of the new distribution model, we could inquire about the proportion of investment made by the retailers versus the program investment. If most retailers invest a significant amount of their own resources, it could be an early indicator of sustainability. [40]

Example Rubric
1. Adopters invest none of their own resources
2. Few of the resources are from adopter investment
3. Some of the resources are from adopter investment
4. Most resources are from adopter investment
5. All of the resources are from adopter investment

Capacity of the system to maintain the change

Perspective 2. The capacity of the system to maintain the change.
Another perspective on sustainability assesses whether the system can maintain the desired changes and outcomes over time. The assessment of capacity can be complex. Therefore, we suggest prioritizing critical areas of uncertainty and concern to focus your efforts on assessing change on those key risks.

Figure 11: Proxy Indicators to Assess Capacity of the System to Sustain Change

Consideration	Question	Assessment
Economic	*What are the costs and benefits of the change?*	The cost-benefit or the business case of the return on investment of the change.
Organizational	*Are local organizations able to maintain the change?*	The capacity and interest of organizations to sustain the change within the local context.
Networks	*Are networks mobilized to support the systems change?*	The structure of networks and effectiveness of collective efforts to sustain the change.
Institutions	*Are authorities and politics supportive of the change?*	The political context and stakeholders' interests in sustaining the change over time.

---

[38] MarketShare Associates for USAID. (2016). Practical Tools in Measuring Systems Health. Retrieved from https://www.marketlinks.org/resources/practical-tools-measuring-system-health
[39] Nippard, D., Hitchens, R. & Elliott, D. (2014). Adopt-Adapt-Expand-Respond: A Framework for Managing and Measuring Systemic Change Processes. The Springfield Centre. Retrieved from https://beamexchange.org/resources/130/
[40] *This metric was used in the Feed the Future Mozambique Inova Activity.*

In the input case, we could gather data on retailers' profitability and analyze what percentage have financially sustainable margins. This measure could serve as proof of the long-term viability of these organizations to sustain the change.

Ability of the system to sustainably meet wider needs

Perspective 3. The ability of the system to sustainably meet wider needs.
This third perspective on sustainability highlights the importance of considering the wider context and long-term implications of changes and practices. This includes considering the societal, technological, political, or environmental factors in both the present and future context. By considering how the system is evolving, we can determine whether changes that seem sustainable today will remain sustainable in the long run.

Figure 12: Proxy Indicators to Assess System's Ability to Sustainably Meet Wider Needs

Consideration	Question	Assessment
Resilience	*What is the capacity of the system to manage risks that threaten the change?*	The degree to which the change endures as the system absorbs, adapts, and transforms in the face of ongoing shocks or stressors.
Society	*Does society strongly prefer the benefits the change generates over the status quo?*	Societal interests and norms regarding the change and whether the change fits within expectations of the benefits (and costs) to society.
Environment	*Does change exceed its available resources or harm the environment and people?*	The degree to which the change pushes the system to overshoot the responsible environmental use of Earth's limited resources or harm to people. [41]

In the input case, we might monitor for floods or droughts affecting farm production. Following a flood or drought, we could assess the distributors' and retailers' ability to resume input distribution and whether farmers can re-invest and recover.

IMPACT VALUE:

Gather feedback from diverse populations to understand their perceptions of the impact of change. Multiple viewpoints inclusive of various preferences, values, and beliefs give credibility and legitimacy to the assessment. These views can help understand unintended consequences, trade-offs, or conflicts that might have occurred due to the change.

There are several opportunities to gather perspectives on the value of impact.

---

[41] Raworth, K. (2017). Doughnut economics: Seven ways to think like a 21st-century economist. Random House.

Opportunity 1. Gathering perspectives through regular monitoring surveys.
Annual results and regular monitoring surveys offer an opportunity to collect participant feedback. These questions might provoke participants' reactions on how they perceive the significance or relevance of changes or assistance received.

- *How significant has [the change] been to the well-being of you and your family?*
- *How relevant has [the change] been to solve your real needs and problems?*

In the input system case, we could ask smallholder farmers how relevant the new stocking strategies were to their production. Perhaps it was very relevant because it increased reliable access to diverse seeds and products that were otherwise unavailable when they needed them.

Example Rubric
1. Not relevant at all for smallholder farmers
2. Slightly relevant for smallholder farmers
3. Somewhat relevant for smallholder farmers
4. Very relevant for smallholder farmers
5. Extremely relevant for smallholder farmers

It's helpful to add follow-on, open-ended questions that explore the reasons behind these assessments. In particular, personal stories are practical narrative formats to harvest information on other unexpected impacts.[42]

- *Looking back over the last month, what was the most significant change in [domain]?*

Opportunity 2. Engage diverse participants in the actual assessment of change.
When partner organizations actively participate in the assessment process, it can result in buy-in and support for future changes while also making the assessment process more transparent. We suggest involving teams, partner organizations, and participants in workshops to discuss the significance of these changes and capture the underlying reasons for why (or why not).

> *'Self-evaluation, and the learning it engenders, is necessary for successful program management in complex environments.'[43]*

Opportunity 3. Validate your assessments with diverse key informants.

Share these outcome statements and a description of the significance of change with key informants. These should be individuals who deeply understand the change and can provide valuable insights into the outcomes, their significance, and the intervention's contribution (See Section 4). These informants may respond to the outcome statements in various ways, including through emails, surveys, or telephone calls. Their responses should then be aggregated to validate the outcomes.

