



**USAID**  
FROM THE AMERICAN PEOPLE

# EVALUATING SYSTEMS AND SYSTEMIC CHANGE FOR INCLUSIVE MARKET DEVELOPMENT

LITERATURE REVIEW AND SYNTHESIS

**LEO**

REPORT NO. 3

Leveraging Economic  
Opportunities

**JUNE 2014**

This publication was produced for review by the United States Agency for International Development. It was prepared by Ben Fowler of MarketShare Associates and Elizabeth Dunn of Impact LLC for ACDI/VOCA with funding from USAID/MPEP's Leveraging Economic Opportunity (LEO) project.

# EVALUATING SYSTEMS AND SYSTEMIC CHANGE FOR INCLUSIVE MARKET DEVELOPMENT

LITERATURE REVIEW AND SYNTHESIS



## DISCLAIMER

The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

# CONTENTS

<b>I. INTRODUCTION</b>	<b>1</b>
<b>II. SYSTEMS AND SYSTEMIC CHANGE</b>	<b>2</b>
<b>III. EVALUATION TYPOLOGIES AND CHALLENGES</b>	<b>5</b>
<b>IV. TOWARD A FRAMEWORK FOR EVALUATING MARKET SYSTEM FACILITATION</b>	<b>8</b>
<b>V. INDICATORS OF SYSTEMIC CHANGE</b>	<b>13</b>
<b>VI. SUMMARY AND CONCLUSION</b>	<b>19</b>
<b>REFERENCE LIST</b>	<b>21</b>
<b>ANNEX: ANNOTATED BIBLIOGRAPHY</b>	<b>24</b>

# ACRONYM LIST

AECF	Africa Enterprise Challenge Fund
AMAP	Accelerated Microenterprise Advancement Programme
DCED	Donor Committee for Enterprise Development
DFID	Department for International Development
GTZ	German Agency for International Development
IE	Impact evaluation
LEO	Leveraging Economic Opportunities
M&E	Monitoring and evaluation
MaFI	Market Facilitation Initiative
MAP	Market Assistance Programme
M4P	Making Markets Work for the Poor
MSC	Most significant change
NGO	Nongovernmental organization
OM	Outcome mapping
PDIA	Problem-Driven Iterative Adaptation
SDC	Swiss Agency for Development and Agency
USAID	United States Agency for International Development

# I. INTRODUCTION

There is increased interest in systems and systems thinking within the development community, based in part on an emerging recognition that scale, impact, and sustainability can all be linked to systemic change. Inclusive market system development, in particular, seeks to modify both the structure and dynamics of market systems in ways that contribute to inclusive growth. Donors and practitioners are working to improve their understanding and application of systems concepts within inclusive market system development while also seeking better ways to detect, measure and evaluate systemic changes. USAID's recently published framework for working with local systems cites the need for appropriate monitoring and evaluation methods while encouraging the development of new and better ways to measure systems change (USAID 2014). At the same time, there is growing application of systems thinking among evaluation experts (Reynolds et al. 2012; Stern et al. 2012; Patton 2011; Williams and Hummelbrunner 2011; Hargreaves 2010).

This report summarizes key findings from a review of selected literature on evaluating systems and systems change. The purpose of the review was to inform the evaluation research agenda under USAID's Leveraging Economic Opportunities (LEO) project. In particular, this review highlights findings that can contribute to the development of, first, an evaluation framework for interventions designed to facilitate inclusive market systems development and, second, empirical approaches for identifying and monitoring systemic changes.

While focused primarily on evaluation and systems concepts, the review also included literature related to complexity analysis, resilience, and specific monitoring approaches. In addition to the reference list at the end of the document, an annotated bibliography is included as an annex, providing short summaries for most of the documents included in the review.

## **BOX I: LEVERAGING ECONOMIC OPPORTUNITIES**

Leveraging Economic Opportunities (LEO) is a three-year contract to support programming that fosters *inclusive growth* through markets. Building on USAID's value chain approach, LEO focuses on:

- (1) *a systems approach* to markets, acknowledging the complex interrelationships among market actors, market and household systems, climate change, nutrition, the policy environment, and sociocultural factors, including poverty and gender; and
- (2) *inclusion*, recognizing the role that a spectrum of actors—from resource-poor households and small-scale enterprises to larger and more formal firms—play in catalyzing market change and growth that benefits the poor.

## II. SYSTEMS AND SYSTEMIC CHANGE

Within the context of evaluation, systems thinking is more of a conceptual paradigm than it is any specific type of tool or method (Reynolds et al. 2012). While evaluators do not share universally accepted definitions of systems and systems change, there is general agreement that systems can be described in terms of three concepts (Williams and Hummelbrunner 2011):

- **Relationships:** Relationships (or interrelationships) are the oldest and best known of the systems concepts, referring to interconnected processes that define linkages between actors and influence individual behavior and system-level results.
- **Perspectives:** Perspectives shape actors' understandings of the system and its parts, along with their beliefs about system performance, ways to change the system and incentives for promoting change.
- **Boundaries:** Boundaries define the limits of the system being studied, which helps to keep the system manageable for analytical purposes but may result in excluding relevant components.

While the definitions of systems used by USAID, DFID and SDC are similar in several ways, they differ in terms of where system boundaries are drawn. USAID defines a local system<sup>1</sup> in terms of its result (or outcome): “those interconnected sets of actors—governments, civil society, the private sector, universities, individual citizens and others—that jointly produce a particular development outcome” (USAID 2014, p. 4). As an approach for understanding systems, the “five Rs” relate to structural and governance features which, in the case of market systems, include the flows of products, payments and information between market actors (box 2).

DFID and SDC define a system more broadly as “the multi-player, multi-function arrangement comprising three main sets of functions (core, rules and supporting) undertaken by different players (private sector, government, representative organizations, civil society, etc.) through which exchange takes place, develops, adapts and grows”

### BOX 2: THE FIVE R'S OF LOCAL SYSTEMS

- **Resources:** Local systems transform resources—such as budgets or raw materials—into outputs.
- **Roles:** Most local systems involve a number of actors taking on defined roles, such as producer, consumer, funder or advocate.
- **Relationships:** Interactions between actors in a system establish various types of relationships, such as commercial, administrative or hierarchical.
- **Rules:** Rules govern a system by defining or assigning roles, determining the nature of relationships, and establishing terms of access to resources.
- **Results:** These include measures of system strength as well as traditional outputs and outcomes.

*Source:* USAID (2014, p. 8)

---

<sup>1</sup> The “local” in a local system refers to actors in a partner country. As these actors jointly produce an outcome, they are “local” to it. And as development outcomes may occur at many levels, local systems can be national, provincial or community-wide in scope.

(DFID and SDC 2008, n.p.). Given the diversity of environments in which market system programs operate, implementers and evaluators also seek flexibility to define a system in ways that are locally appropriate.

Given the lack of consensus on how to define a system, it is not surprising that there is no agreement about how to define systemic change. Indeed, many publications refer to systemic change without defining it at all (Marks and Wong 2010). Table 1 lists three definitions of systemic change. The first two definitions refer to change within systems in general, while the third refers specifically to change within market systems.

**Table 1: Definitions of Systemic Change**

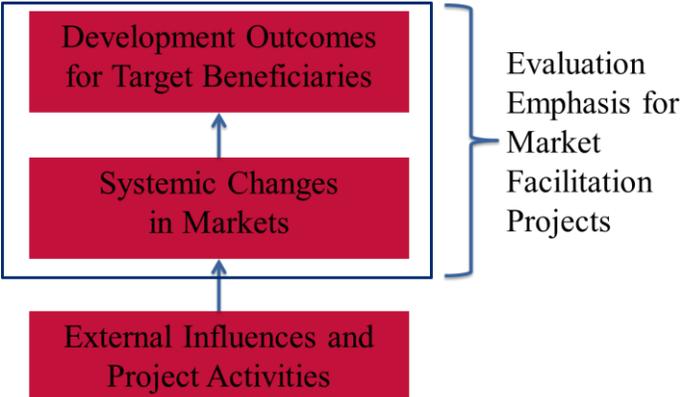
Source	Definition of Systemic Change
Parsons and Hargreaves (2009, n.p.)	“[S]hifts in patterns (similarities and differences) of system relationships, boundaries, focus, timing, events and behaviors over time and space.”
The SEEP Network (Osorio-Cortes and Jenal 2013, p. 7)	“Transformations in the structure or dynamics of a system that leads to impacts on large numbers of people, either in their material conditions or in their behavior.”
DFID and SDC (2008, n.p.)	“Change in the underlying causes of market system performance – typically in the rules and supporting functions – that can bring about more effective, sustainable and inclusive functioning of the market system”

Other descriptions of systemic change focus on the importance of aligning stakeholders’ perspectives. According to this view, systemic change occurs “when both the objective and approach of the social entrepreneur/innovation are adopted or supported by key stakeholders as a priority social issue and best in class solution” (Marks and Wong 2010, p. 5). Version VI of the Donor Committee for Enterprise Development’s Standard for Results Measurement, also known as the DCED Standard, views systemic change as being caused by “introducing alternative innovative sustainable business models at support market level (such as in private sector, government, civil society, public policy level)” and it further outlines several results of systemic change: “widespread indirect impact by crowding in at support market levels and copying at final beneficiary level” (DCED 2013, p. 18).