- *To what extent do you believe that the change described is significant?*
    [ ] Strongly agree [ ] Agree [ ] Disagree [ ] Strongly Disagree [ ] No opinion
- *What makes the change significant or not?* (as a follow-up)

3.3 Assess change at the level of the system as a whole.

In the sections above, we identified and assessed the degree of change for each outcome. The next step is to determine the degree of change for the system as a whole. Note that parts of the system may have changed, but this does not necessarily mean the system as a whole has changed significantly.

---

[42] Davies, R., & Dart, J. (2005). The 'Most Significant Change' (MSC) Technique. Retrieved from https://www.mande.co.uk/wp-content/uploads/2005/MSCGuide.pdf
[43] Douthwaite, B., Kuby, T., Van de Fliert, E., & Schulz, S. (2003). Impact Pathway Evaluation: An Approach for Achieving and Attributing Impact in Complex Systems. Agricultural Systems, 78, 243-265.

There are **THREE TESTS**[44] we find useful to help assess whole-of-system change:

Test #1	Test # 2	Test # 3
The outcomes collectively have reached a threshold of significance of change.	We are observing significant changes in other parts of the system.	The functionality of the system has improved, generating desired impacts.

Test # 1. The outcomes, collectively, have reached a threshold of significance of change.
In the first test, we find it useful to review all of the outcomes identified as necessary and sufficient for the system to change and assess the degree of change of these outcomes. The critical reflection is whether these changes are significant enough to have impacted the overall system.

	Question	Follow-up (if no)
#1	Are these outcomes in themselves sufficient for systems change?	If not, what other outcomes are necessary for the system to change?
#2	Have these outcomes reached the necessary degree of change?	If not, where is more progress needed for the system as a whole to change?
#3	Do we have enough evidence to assess these outcomes?	If not, what other evidence do we need to answer these questions reasonably?

A confidence scale (like the one below identified in the paper No Royal Road[45]) can help frame our certainty in these assessments—although we will rarely be entirely certain.

Definitely Not the Case	Highly Unlikely	Possible	Probable	Strongly Likely	Certain
○	○	○	○	○	○
There is definitive evidence to rule out this conclusion.	*There is little basis for supporting this conclusion, but it can't be ruled out.*	*There is a viable case for this conclusion, but the balance of evidence is against it.*	*The balance of evidence and credibility points to this conclusion.*	*The conclusion can be made with a high level of confidence.*	*There is definitive evidence to be sure of this conclusion.*

In the input systems case, we could assess the following outcomes with the technical team and partners:
- *Input distributors adopt new distribution model to stock retailers.*
- *Retailers supply diverse input products to farmers.*
- *There is information available on agricultural input products.*
- *Farmers change their purchasing patterns for agricultural inputs.*
- *Farmers adopt new practices and technologies.*
- *Retailers develop stronger relationships with input distributors.*

---

[44] Monitoring at this level of the system is one of three tests used by the USAID/Honduras TMS Activity to define if systemic change has happened and the degree to which TMS has influenced that change.
[45] Coe, J., & Schlangen, R. (2019). No Royal Road: Finding and Following the Natural Pathways in Advocacy Evaluation. Retrieved from https://www.evaluationinnovation.org/publication/no-royal-road-finding-and-following-the-natural-pathways-in-advocacy-evaluation/.

Participants might suggest that these outcomes are insufficient for systems change and that new regulations are necessary to prevent counterfeit inputs. Alternatively, they could argue that the information provided by retailers is of such poor quality that farmers are unlikely to increase yields sustainably. Moreover, the threat of counterfeit inputs and disingenuous promotional strategies could harm the long-term profitability of farmers.

If that is the case, participants might strongly feel that the following outcomes are necessary.
- *New regulations are implemented to prevent counterfeit inputs.*
- *Extension services are made available to smallholder farmers.*

Example of Applied to Illustrative Case

Outcome	Scale	Sustainability	Impact Value
Input distributors adopt new trucking model to stock agro-dealers	●●●●	●●○○	●●●○
Agrodealers supply diverse input products to farmers	●●●●	●●●○	●●●●
There is information available on agricultural input products	●●○○	●●○○	●○○○
Farmers change their purchasing patterns for agricultural inputs	●●●○	●●○○	●●○○
Farmers adopt new practices and technologies	●●●○	●●○○	●●●○
Agro-dealers develop stronger relationships with input distributors	●●●●	●●●●	●●●○

● = This is an illustrative rating of the outcome by the dimension rubric on a 1 to 4 scale, with 1 being that the change had reached an initial degree of change and 4 being that the change had reached a full degree of change.

Test # 2. We are observing significant changes in other parts of the system.

If the outcomes have significantly impacted the system, there should be second and third-order effects or other downstream impacts. These types of changes are rarely part of our regular monitoring and evaluation. Instead, they require generative methods to trace causal evidence of how change has amplified in the system. This can be done through open-ended qualitative interviews to find clues of the effects of effects. Often, we rely on frameworks (See Annex 1) to make sense of these changes.

In the input systems case, we may notice broader shifts beyond the input system. For example, transportation companies may open new routes and services to rural areas tied to input distribution. Increased farm yields might result in greater food security and spending on health and education, etc.

Test #3. The functionality of the system has improved, generating desired impacts.

Ultimately, we want to know whether the system's functionality has improved and how it has impacted our target population groups. In Section 1, we defined system functionality and assigned those key metrics that mattered for us to report on functionality and impact, e.g., input sales and yields.