The definitions and descriptions of systemic change share an emphasis on shifts in the underlying structural elements and patterns that characterize the system. While these structural changes may be instrumental to achieving the desired development outcomes for the target beneficiaries, it is important to remember that systemic change is not the final objective. Instead, systemic change is treated as an intermediate outcome that derives its significance from the role that it plays in achieving inclusive economic growth at scale.

Evaluations that are sensitive to systemic change, therefore, should attend both to identifying the systemic changes that have occurred as well as to analyzing the contribution of those changes to achieving final outcomes for the target population. This relationship between a) the evaluation of systemic changes in markets and b) the evaluation of final development outcomes for target beneficiaries is illustrated in figure 1 below. In addition to clarifying that systemic change is an intermediate outcome, the literature also reminds that systems can be inherently dynamic and that systemic change often emerges independently of any donor intervention. Complexity in market systems prevents knowing with certainty how the system will respond to an intervention, and emergent properties of a system can be both positive and negative from the perspective of their contributions to achieving development outcomes.

**Figure 1: Dual Emphasis in Evaluating Market Systems Facilitation**



# III. EVALUATION TYPOLOGIES AND CHALLENGES

USAID Evaluation Policy defines evaluation as “systematic collection and analysis of information about the characteristics and outcomes of programs and projects as a basis for judgments, to improve effectiveness, and/or inform decisions about current and future programming” (USAID 2011, p. 2). This definition lists evaluation purposes that are consistent with the three evaluation types described below, which are grouped according to the purpose of the evaluation: summative, formative or developmental.

## A. EVALUATION TYPES BASED ON PURPOSE

Three general types of evaluation can be distinguished based on characteristics related to the purpose of the evaluation and conditions (stage or status) of the intervention:

- **Summative evaluations** seek to make an overall judgment about the merit and worth of a program (Patton 2011). This would include impact evaluations and the types of performance evaluations that focus on program achievements. As described in box 3, impact evaluations are designed to determine the extent to which observed changes in outcomes for the target population can be attributed to the intervention. While performance evaluations also might include measurements on outcomes for the target population, the observed outcomes are not compared to a counterfactual.
- **Formative evaluations** assess the performance of program models and the fidelity or adaptation involved in implementation. Also known as implementation, performance or process evaluations, formative evaluations are typically applied when the context is well understood and it is assumed that a model can be developed with fairly standard inputs leading to predictable outcomes. Findings from formative evaluations are used to improve interventions while standardizing the approach.
- **Developmental evaluations** support continual improvement of an intervention approach that has not been or cannot be standardized, when experimentation is still being done to identify the proper approaches, and when there is continual emergence of new “questions, challenges, opportunities, successes and activities” (Preskill and Beer 2012). Designed to be flexible, developmental evaluation can be applied in complex settings when outcomes are unknown.

### BOX 3: USAID EVALUATION TERMS

**Impact evaluations** measure the change in a development outcome that is attributable to a defined intervention; impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change.

**Performance evaluations** focus on descriptive and normative questions: what a particular project or program has achieved (either at an intermediate point in execution, or at the conclusion of an implementation period); how it is being implemented; how it is perceived and valued; whether expected results are occurring; and other questions that are pertinent to program design, management and operational decision-making.

**Performance monitoring** follows changes in indicators (for outputs and/or outcomes) to reveal whether desired results are occurring and implementation is on track.

*Source:* USAID Evaluation Policy (2011)

## B. CHALLENGES IN EVALUATING MARKET SYSTEM FACILITATION

As donor interventions in private sector development have shifted toward the use of facilitation approaches, a number of evaluation challenges have emerged. For the purpose of this discussion, market system facilitation is defined as intervening to

“...stimulate change in market systems without the project taking a direct role in . . . the system. Practitioners and donors using this approach try to minimize direct provision of goods and services by the project—focusing instead on changing relationships between actors in the value chain or introducing new ways of doing business that increase the local availability of needed goods and services” (USAID 2012, p. 1).

An initiative under USAID’s AMAP Project (Creevey et al. 2010) identified several evaluation challenges associated with market system facilitation. Some of these challenges are listed below and, while they may not be unique to market system facilitation, they arise from common strategies used in implementing the facilitation approach, also known as the value chain or M4P approach. The incorporation of systems thinking—with the conceptualization of the donor intervention as an endeavor to facilitate systemic change—helps to sharpen the dimensions for some of these evaluation challenges:

- **Emphasis on indirect beneficiaries and targeting of secondary contacts can limit the options for establishing the counterfactual.** Facilitation approaches seek to create strong demonstration effects that elicit imitation among non-supported firms, and create benefits that extend to firms and individuals outside of any initially defined treatment groups. As is outlined in section VI below, this imitation is considered by many practitioners as an important indication of systemic change. Over time, the evaluator’s goal of maintaining a valid control group can conflict with the implementer’s goal of attracting as many imitators as possible. In fact, many market systems interventions include components to change macro- and meso-level variables (e.g., policy, infrastructure) that affect all market actors at the microeconomic level; in these cases it is difficult to construct a valid control group since everyone is affected. Moreover, implementers using a market facilitation strategy are not able to play a direct role in selecting treatment group beneficiaries. Instead, buyers, suppliers and other firms select their own commercial partners, based on considerations related to profits, transaction costs, trust and other factors. This affects not only the identity of the secondary contacts, but also their locations, which makes it impossible to predict with certainty which individuals or groups will or will not “participate” as beneficiaries in the market system intervention. It also means that participants can be expected to differ in significant (if unknown) ways from non-participants, since they are selected based on the perspectives and criteria of these independent actors (other firms) in the market system.
- **Multilevel, sequenced interventions can create a treatment effect that varies over time and space.** Changing the dynamics of a system often requires intervening to address several issues (e.g., policies, firm-level behavior) simultaneously. The resulting scope of some market system facilitation programs, implementing multiple interventions at different levels (micro, meso and macro) and across different geographic areas, impedes evaluation: “in practice it is difficult to conduct comprehensive impact evaluations of a project operating in multiple locations and multiple value chains where timing, conditions, and types of interventions are different” (Creevey et al. 2010, p. 2). In programs with several interventions, individuals may participate in some opportunities but not others. Where interventions complement and build on each other over time, there can be a graduated degree of participation (“degree of treatment”) that varies

over the target population and may be conditional on completion of other project components. For example, the benefits that farmers receive from better accessibility of quality input supplies would be conditional on their knowledge of the benefits of application and good practices for doing so.

- **Adaptive implementation approaches may negate evaluation methodologies and findings.** The nature of systems means that the “correct” way to facilitate systemic change is often unknowable *ex ante*, and evolves with time. Effective facilitation therefore implies an adaptive approach that can respond to dynamic and unpredictable events. Dynamics in the market system might be a response to the intervention, due to inherent features of the system that are independent of the donor-funded intervention, or caused by external shocks or other non-program influences. In adjusting their strategies for facilitating market systems change, project implementers might change the geographic location of their activities, the identity of their private sector collaborators, or the sectoral focus of their interventions. These kinds of shifts reduce the value of baseline data and create continuity problems for evaluation. For example, the evaluation plan for USAID’s PROFIT project in Zambia had to be revised after implementers relocated their cotton sector activities in response to changes in international competition (Creevey et al. 2010). As Ruffer and Wach note, “the adaptive nature of M4P programs means that they cannot rely too heavily on data sets (e.g., baseline and control groups) identified *ex ante*”. This creates the need for flexible evaluation designs that are sensitive to emergent conditions.
- **The unpredictable pace of systemic change challenges evaluation timing.** The systems literature indicates that systemic change is nonlinear, path-dependent and episodic. A system may remain latent for long periods before experiencing large-scale change once a “tipping point” is reached. This point may occur following the end of a project and its final evaluation. Evaluation findings may therefore underestimate a program’s true impacts if conducted prior to the tipping point (Johnson and Boulton 2014). Where facilitation initially causes a negative response that is subsequently reversed, an early evaluation would indicate the initiative is harmful and might suggest cancellation.