For example, in an input system, we defined the function of the input system to deliver more localized support and access to quality inputs measured through sales of seeds and fertilizers to smallholder farmers. As a result of the input system functioning better (more competitively, inclusively, or resiliently, for example), we would want to see that the system has significantly impacted smallholder farmer yields and livelihoods (income, etc.). If it has not, perhaps change has not progressed to the necessary degree, other changes are needed, or we may not be intervening in the right systems. These are some of the critical considerations for a program and why openness and flexibility to adapt at multiple levels are often critical to achieving our development impacts.

PART 4. ESTABLISHING CONTRIBUTION

KEY POINTS

- ✓ Establish the causal links between the observed outcomes and the specific program interventions and include evidence to substantiate that theory with a degree of certainty.
- ✓ Make a clear and concise contribution claim and validate this case with key informants. The question we are asking is, *"Did the program help make the change happen; if so, to what extent?"*
- ✓ Exploring these causes behind outcomes and the program's contribution is critical to better understanding how to adapt our approach to bringing about systems change.

THE CHALLENGE

We are not the creators of change; we are the mentors, coaches, and advisors. The people in charge of change are the individuals and organizations within the system. This reality throws a wrench into traditional ideas around attribution. Failing to recognize the people and issues contributing to system change would mean we are functioning in a neo-colonial bubble. [46]

Other confounding aspects of systems change further complicate attribution. For example, systems change can be unpredictable and erratic at times. Furthermore, change is almost always happening in systems, for better or worse, with or without us, as part of more significant societal transitions.

Systems change is emergent. Attributing the direction or speed of change to an intervention distinct from the multiple people and issues that had some role in that change in myriad ways can be difficult.

> **Key Reflection:** Understanding our contribution[47] to systems change facilitates learning by asking: *Are we doing the right things to catalyze systems change?* This can significantly improve our programming.

WHAT APPROACHES ARE PRACTITIONERS USING?

The approaches that we see most practitioners using are counterfactual and theory-based. There has been a lot of debate over which sets of methods work best in measuring complex systems change. [48] At the risk of oversimplification and recognizing that some programs may use these two approaches together in a complementary fashion, we provide some general differences between them.

	Counterfactual	Theory-Based
What?	A counterfactual approach estimates what would have occurred if the intervention had not occurred and then compares it to the actual change observed with the intervention. Typically, experimental designs employ control groups, while quasi-experimental methods use comparison groups. Statistical techniques	Theory-based methods estimate the influence of intervention by identifying evidence that supports or contradicts the relationships in a causal theory. [50] Methods may also seek to eliminate alternative explanations or theories or estimate the influence of other contributing factors. Example methods include process tracing, contribution analysis, and outcome harvesting.

---

[46] Attributed to Luca Crudeli, L. (n.d.). Pushing Structural Changes Over Short-Term Changes in Behavior in Market Systems. Marketlinks. Retrieved from https://www.marketlinks.org/blogs/pushing-structural-changes-over-short-term-changes-behavior-market-systems.
[47] We use the term contribution throughout to include both contribution and attribution.
[48] Jenal, M., & Liesner, M. (2017). Causality and Attribution in Market Systems Development. The BEAM Exchange.
[50] Rogers, P. (2014). UNICEF Office of Research. Methodological Briefs Impact Evaluation No. 6 Overview: Strategies for Causal Attribution.

	Counterfactual	Theory-Based
	are used to evaluate and contrast the performance of the different groups. [49]	
Pros	Counterfactual methods are widely used, highly valued, and prized across multiple fields and organizations. They provide strong causal inferences to identify the effect of the intervention as they rely on comparison groups that did not receive support.	Theory-based methods focus on why change happens, allowing teams to understand better how systems change happens. Theory-based methods are generally more flexible and applicable in complex contexts when establishing a counterfactual is difficult or impossible.
Cons	Counterfactual methods have to simplify parts of programs and contexts to create a controlled environment making them less useful in complex situations. They are less helpful in explaining why the change happened, offering less insight than theory-based methods into how systems change.	Theory-based methods can be more subjective as they depend mainly on the perspectives of those who developed the theory of change. There is often less certainty in causality in theory-based methods due to a lack of comparisons and less precision than in counterfactual methods.
	To utilize counterfactual methods effectively in complex settings, we suggest seeking guidance from an experienced evaluator with a proven track record.	☑ **Recommendation** A theory-based approach best fits most systems change programs and is the one reflected in this guide. Developing a theory or logic model, evaluating change based on this model, and involving stakeholders in this process can provide valuable insights into program effectiveness and contribute to learning and adaptation.

RECOMMENDATIONS

Below are five key recommendations and steps we found to help assess contribution.

1. Establish the context in which the outcome happened.
2. Identify the theorized causal logic for why the outcome did or did not happen.
3. Collect and assess the evidence on the theorized causal logic.
4. Develop a clear, evidence-based claim of the program's contribution.
5. Validate the contribution claim with key informants.

4.1 Establish the context in which the outcome happened.

Describe the context and circumstances of the change. Identify the trends, events, and people that helped or hindered the change. Identifying these other factors provides a grounded understanding of the significance of the program's role and intervention.

On the next page, we present **THREE HELPFUL LENSES** for thinking through the context of change.

---

[49] White, H., & Sabarwal, S. (2014). Quasi-Experimental Design and Methods. UNICEF. Retrieved from
https://www.unicef-irc.org/KM/IE/img/downloads/Quasi-Experimental_Design_and_Methods_ENG.pdf

Has this happened elsewhere?