These and other challenges contribute to the difficulty of evaluating change under market system facilitation. Empirical evidence suggests that few evaluations to date have adequately assessed systemic change. In a recent review of 14 evaluations of market systems programs that were funded by donors including USAID, DFID, SDC and GIZ, the authors note that just five did so in a way that triangulated multiple sources of quality evidence (Ruffer and Wach 2013). The findings from the review highlight the need to continue working toward the development of evaluation approaches, frameworks and indicators that can help to address some of the challenges associated with evaluating market system facilitation.

# IV. TOWARD A FRAMEWORK FOR EVALUATING MARKET SYSTEM FACILITATION

This section considers topics related to developing an evaluation framework for market system interventions and systemic change under facilitation. It considers the nature of evidence, general evaluation principles, and evaluation-supportive monitoring approaches. While there are different perspectives on the best approach for evaluating market system facilitation, the findings suggests that one path forward is to build on foundational evaluation guidance, but modify the framework to address some of the evaluation challenges specific to facilitation programs and incorporate newer concepts and tools for evaluating systems and systemic change.

## A. RECONSIDERING EVIDENCE AND IMPACT

The overarching purpose of an impact evaluation is to measure changes in key outcome variables and determine how much of the observed changes (if any) can be attributed to the intervention. The previous section described a number of factors that can limit the validity of the counterfactual, including iterative interventions at multiple levels, a shifting range of actors and locations, and active strategies for increasing the number of indirect beneficiaries. Stern et al. (2012) argue that it is difficult in many of these contexts to make a credible argument for attributing observed changes to a specific intervention. Given the role that factors outside of a development intervention have on final results, the authors argue that evaluation questions should focus on understanding the contribution of particular interventions in conjunction with other causal factors, a process assisted by focusing on the following four questions.

- **To what extent can a specific (net) impact be attributed to the intervention?** Developmental impacts are typically caused by multiple factors, which may or may not be necessary or sufficient for those impacts to occur. While experimental designs, such as randomized control trials, can enlighten us on the effectiveness of an intervention, they require many conditions to be present in order to be useful, and Stern et al. (2012) estimate that these exist in about five percent of projects. When these conditions do not hold, or cost considerations prohibit conducting multiple RCTs across different environments, then case-based designs and comparative analysis may be applied.
- **Did the intervention make a difference?** This question considers the necessity and the sufficiency of an intervention to result in the observed impact. An intervention may be both, one, or neither of these. Where one of a number of contributory factors, it is important to understand the role of the intervention. Statistical approaches are more challenging to apply in such cases, where strong project monitoring systems may be critical to ensuring course correction.
- **How has the intervention made a difference?** The amount of existing knowledge about an intervention will influence how to best answer this question. Theory-based evaluation is often appropriate, particularly where there is little existing knowledge.
- **Will the intervention work elsewhere?** This last question emphasizes the importance of analyzing and strengthening the external validity of the evaluation findings.

## B. EVALUATION PRINCIPLES

A number of important principles have been codified into USAID's Evaluation Policy (see box 4). DFID's International Development Evaluation Policy (2013) is broadly consistent with USAID evaluation principles, but DFID's policy additionally focuses on ethics, requiring that evaluations do not harm participants, respect their privacy, and only involve willing subjects.

While many of the principles apply equally to a systems context, there are some controversial points. One example relates to the principle of maintaining independence between the evaluator and evaluation process, on the one hand, and the program being evaluated on the other hand. Ruffer and Wach (2013) argue that evaluators of market system facilitation programs need an in-depth understanding of the program in order to evaluate it. Close collaboration between the evaluation and implementation teams can support this understanding, yet may compromise the principle of independence in evaluation. Similarly, implementers may play a role in collecting evaluation data that is audited by external evaluators. Evaluators are in a better position to develop a thorough understanding of a program if they are contracted at the beginning of the program and engage with staff periodically during the course of implementation to maintain the relevance of the evaluation design.

The Degrees of Evidence framework, developed under USAID's AMAP project (Creevey et al. 2010), considers more specifically the inherent evaluation challenges associated with market systems facilitation (see discussion in section III.B above). In addressing these challenges, it outlines five principles for ensuring evaluation quality. Some of these are similar to the principles in the USAID and DFID guidance. The Degrees of Evidence Principles offer the following evaluation guidelines:

- The evaluation should be grounded in a plausible causal model.
- Evaluation methods should be assessed relative to four standards of methodological validity: internal validity, external validity, construct validity, and statistical conclusion validity.
- Evaluation findings should be triangulated to determine the preponderance of evidence.
- Evaluation methods should follow sound data collection practices.
- The evaluation methods used, along with their strengths and weaknesses, should be transparently presented to the end user(s).

### BOX 4. USAID EVALUATION POLICY PRINCIPLES

- **Integrate evaluations into up-front program design.** Establish a clear theory of change and evaluation questions at the beginning and collect baseline data prior to initiating program activities.
- **Seek to eliminate bias.** Use independent evaluators to reduce the perception of and potential for prejudiced decision-making.
- **Be relevant.** Evaluation questions should respond to the needs and interests of key stakeholders.
- **Use best tools and methods for purpose and context.** Consider feasibility and robustness when selecting an evaluation method, with particular consideration for mixed methods.
- **Ensure transparency.** Share findings widely and include a detailed description of methods.
- **Build local capacity.** Incorporate experts from the context in evaluation design and implementation.

*Source:* USAID Evaluation Policy (2011).

Finally, drawing from a review of 14 evaluations of market systems programs, Ruffer and Wach (2013) provide a number of recommendations for evaluating market systems programs:

- **Evaluation timeframe** should cover both implementation and the period following the close of the program. Many program evaluations share the same end date as the projects themselves, which limits the ability to assess the post-project sustainability of change. The timing of evaluation activities should coincide with the milestones anticipated in the program’s theory of change, but with flexibility built into the design for inevitable programmatic shifts.
- **Evaluation methods** should be selected that can verify the links in a programs’ theory of change, using mixed methods. The theory of change should incorporate external stakeholder perspectives, be regularly reviewed and embrace complexity.
- **Systemic change** should be incorporated into the theory of change for the intervention. It should define the boundaries of the system and the assumptions of how the system will change.
- **Evaluation of sustainability** should consider both static and dynamic concepts of sustainability<sup>2</sup> and incorporate a post-project assessment of impacts.
- **Evaluation of unintended consequences** (i.e., changes that a program did not purposely try to create) should be included in the analysis. Examples of negative unintended consequences include displacement of non-target populations, environmental damage, and changes in power or social relations.
- **Evaluation of indirect impacts** should be incorporated for the purpose of measuring the full positive and negative impacts of the program.

## C. EVALUATION-SUPPORTIVE MONITORING SYSTEMS

The monitoring community has been at the forefront of efforts to better understand how program interventions contribute to systemic change. Given the need to make decisions in the face of complexity that is implicit in taking a systemic approach, some implementers have refined their monitoring systems to support their learning and program adjustments under market system facilitation. The assumption by practitioners of a much more active role in articulating their theory of change and collecting data about all levels of results can greatly assist the ability to evaluate market systems programs.

The DCED Standard represents the leading effort to support this approach. The DCED Standard has been applied in over 30 market systems programs. It outlines a process that market systems programs can use for monitoring and developmental evaluation of their initiatives (DCED 2013). The eight components of the DCED Standard are as follows:

1. **Articulating Results Chains.** Results chains visually represent the change process through which project activities are expected to lead to intended impacts, showing the anticipated causal links and

---

<sup>2</sup> Static concepts of sustainability relate to “the extent to which the *status quo* (in terms of the results achieved through an intervention) will be maintained after external support is withdrawn”, while dynamic concepts of sustainability relate to “structural change to the market system which enhances its resilience to shocks and stresses through evolution or innovation in response to changing external factors” (Ruffer and Wach 2013).

relationships between them. They clearly demonstrate what the project is doing and the sequence of changes that are expected as a result.