Comparing and contrasting parallel experiences from elsewhere can be a valuable approach to shed light on the causal contribution of the intervention or program. *In the input system example, you could explore cases of wholesalers who experimented with distribution methods in different regions. In similar situations, you may discover that the distribution models have happened in densely populated areas. That this change occurred in sparsely populated regions is an important context clue to the program contribution.*

What happened before?

Since the elements of systems change are often deeply rooted, it is helpful to look to the past to understand the historical context of the change. Look for any deviations in speed or direction of change that may be attributable to the program. *The input system example may show how wholesalers gradually strengthened their ties with retailers before the program. They may already have a preferred supplier program or provide technical training. By embracing the new distribution model promoted by the program, the wholesalers would have built on their established relationships with retailers.*

What else happened at the time?

Identifying the factors beyond the intervention that drove or hindered change is important to gain insight into how the program contributed to the change. *You could identify a government subsidy for farm inputs in the input system example. The rapid expansion of the new distribution model introduced by the program could partly be due to the push by wholesalers to take advantage of this government program.*

4.2 Identify the theorized causal logic for why the outcome did or did not happen.

Start with the outcome and list its possible causes. Then repeatedly ask "why" or "why not" to trace back to the theorized root or underlying causes. Identify where the program had a specific role in influencing these underlying causes (and where it did not). [51]

We recommend starting with the change that occurred rather than with the intervention itself (or the theory behind it). By focusing on the change, we can better identify the external and contextual factors that influenced the systems change. This approach helps to avoid oversimplifying (and overstating) the causal relationships between program interventions and outcomes.

In the input system example, we may identify the principal causes behind the adoption of the new input distribution model to be (1) farmer demand for inputs, (2) the innovation in the distribution model, and (3) the levels of trust between wholesalers and retailers. For each principal cause for the outcome, we may ask 'why' or 'why not' to dig deeper. For example, in levels of trust, we may identify (1) a history of working together (+), (2) failure to repay input credits (-), or (3) a new digital platform for ordering. And so on. The program's role may have been in developing the new digital platform. The program may also have had other interventions to build relationships by sponsoring forums for wholesalers and retailers.

---

[51] "Tools and Tips for Implementing Contribution Analysis: A Quick Guide for Practitioners." Retrieved from
https://www.evaluationinnovation.org/publication/tools-and-tips-for-implementing-contribution-analysis/

Figure 13: A Team Exercise To Evaluate Outcomes

One way to launch this exercise is to place the outcome on a board and then get diverse stakeholders (not just staff!) involved in the process to evaluate the outcome. In addition to asking why an outcome happened, ask why not. There may be other limiting factors or binding constraints for change that are not being met that are good discussion points for how we need to adapt.

- ✓ Start with the final outcome and work backward to identify the intermediate outcomes that led to it. Identify the outcomes that led to those intermediate outcomes, and so on. Continue this until all the significant and possible causal links between intermediate outcomes and the final outcome are identified.

- ✓ The logical flow between outcomes is commonly referred to as **change pathways** or processes by which the final outcome emerged. It is important to clarify the evidence that substantiates these causal linkages and provides credibility to the theory or provides a basis for further data collection and testing.

- ✓ Once the change pathways are identified, specify where the **program interventions** generated certain outputs that **influenced** these outcomes. It is important to establish the outputs are relevant to the outcomes through some basis of credible evidence and incorporating participant perceptions.

- ✓ **Credible evidence** may from a range of sources, including indicators established as part of the original theory of change, secondary data, and other qualitative evidence as relevant.

Overall, this method provides a structured approach for assessing factors that contributed to systems change and gaining insights into change mechanisms helping to identify opportunities to improve program interventions and improve our impact.

__Comparing Guidance:__ The [DCED](#), [M4P](#), and [Pragmatic Guidance](#) recommend using result chains to map out causal relationships between program activities, outputs, outcomes, and impacts in similar ways to this guidance. We acknowledge that developing such causal logic prospectively (i.e., planning for change) can be difficult, especially when using 'fail fast' and more localized approaches. We encourage teams in complex situations to focus on more retrospection and to take a broader systemic perspective, considering external factors and pathways influencing change beyond what we did.

4.3 Collect and assess the evidence on the theorized causal logic.

Collect evidence on the causal pathways and processes, including alternative theories of change. In certain instances, the evidence supporting causal relationships may be inconclusive or subject to controversy. It's common to realize that we need additional evidence to substantiate a theory. Keep in mind that the evidence available may have limitations, and we may also face time and resource constraints to collect more. There will always be some degree of uncertainty in understanding the contribution of program interventions to systems change, so we should state confidence in assessments.

Below are six recommended categories to state our level of certainty. [52]

Definitely Not the Case	Highly Unlikely	Possible	Probable	Strongly Likely	Certain
○	○	○	○	○	○
There is definitive evidence to rule out this conclusion	*There is little basis for supporting this conclusion, but it can't be ruled out.*	*There is a viable case for this conclusion, but the balance of evidence is against it.*	*The balance of evidence and credibility points to this conclusion.*	*The conclusion can be made with a high level of confidence.*	*There is definitive evidence to be sure of this conclusion*

In the input system example, conducting surveys to gather extensive evidence on the levels of trust between wholesalers and retailers may be impractical. We can seek input from a few wholesalers and program technicians to gauge the impact of trust in the adoption of the distribution model. This can be done by asking distributors about their perceptions and experiences with retailers. Based on the responses, it may be reasonable to conclude that trust probably, maybe even strongly likely, played a significant role in the outcome. However, we cannot say with absolute certainty because there may be some uncertainty about response bias, etc.

4.4 Develop a clear, evidence-based claim of the program's contribution.

Develop a clear, evidence-based claim that communicates the role of the intervention in achieving the outcome while acknowledging the role of other factors. We found the guidance on contribution claims from Clear Horizons Academy's What-If Tool[53] useful.

1. Outcome: The outcome achieved with evidence of the degree of change.
2. Intervention: The intervention that made a significant impact on the outcome.
3. Context: The context of change (e.g., what happened before, elsewhere, and concurrently).
4. Evidence: The evidence to substantiate the outcome and role of the intervention.
5. Conclusion: A short claim stating the contribution of the intervention.

Some have found it useful to create a scale of the extent that the intervention 'made a difference.'[54]

Low Contribution	Medium-Low Contribution	Medium Contribution	Medium-High Contribution	High Contribution
○	○	○	○	○
The outcome would have probably happened in the same way regardless of the intervention.	*The outcome would probably not have happened in the same way without the intervention.*	*The outcome could not have happened in the same way without the intervention.*	*The outcome would probably not have happened at all, and not in the same way without the intervention.*	*The outcome could not have happened at all without the intervention.*

Here is an example of a sample contribution claim related to the input case.

(1) In rural areas, input distributors have used a new agent model to expand retail points for sales. More than 45 retail agents have been established in these zones, and more than $20 million in fertilizers and pesticides have been distributed through this agent network. (2) To implement the new distribution model, wholesalers and retailers used a new digital ordering platform developed through the program intervention. (3) Twenty agents

[52] Coe, J., & Schlangen, R. (2019). No Royal Road: Finding and Following the Natural Pathways in Advocacy Evaluation. Retrieved from https://www.evaluationinnovation.org/publication/no-royal-road-finding-and-following-the-natural-pathways-in-advocacy-evaluation/.
[53] *Dart, J. (n.d.). What Else Test. A Basic Tool for Strengthening Contribution Claims. Clear Horizon Academy.* Retrieved from http://www.clearhorizon.com.au/f.ashx/%24186820%24The-What-Else-Test.pdf
[54] Adapted Thomas Aston. (2020). *Aston, T. (2020). Contribution Rubrics. Medium.* Retrieved from: https://thomasmtaston.medium.com/contribution-rubrics-a76104ecebcb

already had prior distribution relationships with the distributor, providing some foundation of trust. A new government input subsidy program also incentivized wholesalers to increase their distribution. (4) The input distributor agreed that "I was unable to adopt the new distribution model without the new digital platform [developed through the program intervention]. I don't believe I could have managed this agent model without it." (5) To summarize, the evidence suggests that there is a considerable likelihood that the program's digital platform had a significant impact on the expansion of this new agent model by the input distributor.

4.5 Validate the contribution claim with key informants.

Reasonable people may disagree with the conclusion made in the contribution claim. This level of disagreement can happen because of the complexity of systems change and our particular points of view, values, and preferences. We recommend gathering diverse viewpoints to judge the plausibility of contribution claims more reliably.

We recommend that contribution claim statements are shared in some form—whether by email, by phone, or in an interview—with a set of key informants to assess their agreement with claims.

To what extent do key informants agree on the contribution claim.

Strongly Disagree	Disagree	Agree	Strongly Agree	Undecided/ No Comment
◯	◯	◯	◯	◯
Respondents dispute key elements of the contribution claim	*Respondents tend to disagree with parts of the contribution claim.*	*Respondents agree in many parts with the contribution claim.*	*Respondents overall agree with the contribution claim.*	*Respondents cannot comment or are uncertain of the claim.*

ALTERNATIVE APPROACHES

Due to their complexity, we do not discuss counterfactual methods in this guidance, such as randomized controls and quasi-experimental methods. There are obviously valid reasons for using counterfactuals, and if you do so, we suggest seeking the advice of a skilled evaluator who is familiar with applying these techniques in complex contexts.

Nevertheless, many programs will employ less formalized types of action research to investigate causality in a formative assessment. As an example, programs may deliberately mix or combine interventions in different ways to look for differences in effects on different populations. This strategy works best when one or more discrete interventions are piloted with well-defined population groups. Programs can separate participants into various groups based on community, cohort, or other segments, for instance, this can assist programs in determining the relative effects of various support combinations.

Mercy Corps in Uganda has used registration data from technical assistance events to sample different "intensity" groups while using secondary data to look at comparison groups. The researchers then compared outcomes among these groups using statistical analysis. This process helped the team understand different scenarios and adapt intervention designs.

✓ **Comparison Group, "Non-Targeted Low Intensity:"** Communities that are located outside the program implementation area; however, are located in area receiving program resilience-strengthening system-level interventions.

✓ **High-Exposure and Low-Participation, "Targeted Medium-Intensity;"** Communities within the program area that are receiving program resilience-strengthening system-level interventions, however, have limited engagement in direct-level program participation.

✓ **High-Exposure and High-Participation, "Targeted High Intensity;"** Communities within the program area that are receiving both program resilience-strengthening system-level interventions and are highly engaged in direct-level program participation.

CONCLUSIONS

No one knows how to change complex systems. We must discover how to do it.