2. **Defining indicators of change.** An indicator specifies what projects will measure in order to see whether change has occurred. Defining indicators on the basis of the results chain allows projects to develop an appropriate monitoring plan.
3. **Measuring changes in indicators.** Once indicators have been defined, projects develop and implement a monitoring plan that conforms to good research practice.
4. **Estimating attributable changes.** Once a change is measured, the extent to which that change is due to the intervention, rather than to other influences, should be assessed. For example, an increase in jobs may be due to an intervention, to exogenous factors, or to a combination of the two.
5. **Capturing wider changes in the system or market.** Many programs aim to affect entire market systems, and, where this is the case, the results of these changes need to be captured.
6. **Tracking program costs.** In order to assess the value for money of the intervention it is necessary to know how much was spent in achieving the reported results.
7. **Reporting results.** Findings should be communicated clearly to funders, local stakeholders, and to the wider development community, where possible.
8. **Managing the system for results measurement.** The results measurement system should be sufficiently resourced and integrated into project management, informing implementation and guiding strategy.

Projects applying the DCED Standard determine the appropriate amount of documentation that serves management purposes while enabling flexibility. The DCED Standard includes a mechanism for programs to arrange an external audit by independent auditors. This audit provides a means of certifying the quality of the measurement system in place and thus the quality of the results. This audit process is, however, not intended to replace the role of an independent evaluator. The methodology advocated by the DCED Standard supports independent evaluations in the following ways (Calvert 2014): i) by improving the quality of monitoring data, ii) by articulating detailed theories of change at the intervention level, iii) by estimating how much of the observed changes can be attributed to the intervention, iv) by attempting to measure systemic change, and v) by tracking data on program costs.

A recent paper examining the implications of complexity for monitoring in complex environments suggests a number of monitoring approaches (Britt 2014). The paper argues for establishing indicators that explain progress towards systemic change (i.e., “leading indicators”), and remaining vigilant to identify unexpected positive and negative change and alternative causal pathways. Evidently, systems change evaluation needs to have strong links with program monitoring.

## **D. PRACTICAL GUIDANCE**

The discussion above suggests the need for an evaluation framework that addresses the challenges of evaluating systems and systemic change. Surprisingly little guidance currently exists that is specifically tailored to evaluating market system interventions. On a more general level, Hargreaves (2010) outlines a helpful three-step approach for incorporating systems thinking into evaluation planning (a planning worksheet for using this approach is included in the annotated bibliography):

1. **Assess the dynamics of the system.** The context shapes how and if systemic change will or will not occur. Evaluators should identify relevant stakeholders, their relationships, and perspectives relating to potential shifts. The nature of relations between actors within a system and their attitudes are critical to allowing or blocking change. Evaluators should draw from this understanding to set the boundaries of the evaluation;
2. **Determine the dynamics of the intervention.** A second driver of the evaluation approach is the nature of the intervention seeking to create systemic change. Evaluators should understand how the intervention is governed (i.e., the management structure, the mix of partners involved), the intervention's theory of change, and the anticipated outcomes; and
3. **Select the appropriate systemic change evaluation approach.** The final step is to understand who the users of the evaluation will be and the purpose of conducting the evaluation. This will suggest the most appropriate evaluation methods.

While the literature review did not yield a comprehensive framework to guide evaluation of market system facilitation, it did imply that an essential component of such a framework would include a causal model that answers the following questions:

- What specific actions will the intervention undertake?
- How is the market system expected to change as a result of these actions?
- How will these systemic changes contribute to achieving inclusive growth outcomes?

# V. INDICATORS OF SYSTEMIC CHANGE

This section describes several frameworks that are being used to identify indicators of systemic change. None of these frameworks were developed specifically for use in evaluation, and several are in draft form. Nevertheless, they each contain elements that can inform USAID’s evaluation policies and practice around approaches for measuring systemic change. None of the following resources focus on indicators that could measure the ultimate impacts on target beneficiaries; rather they focus on testing the changes in systems that could ultimately lead to those impacts. The omission of indicators focused at the target beneficiary level does not imply that measuring such changes is not important, but rather that it is measuring systemic change that has been less thoroughly studied to date and thus where focus is particularly important.

## A. EVALUATION FRAMEWORKS

### DCED INDICATORS

The DCED’s guidance on systemic change (Kessler and Sen 2013) outlines five aspects of systemic change:

- **Crowding in:** The program helps targeted enterprises provide a new service, by supplying training or improving the market environment. Other enterprises see that this service can be profitable, and start supplying it as well. For example, a program helps agricultural suppliers start up pesticide spraying services. Other agricultural input suppliers, who did not receive any direct input from the program, may then start up a similar pesticide spraying service.
- **Copying:** The program improves the practices of targeted enterprises, to improve the quality or efficiency of production. Other entrepreneurs can see the positive impact of these new practices, and adopt them in their own business. For example, a shoe-making entrepreneur who sees that his rival has improved the quality of his shoes copies the quality improvements and so also gets higher prices for his goods.
- **Sector growth:** Program activities cause the targeted sectors to grow. Consequently, existing enterprises expand their businesses and new entrants come into the market.
- **Backward and forward linkages:** Changes in the market can trigger changes at other points along the value chain. For example, a program increases the amount of maize cultivated. This benefits not just farmers, but others in the value chain, such as truck drivers who transport maize. They receive more business as there is a greater volume of maize to transport.
- **Other indirect impact:** As a result of program activities, other indirect impacts may occur in completely different sectors. For example, if a program increases the income of pig producers, they may spend more on consumer goods, benefiting shops in the local area.

Of the five aspects of systemic change, the first two (crowding in and copying) represent imitation and replication of business models, technologies and behaviors by other market actors. The last three indicators describe second-order or multiplier effects that are created by the first two. These last three aspects of systemic change—sector growth, backward and forward linkages, and other impacts—are different from the first two

in that they describe increases in income and business growth rather than replication of a specific business model.

## AECF INDICATORS

The Africa Enterprise Challenge Fund (AECF), a multi-country initiative in Africa that funds innovative business proposals, has developed its own indicators of systemic change. These are similar to and draw from the DCED indicators, but incorporate additional aspects, such as innovation that occurs when market actors go beyond any of the new practices promoted by a project (Kessler 2013). In addition, there are inconsistencies in the definitions used in the AECF and DCED indicators, most notably the definitions for copying and crowding in (see table 2).

**Table 2: African Enterprise Challenge Fund's Indicators of Systemic Change**

Type of Systemic Change	Example
<p><b>Copying by other businesses</b></p> <p>Other businesses see the benefits of the grantee's business model, and so copy the idea.</p>	<p>The challenge fund provides seed finance to support an out-grower scheme, which purchases tomatoes from poor small-holder farmers. This is a financial success, and other companies copy the business model and begin to work with small-holder tomato farmers.</p>
<p><b>Crowding in</b></p> <p>Other businesses are encouraged into the space created by the grantee. The distinction between this and the previous category is that other businesses do not copy the business model, but offer supplementary services which are only viable because of the AECF grantee.</p>	<p>The challenge fund provides a grant to a seed supplier to set up shops in rural areas. A financial service provider, not funded by the challenge fund, works with the seed supplier to provide microfinance to farmers who wish to buy the seed.</p>
<p><b>Copying successful practice</b></p> <p>People who are not working with the project copy the behaviors or technologies that the project introduced. While the previous two categories refer to behavior change in businesses, this refers to behavior change among farmers and others.</p>	<p>The challenge fund provides a grant to an outgrower scheme, which teaches sustainable farming techniques to participating farmers. Other nearby farmers copy these techniques and thus improve their yields.</p>
<p><b>Business regulatory environment</b></p> <p>All projects work within a regulatory environment, principally defined by the government. They must follow laws and regulations, and work with government officials to gain permission to work, export, etc. Many companies seek to improve the regulatory environment, to make it easier for them to do business.</p>	<p>The challenge fund provides a grant to a number of livestock businesses that import vaccines. Regulations for importing vaccines are time-consuming and cumbersome to follow, and government officials regularly ask for bribes. The businesses join together to pressure the government to bring about changes in regulations and reform in government practices.</p>
<p><b>Factor and other market systems</b></p> <p>Changes in factor market systems are changes that the project causes in the main factor market systems of land, labor and capital, but also include ancillary markets such as information.</p>	<p>In the above example for crowding in, financial service organizations provided financial services to customers of a seed supplier. If those organizations also begin to provide financial services to other people and businesses unrelated to the grantee, this indicates a change in the financial market system, as there is improved access to finance.</p>