We begin by understanding the system in a limited way. However, as we intervene and learn, we can progress toward a more complex and complete understanding. This requires an abundance of humility and endless curiosity. Assessing systems change is about connecting the dots and gaining insights into how we can more effectively facilitate systems change. We emphasize the importance of learning from this assessment process as much as the actual evaluation of systems change.

> **Comparing Guidance:** *In line with other guidance, we agree that the primary purpose of systems change assessment is to support decision-making. There is no known pathway to change a given system; it must be discovered through collaborative innovation, learning, and adaptation with local partners and actors. Therefore, the more influential the feedback from systems change assessment, the greater the likelihood we can collectively advance systems change.*
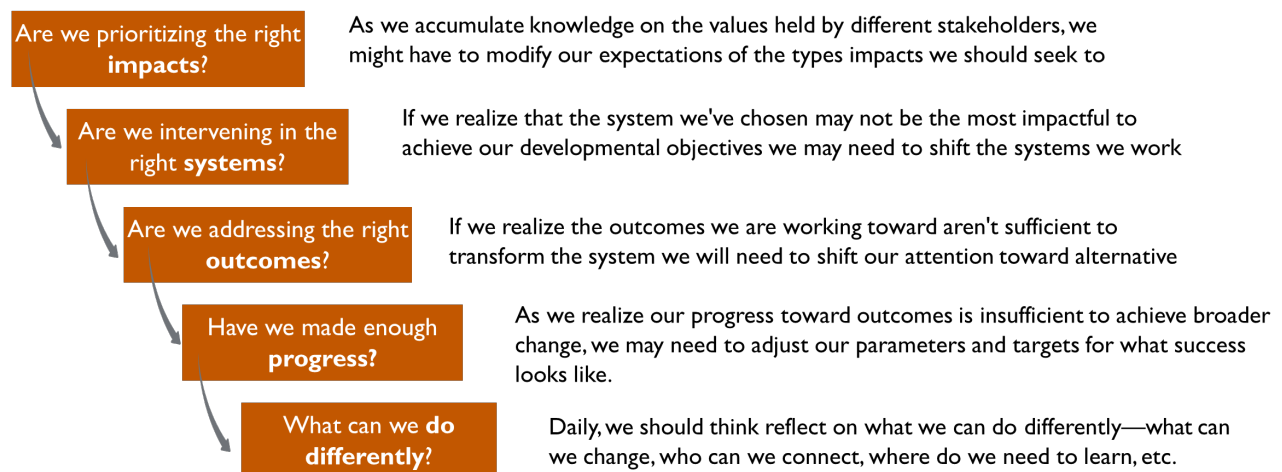
To enhance this learning process, we have provided some helpful reflection questions to include in regular day-to-day Pause and Reflect sessions with your teams.

Figure 14: Team Pause and Reflect Exercise: Discussion Prompts (1/2)

1. What is showing signs of becoming truly systemic?
2. What could we do differently to help this change scale?
3. What faulty assumptions or logic do we need to fix?
4. What areas do we need to explore and find out more?

Furthermore, we must be prepared to make significant changes to our interventions in response to system change assessments. We only get so many chances to adapt substantially in our program lifecycle. Therefore, we need to reflect at all levels of our work at least annually (if not more frequently).

Figure 15: Team Pause and Reflect Exercise: Discussion Prompts (2/2)

Are we prioritizing the right impacts?	As we accumulate knowledge on the values held by different stakeholders, we might have to modify our expectations of the types impacts we should seek to
Are we intervening in the right systems?	If we realize that the system we've chosen may not be the most impactful to achieve our developmental objectives we may need to shift the systems we work
Are we addressing the right outcomes?	If we realize the outcomes we are working toward aren't sufficient to transform the system we will need to shift our attention toward alternative
Have we made enough progress?	As we realize our progress toward outcomes is insufficient to achieve broader change, we may need to adjust our parameters and targets for what success looks like.
What can we do differently?	Daily, we should think reflect on what we can do differently—what can we change, who can we connect, where do we need to learn, etc.

We hope this guide provides practical tools and resources for assessing systems change. The stakes of the issues we are working on are high, and we recognize the need for clear guidance in navigating this complex terrain. There are, however, endless opportunities to explore and discover what works together. We invite you to join our collective effort and contribute your expertise and insights.

We tell ourselves stories about the way that things work in a system. These stories convey our understanding of the world around us.

When through this assessment process, we find evidence that is inconsistent with a story, or we connect the dots between pieces of information, or out of desperation, we drop an assumption that breaks an impasse in our understanding; we are discovering something new.

This is the path to discovery and insight.[55]

---

[55] Adapted from: Seeing What Others Don't: The Remarkable Ways We Gain Insights. (2013). Gary Klein.

ANNEX 1: SYSTEMS CHANGE FRAMEWORKS

The methodology we use to assess systems change involves implicit hypotheses and assumptions about how systems change occurs. Different frameworks articulate the underlying theories that explain how systems change. Below are some key distinctions that we've noticed between frameworks.

|	Points of Difference		Practical Implications
#1	One change or multiple changes	Some frameworks focus on adopting and scaling a single change, such as a new technology, business model, or behavior. Other frameworks evaluate systems change based on a necessary set of outcomes achieved for systems change.	If we consider one change scaling as systemic, we can focus our assessment on the trajectory of that change. If we believe multiple changes are needed, we need to regularly assess those changes in relation to whole-of-system function.
#2	Predictable versus unknowable	There are different ideas about whether the change would occur incrementally and predictably, as opposed to non-linear change, where the course is unpredictable and may be abrupt or periodic.	If change is predictable, we can monitor it based on how we expect it to happen. If change is non-linear, we may need a more generative approach to develop a deeper understanding of the change.
#3	Changes that matter more	Some frameworks prioritize certain types of change, assigning them greater significance. Other frameworks are less specific on the kind of change but on the perspectives of stakeholders.	Based on our theory, it makes sense to prioritize changes that are considered most valuable. Different perspectives can also help us assign significance to change when we're uncertain.

Although our paper was not explicitly focused on systems change frameworks, we all have certain assumptions and mental models that underlie our understanding of how systems change. It is important to articulate those assumptions for others and to be transparent in our assessments. It is good practice to pick a framework or create your own to make explicit those assumptions and mental models.

We have seen five interesting frameworks and how they assess systems change differently.

BEFORE & AFTER WITH INTERMEDIATE PHASES

The Market Development Facility (MDF) assessed the degree of systems change based on four stages: Initial, Intermediate, Advanced, and Mature. The framework makes explicit an expected logical progression of change as it scales from direct partners to the broader system.

MDF Systemic Change Pathway
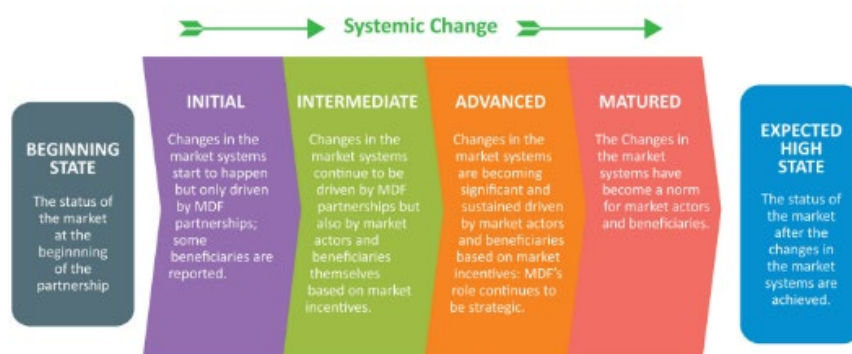
RIPPLES ACROSS MULTIPLE LEVELS OF THE SYSTEM

The Socio-Ecological Model assesses change across multiple levels of a system. The framework explicitly values change that occurs beyond individual behavior to wider communities and institutions. [56]

Based on Socio-Ecological Model conceived in 1988 by Dr. Kenneth McLeroy
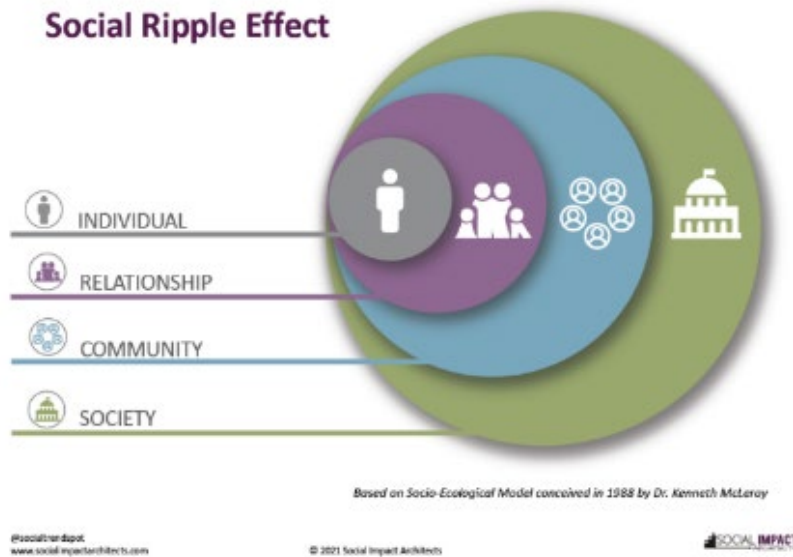
OUTCOMES NECESSARY FOR THE SYSTEM TO CHANGE

The Disrupting Systems Dynamics Framework establishes dimensions to systems change, including depth and strength, to value individual changes as more (or less) systemic. The framework also assesses these changes in relation to the overall impact on target populations. [57]
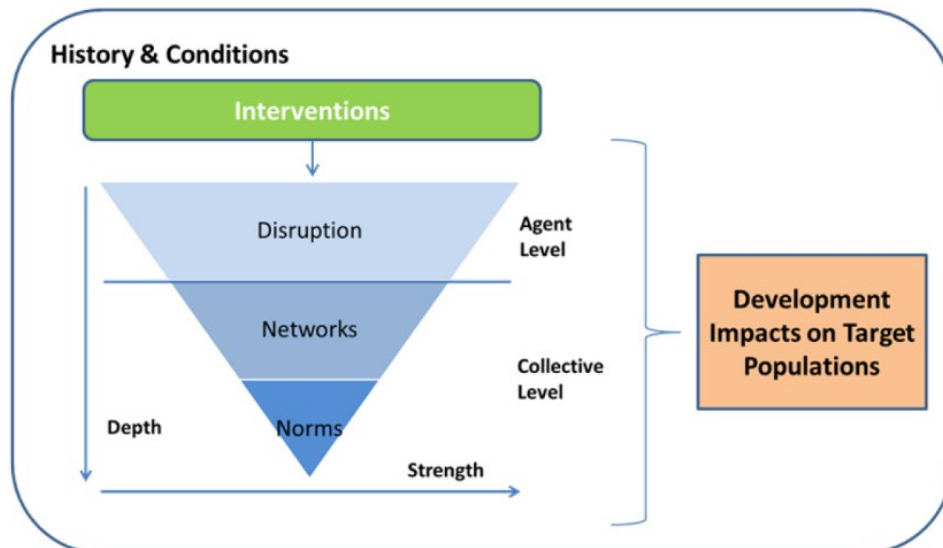
[56] Smith, S. (2021). The Socio-Ecological Model Makes Large-scale Social Change Possible. Social Impact Architects. Retrieved from https://socialimpactarchitects.com/social-ripple-effect/.