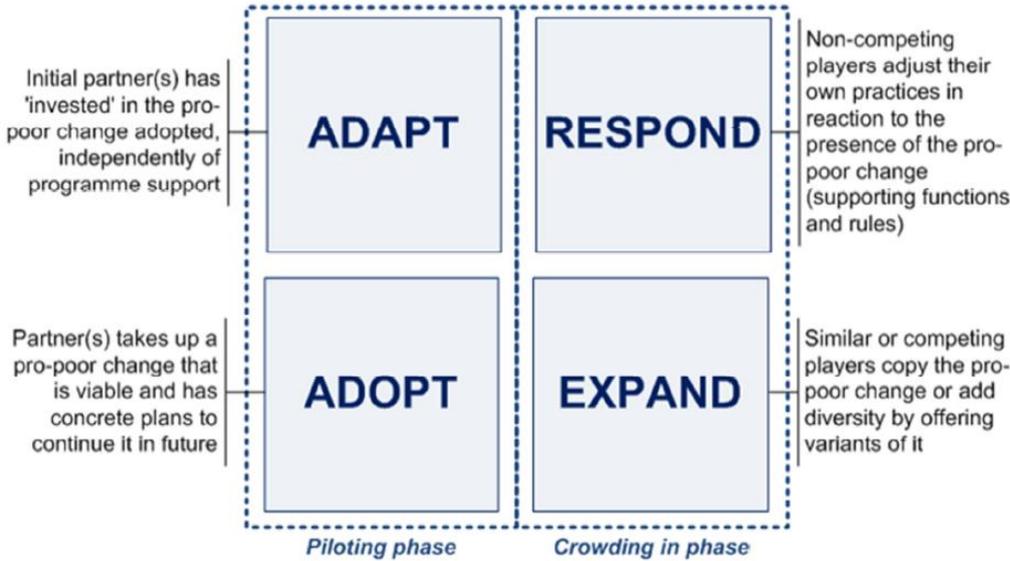
<p><b>Innovation</b></p> <p>The grantee introduces additional innovations that were not in the original business plan, but which were developed as a result of the AECF-funded project.</p>	<p>The challenge fund provides funding to a pesticide company to develop a new type of organic pesticide for a certain pest. Although the original design did not work, it led to the creation of a new type of pesticide effective against a different pest.</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Source: Adapted slightly from (Kessler 2013).

**SPRINGFIELD/KATALYST SYSTEMIC CHANGE FRAMEWORK**

The Springfield Centre, in conjunction with the Katalyst program, has developed a draft systemic change framework (Springfield Centre 2014). The framework outlines four elements that indicate systemic changes that occur as a market system is evolving. The “adopt” stage is not considered a systemic change itself, but rather an initial step that may then lead to change in any of the other three elements. The Springfield/Katalyst framework is presented in figure 2. The four elements in the framework were used to identify a number of systemic change indicators, which are listed in table 3.

**Figure 2: Springfield Centre/Katalyst Systemic Change Framework**



**Table 3: Springfield/Katalyst Systemic Change Indicators**

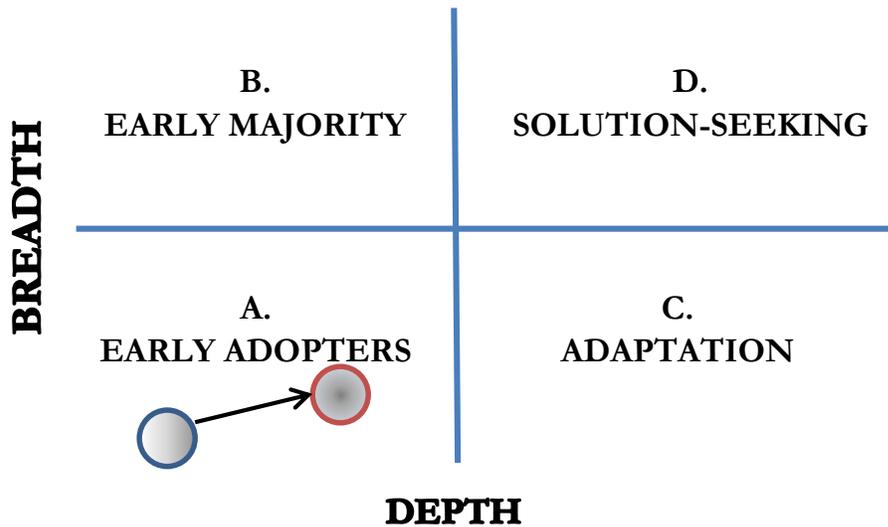
Elements	Suggested Indicators
Adopt	Independent investment Target group benefits sustained
Adapt	Partner contribution to the pilot Long-term viability/benefit of practice change Partner satisfaction and intent to continue Partner ability to continue Target group’s satisfaction and benefit
Expand	Competitors or similar types of organizations ‘crowd-in’ Ability to accommodate competition or collaboration (depends on the nature of the system)
Respond	System responsiveness and receptiveness

Source: Springfield Centre (2014)

### KENYA MAP SYSTEMIC CHANGE BENCHMARKING TOOL

The Market Assistance Programme (MAP) in Kenya, implemented by Kenya Markets Trust, has developed a four-part framework for mapping behavior changes that indicate systemic change. The framework is informed by the idea that “the presence (or absence) of continuous adaptation to external opportunities or threats is a sign of a competitive, solution seeking system” (Osorio-Cortes et al. 2013). The framework, as shown in figure 3, distinguishes between the breadth of systemic change (i.e., change across the sector, such as the number of players that are adopting a new behavior), and the depth of systemic change (i.e., change within the firm, such as the types of behavior changes that market players are adopting). Two of the categories—early adopters and early majority—were drawn from the technology diffusion literature. The framework was used to identify the indicators listed in table 4.

**Figure 3: Kenya Market Assistance Programme Behavior Change Framework**



**Table 4: Kenya Market Assistance Programme Systemic Change Indicators**

Indicator Category	Area of Focus
Behavior change	Investment patterns Technology use Relationships with suppliers and buyers Strategy change: from price focus to value-addition focus
Trust	Transparency about quality of agricultural inputs Possibility and freedom to choose between different types of products, qualities, and prices Win-win outcomes Friendship and strategic alliances Convergence of objectives, mainly around mutual growth
Loyalty	Long-term relationships based on mutual interests and policies or norms that promote and enforce the rule of law

Consumer awareness	Consumers' appreciation of value addition by the businesses from which they buy
Business management patterns	Human resources Production processes Information Decision making
Participation in policy change and advocacy	Who participates Who should participate, and why they are or are not participating Interactions and collaborations to change policies Accountability mechanisms Enforcement mechanisms
Relationships between actors	Improved or new relationships The factors and motivations that bring the actors together
Repeat sales	Improved or new relationships Sustainability of relationships Changes in investment patterns Increased freedom of choice Increased product/service quality Client-oriented business strategies
Perceptions and preconceptions	Of other actors Of self (how actors perceive themselves) Stigma Peer pressure
Knowledge nodes, structures, and flows	Who produces, stores, and keeps knowledge up to date How information and knowledge are flowing throughout the system How existing knowledge is combined to produce new knowledge How collaboration for innovation is happening and who is participating

*Source:* Osorio-Cortes, Jenal and Brand (2013)

## B. COMPARISON OF INDICATORS

The proposed indicators of systemic change, as outlined above, exhibit both similarities and differences. The Springfield/Katalyst and MAP indicators explicitly include adoption of a program-supported business model among their indicators of systemic change, while the others consider that to be program-created and reversible. Further, Springfield/Katalyst, MAP and AECF consider firms' innovations to a program-introduced business model as evidence of systemic change. Only the Springfield/Katalyst indicators explicitly include innovations by other market actors as a type of systemic change, caused by the introduction of new business models. The replication of a program-supported business model by other businesses through a process of crowding-in is considered an indication of systemic change by all four entities. Only the DCED and AECF explicitly define copying of behavior at the target beneficiary level as a type of systemic change. Finally, the DCED is the only entity that includes the multiplier effects of additional spending and economic activity generated by a program on the growth of other businesses in the target sector and non-target sectors. The four sets of proposed indicators are compared in table 5.