[57] MarketShare Associates for USAID. (2016). Disrupting System Dynamics: A Framework for Understanding System Changes. Retrieved from https://www.marketlinks.org/resources/disrupting-system-dynamics-framework-understanding-systemic-changes

SHIFTS IN THE CONDITIONS THAT HOLD PROBLEMS IN PLACE

FSG's Water of Systems Change Framework[58] identifies multiple types of changes (e.g., resource flows, policy change, etc.) that are conditions for systems change. Donna Loveridge created a maturity model[59] to describe the expected characteristics of the significance of those types of systems changes.

	1) Beginning	2)	3) Strengthening	4)	5) Significant
Policies	There is no or little awareness of the need for or benefits of a rule change. Or there are no active efforts to achieve a rule change.	There is some awareness of the need for or benefits of a rule change and some early efforts by a limited number of actors to achieve a rule change.	Desire for an inclusive rule change has begun to increase and various proposals are being discussed.	Notable progress is made towards formalising pro-inclusive growth rule changes e.g. draft policies, legislation or regulations.	A significant pro-inclusive growth rule change has been made and interest in the issue is beginning to grow and strengthen
Practices	Negligible presence of inclusive business models or practices in the market.	There are a limited number of inclusive business models in the market but they are not wide spread.	Several inclusive business models or practices in the market.	Significant presence of inclusive business models or practices in the market.	Dominant, widely adopted inclusive business models or practices in the market
Resource flows	Negligible or single examples of investment plans and investments of market actors (government and private).	A limited number of investment plans and investments of market actors (government and private) are being made.	Several market actors are making moderate to large investments, relative to their size and role in the market and others are planning to invest.	Many market actors are making moderate to large investments relative to their size and role in the market, but the investment may not be spread across all sizes/roles of market actors.	Wide-spread investment (financial and non-financial markets) by a range of market actors that is significant according to their role in the market.

NON-LINEAR TRANSITIONS FROM THE STATUS QUO

The Three Horizons Framework[60] articulates systems change as a transition from the current status quo to a future successor system with periods of uncertainty in between. The Finance Innovation Lab and Smart CSOs Labs[61] build from Transitions Theory to describe these transitions at multiple levels.
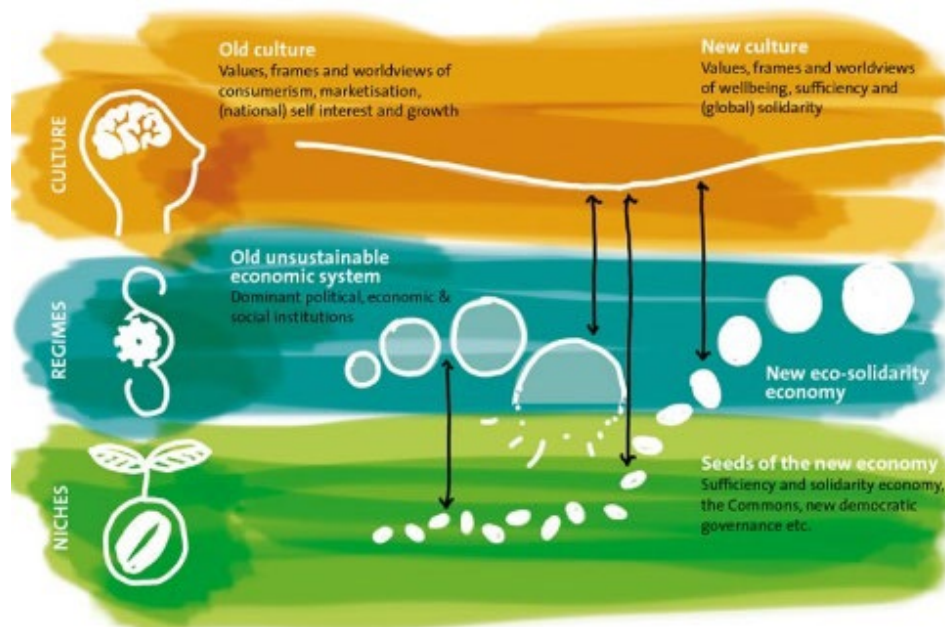
[58] Kramer, M. R., Kania, J., & Senge, P. (2018). The Water of Systems Change. FSG.
[59] Loveridge, D. (2022). Market Systems Change Rubric. Oxford Policy Management. Retrieved from https://beamexchange.org/resources/1658/.
[60] Hodgson, A., & Midgley, G. (2015). Bringing Foresight into Systems Thinking: A Three Horizon Approach. Political Science, 67(1), 1-24.
[61] The Finance Innovation Lab. (n.d.). A Strategy for Systems Change. Retrieved from https://financeinnovationlab.org/wp-content/uploads/2015/04/FIL_SystemsChange-Web-Final.pdf.