**Table 5: Comparison of Proposed Systemic Change Indicators**

Entity	DCED	Springfield/ Katalyst	MAP	AECF
Initial adoption by program-supported partner		X	X	
Adaptation or innovation by program-supported partner		X	X	X
Innovations by other market actors		X		
Crowding-in by other market actors	X	X	X	X
Copying by target beneficiaries	X			X
Multiplier effects on other businesses	X			

Source: Fowler (2014)

## C. CATEGORIES OF INDICATORS

A review of some of the most recently proposed indicators of systemic change suggests two relatively universal categories of systemic change:

1. **Buy-in indicators** measure the degree to which market actors have taken ownership over the new business models, technologies, practices and behavior changes that were introduced and/or supported by the intervention. Some examples of buy-in indicators include the following:
  - Adaptation or innovation to the original, program-sponsored model(s)
  - Continued, independent investment after program sponsorship ends
  - Repeat behavior
  - Satisfaction with program-facilitated changes
2. **Imitation indicators** measure the scale or breadth of program-supported behavior change within a system. There are two prominent examples of imitation indicators:
  - Crowding-in by other businesses that imitate program-sponsored business models originally adopted and demonstrated by business(es) that collaborate with the implementer
  - Copying, mentioned less often than crowding-in, refers to imitation at the target beneficiary level by market actors (firms, farms, households or individuals) that imitate the new practices originally adopted and demonstrated by the target beneficiaries of the intervention

# VI. SUMMARY AND CONCLUSION

This literature review examined current thinking on evaluating systems and systemic change for inclusive market development. It was conducted as the first step in defining a research agenda to inform evaluation practice for market systems facilitation interventions. Individual articles, reports and presentations that were reviewed for this study are summarized in the annotated bibliography, provided as an annex. This section summarizes some of the findings as they relate to defining systemic change, incorporating systems thinking into evaluation frameworks, and identifying useful indicators for measuring systemic change.

## A. DEFINING SYSTEMIC CHANGE

There is no consensus on how to define a system and systemic change. The literature review suggests that definitions of systemic change within the context of market system facilitation should incorporate several elements, including: i) recognition that the causes of systemic change are diverse and overlapping, including donor-funded interventions and emergent solutions from within the system itself; ii) acknowledgement that impacts of systemic changes are equally diverse, including both those that are positive and negative from the perspective of a facilitator's objectives, and iii) understanding that systemic change is an intermediate outcome distinct from but that can contribute to final development outcomes for target beneficiaries. The literature indicates that evaluations of systemic change in market systems programming should assess both systemic changes themselves and also the resulting development impacts for target populations.

## B. FRAMEWORKS FOR EVALUATING SYSTEMS

The literature review suggests that there is no comprehensive framework for evaluating systemic change of market systems interventions. Any of three general categories of evaluation—summative, formative and developmental—can be adapted to incorporate systems thinking through explicit attention to relationships, boundaries and perspectives (Williams and Hummelbrunner 2011; Hargreaves 2010; Britt 2013). Nevertheless, evaluating market systems facilitation interventions is constrained by a number of factors: i) the deliberate push for non-target individuals and firms to replicate project-supported models, which can contaminate control groups; ii) the implementation of multiple, sequenced interventions targeting multiple levels (micro, meso, macro) of the market system, which results in varying degrees of participation and benefit for beneficiaries; iii) the adaptive nature of facilitation interventions, which requires greater flexibility in evaluation methods, and iv) the unpredictability of systemic change, which confounds evaluation timing.

In response to these challenges, there is increasing support for the view that the complexity of attributing change to a specific intervention favors the estimation of the contribution of an intervention to an observed change. With its frequent feedback loops and adaptive flexibility, developmental evaluation (Patton 2011) appears to be well suited to support learning around the early results of systemic interventions under conditions of complexity, where the response of the system to the intervention is unpredictable. In addition, there are arguments in favor of methodological heterodoxy—supporting the use of mixed methods and triangulation of evidence—based at least partially on the need for evaluation results that are strong in terms of both internal and external validity.

Donors provide foundational guidance on evaluating development interventions, including USAID's Evaluation Policy (2011) and DFID's International Development Evaluation Policy (2013). Evaluation frameworks created specifically for value chain and market systems interventions are also relevant, including the Degrees of Evidence principles developed under USAID's AMAP project (Creevey et al. 2010) and the more recent

evaluation principles compiled by Ruffer and Wach (2013). The DCED Standard for Results Measurement (2013) provides a wealth of helpful guidance on how to establish project-managed monitoring systems that support elements of evaluation for interventions based on market system facilitation.

The roles of evaluators are mentioned in ways that are sometimes contradictory. While most foundational frameworks call for independent evaluators (USAID 2011; DFID 2013), Ruffer and Wach (2013) recommend shared evaluation responsibility between evaluators and implementers. Collaboration between evaluators and implementers is recommended at several points in the project cycle, beginning with project design and continuing through process evaluation. The DCED encourages projects to consider contracting independent auditors to assess their compliance with the DCED Standard, and has a published mechanism to share audit results, though it does not suggest that this necessarily replaces the need for independent evaluation. In fact, the implementation of the Standard can actually support such evaluation (Calvert 2013). The unpredictability and slower pace of systemic change suggests that longer evaluation timelines may be needed, stretching often past the closure of project activities.

While the context might be best described as a complex and unpredictable system, there is still a need for theories of change to guide implementation and provide a framework for building an evidence base on the linkages between market system interventions and inclusive economic growth. The literature review suggests that such theories of change should incorporate an initiative's specific activities, the expected changes in the market system, and the anticipated contributions of those systemic changes to development outcomes.

### **C. INDICATORS FOR EVALUATING SYSTEMIC CHANGE**

The selection of indicators for evaluating systemic change can also be informed by distinguishing between the market system and the intervention designed to facilitate changes in the system. Systemic change is not a final outcome, but an instrumental step toward achieving outcomes such as improved incomes, employment and food security, and reduced poverty. This implies that indicators of systemic change should not be defined in terms of final development outcomes. Instead, systemic change indicators should be defined in terms of shifts in the underlying or structural elements and patterns that characterize a system, such as the quality of the relationships between actors.

A review of some of the most recently proposed indicators of systemic change suggests two relatively universal categories of systemic change: i) buy-in indicators, measuring the degree to which market actors have taken ownership over the new business models, technologies, practices and behavior changes that were introduced and/or supported by the intervention, and ii) imitation indicators, measuring the scale or breadth of program-supported behavior change within a system.

These concepts related to buy-in and imitation represent a starting point in identifying indicators for detecting and measuring systemic change. Better ways to evaluate systems and systemic change, along with principles and frameworks that have been adapted to meet the challenges associated with evaluating interventions based on market systems facilitation, would help to promote learning, inform programming and improve outcomes in this important area of development programming.

# REFERENCE LIST

- Andrews, Matt, Lant Pritchett and Michael Woolcock. 2012. “Escaping Capability Traps through Problem-Driven Iterative Adaptation (PDIA).” CGD Working Paper 299. Washington DC: Center for Global Development.
- Boston, Jonathan. 2000. “The Challenge of Evaluating Systemic Change: The Case of Public Management Reform.” *International Public Management Journal* 3(2000): 23–46.
- Britt, Heather. 2013. “Complexity-Aware Monitoring.” USAID Discussion Note, Monitoring & Evaluation Series. Washington DC: DevTech Systems Inc.
- Calvert, Simon. “Evaluation and the DCED Standard for Results Measurement.” 2014. London: Department for International Development.
- Creevey, Lucy, Jeanne Downing, Elizabeth Dunn, Zan Northrip, Don Snodgrass, and Amy Cogan Wares. 2010. “Assessing the Effectiveness of Economic Growth Programs.” AMAP Private Sector Development Impact Assessment Initiative. Bethesda MD: DAI.
- DAI. 2010. PROFIT Zambia Impact Assessment Final Report. Bethesda MD: DAI.
- Deprez, Steff. 2013. “The Use of Outcome Mapping in Value-Chain Development Programmes: The Case of Vredeseilanden (VECO).” OM Ideas No. 7, Outcome Mapping and Learning Community. London: Overseas Development Institute.
- DFID. 2013. “International Development Evaluation Policy.” London and Glasgow: Department for International Development.
- DFID and SDC. 2008. “The Operational Guide for the Making Markets Work for the Poor (M4P) Approach.” London and Bern: Department for International Development and Swiss Agency for Development and Cooperation.
- DCED. 2013. Standard for Results Measurement. Version VI. Cambridge: Donor Committee for Enterprise Development.
- Eoyang, Glenda and Thomas Berkas. 1999. “Evaluating Performance in a Complex Adaptive System.” In *Managing Complexity in Organizations: A View in Many Directions*, M. Lissack and H. Gunz (eds.). Westport CT: Quorum Books.
- Fowler, Ben. 2014. “Systemic Change and the DCED Standard.” Ottawa, Canada: MarketShare Associates.
- Frankenberger, Tim and Suzanne Nelson. 2013. “Resilience Measurement for Food Security.” Background Paper for Expert Consultation, Food and Agriculture Organization of the United Nations and the World Food Programme. Tucson AZ: TANGO International.
- Hargreaves, Margaret B. 2010. “Evaluating System Change: A Planning Guide.” Methods Brief. Princeton NJ: Mathematica Policy Research, Inc.
- Hargreaves, Margaret, Marah Moore and Beverly Parsons. 2010. “Useful Tools for Integrating Systems Concepts into System Change Evaluations.” Professional Development Workshop at the American Evaluation Association Annual Conference, November 8-14, San Antonio TX.

- Johnson, Susan and Jean Boulton. 2014. "Impact Assessment of Financial Market Development through the Lens of Complexity Theory." Centre for Development Studies, University of Bath.
- Kania, John and Mark Kramer. 2013. "Embracing Emergence: How Collective Impact Addresses Complexity." Stanford Social Innovation Review, Blog (January 21). Washington DC: FSG.
- Kessler, Adam. 2013. "Measuring Results in Challenge Funds: Practical Guidelines for Implementing the DCED Standard." Cambridge: Donor Committee for Enterprise Development.
- Kessler, Adam and Nabanita Sen. 2013. "Guidelines to the DCED Standard for Results Measurement: Capturing Wider Changes in the System or Market." Cambridge: Donor Committee for Enterprise Development.
- Marks, James and Pete Wong. 2010. "Catalyzing Systemic Change: The Role of Venture Philanthropy." London, UK: Impetus Trust and Collier Institute of Private Equity.
- Olson, Edwin, and Glenda Eoyang. 2001. *Facilitating Organization Change: Lessons from Complexity Science*. San Francisco, CA: Jossey-Bass/Pfeiffer.
- Osorio-Cortes, Lucho and Marcus Jenal. 2013. "Monitoring and Measuring Change in Market Systems: Rethinking the Current Paradigm." MaFI Synthesis Report. Arlington VA: The SEEP Network.
- Osorio-Cortes, Luis, Marcus Jenal and Margie Brand. 2013. "Monitoring and Measuring Change in Market Systems: The Systemic M&E Principles in the Context of the Kenya Market Assistance Program." MaFI Case Study. Arlington VA: The SEEP Network.
- Parsons, Beverly and Meg Hargreaves. 2009. "Evaluating Complex Systems Interventions." Professional Development Workshop at the American Evaluation Association Annual Conference, November 9-15, Orlando FL.
- Patton, Michael Quinn. 2011. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford Press.
- Preskill, Hallie and Tanya Beer. 2012. "Evaluating Social Innovation." FSD and Center for Social Innovation.
- Reynolds, Martin, Kim Forss, Richard Hummelbrunner, Mita Marra and Burt Perrin. 2012. "Complexity, Systems Thinking and Evaluation - An Emerging Relationship?" *Evaluation Connections Newsletter of the European Evaluation Society* (December): 7-9.
- Ripley, Matthew and David Nippard. 2014. "Making Sense of Messiness: Monitoring and measuring change in market systems: a practitioner perspective." Samarth-NMDP and The Springfield Centre.
- Ruffer, Tim and Elise Wach. 2013. "Review of Making Markets Work for the Poor (M4P) Evaluation Methods and Approaches." DFID Working Paper 41. London: Department for International Development.
- Sarriot, Erin, Jim Ricca, Jennifer Yourkavitch, Leo Ryan, and the Sustained Health Outcomes (SHOUT) Group. 2008. "Taking the Long View: A Practical Guide to Sustainability Planning and Measurement in Community-Oriented Health Programming." Calverton MD: Macro International Inc.
- Sen, Nabanita. 2010. *A Walk Through the DCED Standard for Measuring Results in PSD*. DCED.

- Springfield Centre. 2014. "Good Practices in Facilitation: The System Change Framework." Durham UK: The Springfield Centre for Business in Development.
- Stern, Elliot, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies and Barbara Befani. 2012. "Broadening the Range of Designs and Methods for Impact Evaluations." DFID Working Paper 38. London: Department for International Development.
- TANGO International. 2013. "Summary of the Expert Consultation on Measuring Resilience." Summary of expert consultation held in Rome, February 19-21. Tucson AZ: Tango International.
- USAID. 2012. "Understanding Facilitation." Washington DC: United States Agency for International Development.
- USAID. 2014. "Local Systems: A Framework for Supporting Sustained Development." Washington DC: United States Agency for International Development.
- USAID. 2011. "Evaluation: Learning from Experience." USAID Evaluation Policy. Washington DC: United States Agency for International Development.
- Williams, Bob and Richard Hummelbrunner. 2011. *Systems Concepts in Action: a Practitioner's Toolkit*. Stanford: Stanford University Press.
- Woolcock, Michael. 2009. "Towards a Plurality of Methods in Project Evaluation: A Contextualised Approach to Understanding Impact Trajectories and Efficacy." BWPI Working Paper 73.

# ANNEX: ANNOTATED BIBLIOGRAPHY

**Andrews, Matt, Lant Pritchett and Michael Woolcock. 2012. “Escaping Capability Traps through Problem-Driven Iterative Adaptation (PDIA).” CGD Working Paper 299. Washington DC: Center for Global Development. (26 pages)**

This paper outlines the longstanding failure of donor-funded institutional reform efforts to improve the capacity of developing country governments. It identifies capability traps as a challenge, where “state capability stagnates, or even deteriorates, over long periods of time even though governments remain engaged in developmental rhetoric and continue to receive development resources” (p.2). The authors argue that development funding for institutional reform in certain cases actually contributes to capability traps, by constraining local innovation while enabling governments to engage in “isomorphic mimicry”, in which they make reforms that increase their legitimacy with international stakeholders but do not actually change fundamental decision-making processes (p.2). Seeking to apply international best practice via linear processes, rigid monitoring of inputs, and compliance to a predetermined plan all contribute to capability traps.

To support institutional change, the paper proposes a “Problem-Driven Iterative Adaptation” (PDIA) approach. PDIA is based on four key principles:

1. Local definition of performance problems, rather than importation of “best practice” solutions;
2. Support for local experimentation, rather than requiring adherence to a predetermined implementation plan;
3. Design of tight feedback loops for “rapid experiential learning”, rather than waiting for the results from ex-post evaluation; and
4. Engagement of a broad group of actors designing local solutions to locally defined problems, rather than top-down imposition of solutions from a small group of external experts.

They advocate for M&E systems that allow interventions to evolve with learning, and thus are critical of the overuse of randomized control trials where they restrain rapid innovation and adaptation. They see results measurement as ultimately supporting the development of local solutions, drawing from international good practices but making adaptations and hybrids that respond to the local context.

**Boston, Jonathan. 2000. “The Challenge of Evaluating Systemic Change: The Case of Public Management Reform.” *International Public Management Journal* 3(2000): 23–46.**

The challenges of evaluating systemic change are described in detail and illustrated with the example of New Zealand’s radical transformation of its public sector management system (1985-1990). The author asserts that a thorough appraisal of this systemic reform has never been done because of these evaluation challenges (p.26):

1. “Choosing the appropriate criteria for evaluation and determining what constitutes ‘success’;
2. Determining and securing the relevant evidence;

3. Interpreting the available evidence, including the problem of establishing appropriate counterfactuals and determining causation; and
4. Arriving at an overall assessment.”

Specific examples for each of these four challenges are provided in Table 1 (p.27).

**Britt, Heather. 2013. “Complexity-Aware Monitoring.” USAID Discussion Note (version 2.0), Monitoring & Evaluation Series. Washington DC: DevTech Systems Inc. (19 pages)**

This discussion note is written to inform USAID’s thinking on complexity-aware monitoring as a complement to performance monitoring for complex areas of projects. The resource first outlines the situations when complexity-aware monitoring is appropriate, such as those where the relationships between cause and effect are not well understood. It suggests asking the following questions to identify such situations (p.2):

- “What is the degree of certainty about how to solve the problem?”
- “What is the degree of agreement among stakeholders about how to solve the problem?”

Situations with a low degree of certainty and agreement are characterized as complex. In such situations, the paper suggests three principles:

1. **Synchronize monitoring with the pace of change.** In highly dynamic contexts, using leading or coincident indicators will better enable projects to identify and adapt to change in a timely way.
2. **Attend to performance monitoring’s three blind spots.** The piece argues that USAID’s linear approach to performance monitoring makes it likely to miss the broader range of negative and positive outcomes, alternative causes of observed change, and “the full range of non-linear pathways of contribution” (p.6). A meta-analysis of USAID evaluations found that very few evaluations address these issues.
3. **Attend to relationships, perspectives and boundaries.** The author argues that all three concepts should be applied to understanding complex environments, particularly through participatory monitoring approaches.

The discussion note also recommends five approaches to complexity-aware monitoring as a starting point for experimentation within USAID:

1. **Sentinel Indicators:** a sentinel indicator is defined as “an indicator which captures the essence of the process of change affecting a broad area of interest and which is also easily communicated.” (p.7) These can be selected by mapping out key relationships between the project, other actors and influencing factors, and selecting measures at key leverage points that indicate their important systemic changes. The author acknowledges the tension between “indicator-based monitoring” and monitoring in less predictable, complex environments, where indicators should be expected to evolve with the program.
2. **Stakeholder Feedback:** Collecting information from stakeholders can assist in bolstering project effectiveness in complex environments. This is because knowledge of a system is partial and stakeholders’ perceptions help to shape their behaviors and consequently outcomes. Although the author recognizes several challenges with this approach (e.g., sampling error, cost, technical difficulty), she recommends it as a way to gather valuable information for dealing with complexity. This particularly includes determining the boundaries of a system.

3. **Process Monitoring of Impacts:** This approach mirrors the results chain approach advocated by the DCED Standard by describing the processes that link project outputs to expected outcomes. It is seen to address complex situations by enabling rapid identification of “results-producing processes”. Moreover, by providing a richer description of the project and its environment than traditional USAID tools, it enables the identification of “alternative causes, multiple causal pathways, and feedback loops” (p.10). The author suggests that one drawback is an inability to capture positive or negative unintended consequences, for which users will need to remain vigilant to observe in other ways, such as by incorporating diverse perspectives on project processes and results.
4. **Most Significant Change:** The most significant change (MSC) methodology is a qualitative methodology that allows project stakeholders to report on the most important results that they have observed as a result of project intervention and the reasons for this. MSC outlines “domains of change” that results are grouped under, but keeps them broad to avoid predetermining the types of feedback that will be received. The stories that are collected may be verified and/or quantified.
5. **Outcome Harvesting:** Outcome harvesting is a qualitative approach to capturing outcomes, then linking them to plausible contributors. These causal relationships are then validated by the monitors through processes including the triangulation of perspectives. It is somewhat more defined than the MSC approach, and focused on the application of the information that is generated.

The author finds several divergences between the five approaches and current practice at USAID. For instance, the last two approaches are described as “indicator-free” and “goal-free”, because they do not articulate expected results prior to information gathering. Process monitoring of impacts and stakeholder feedback are indicator-optional. The paper closes with a guide for USAID staff seeking to apply the approaches outlined in the paper.

**Department for International Development (DFID) and Swiss Agency for Development and Cooperation (SDC). 2008. “The Operational Guide for the Making Markets Work for the Poor (M4P) Approach.” London and Bern: DFID and SDC. (124 pages)**

The Making Markets Work for the Poor (M4P) approach is a framework for analyzing and intervening in market systems. The M4P approach is designed to be flexibly applied to a diverse array of systems, including agricultural, educational and health systems. Facilitating systemic change is one of the five M4P components. The term “facilitating system change” is used roughly as a synonym for intervening and seen as core to achieving sustainable impact. Drawing from its market systems framework, the publication outlines four types of systemic change (p. 91):

1. Improved delivery of the core transaction under focus (by increased sales, higher satisfaction, etc.);
2. “Changes in practices, roles and performance of important system players and functions”;
3. “Crowding-in of system players and functions”; and
4. “Demonstrated dynamism of system players and functions (e.g., responsiveness to changed conditions in the system)”.

Among these four types of systemic change, crowding-in is highlighted as critical along the pathway to successfully exiting a market facilitation role. The publication guides practitioners on how to effectively facilitate systemic change in their market systems of focus.

**Deprez, Steff. 2013. “The Use of Outcome Mapping in Value-Chain Development Programmes: The Case of Vredeseilanden (VECO).” OM Ideas No. 7, Outcome Mapping and Learning Community. London: Overseas Development Institute. (8 pages)**

The paper is one in a series of case studies showcasing the real-world application of the Outcome Mapping (OM) methodology in different contexts. It describes the application of the OM methodology to an agricultural value chain program implemented by the Belgian NGO VECO from 2008 to 2013. Importantly, it describes how VECO implemented the OM methodology while also developing and reporting on a logframe as required by its donor. Over phase 1 (2008-2010), the program applied the components of the OM methodology, including outcome challenges, progress markers and strategy maps, for each type of boundary partner: private sector actors, farmer organizations, consumer organizations, NGOs, and network organizations. An innovation that VECO made to the OM methodology was to create a specific objective around organizational improvement for the implementing partner itself.

Learning from experience, VECO changed its approach in phase 2 (2011-2013). Rather than creating outcome challenges, progress markers and strategy maps for each type of boundary partner, it prepared them for each value chain that it was focused on. This reflected the fact that each value chain required a different strategy. Ultimately, VECO created 42 chain intervention frameworks. They also modified the OM methodology by adding a level in the chain intervention frameworks that outlined the explicit changes expected at the farmer level, making it more consistent with the logframe. This differs from normal OM practice, where changes are only defined for boundary partners. VECO developed standard capacity building indicators for its farmer organizations to measure overall change across the portfolio, in addition to tailor-made measures. During phase 2, the logframe was used to summarize progress across the portfolio in each geographical region.

Finally, the paper discusses practical management considerations in applying OM within VECO’s planning, learning and accountability system. For an M&E system to foster learning, it is important to a) agree on the main purposes and intended uses of the system, and b) align the M&E schedule with “organizational spaces and rhythms” that create sharing and learning opportunities. While applying the OM methodology does not automatically guarantee that the program is equipped to deal with complexity (p.7), it contains design elements that can support a process-oriented monitoring approach.

**Eoyang, Glenda and Thomas Berkas. 1999. “Evaluating Performance in a Complex Adaptive System.” In *Managing Complexity in Organizations: A View in Many Directions*, M. Lissack and H. Gunz (eds.). Westport CT: Quorum Books. (21 pages)**

This paper discusses the implications of complex adaptive systems for evaluation of organizational systems, using the assumption that organizations are complex adaptive systems. It describes the characteristics of complex adaptive systems, including:

- Dynamic
- Massively entangled
- Scale independent
- Transformative
- Emergent

The paper then discusses principles for addressing each of the above characteristics when evaluating complex adaptive human systems. It outlines tools and techniques to do so, which include:

- Causal diagrams (i.e., results chains)
- Iterative redesign
- Shorts and simples (short lists of simple rules to guide evaluation)
- Feedback analysis (based on Venn diagrams)
- Time series analysis

Finally, the paper suggests that evaluators need to shift their role by embracing uncertainty during the evaluation process and prioritizing learning as the outcome of an evaluation.

**Frankenberger, Tim and Suzanne Nelson. 2013. “Resilience Measurement for Food Security.” Background Paper for Expert Consultation, Food and Agriculture Organization of the United Nations and the World Food Programme. Tucson AZ: TANGO International. (42 pages)**

Written to inform an expert consultation on measuring resilience in food security and nutrition, this paper notes the absence of robust and reliable indicators of resilience at the household, community and national levels. The authors call for empirical evidence on the factors that support resilience or cause shocks, and in which contexts. The paper suggests there may not be universally relevant indicators of resilience across contexts or implementing agencies; rather, indicators may need to be more shock- or intervention-specific. It provides a detailed overview of current practices by international agencies (e.g., World Food Programme, the Food and Agriculture Organization of the United Nations), presents a resilience assessment framework, and closes with questions to consider in attempting to measure resilience.

**Hargreaves, Margaret B. 2010. “Evaluating System Change: A Planning Guide.” Methods Brief. Princeton NJ: Mathematica Policy Research, Inc. (24 pages)**

“This methods brief provides guidance on planning effective evaluations of system change interventions. It begins with a general overview of systems theory and then outlines a three-part process for designing system change evaluations. This three-part process aligns (1) the dynamics of the targeted system or situation, (2) the dynamics of the system change intervention, and (3) the intended purpose(s) and methods of the evaluation. Incorporating systems theory and dynamics into evaluation planning can improve an evaluation’s design by capturing system conditions, dynamics, and points of influence that affect the operation and impact of a system change intervention. The goal is to provide an introduction to system change evaluation planning and design and to encourage funders, program planners, managers, and evaluators to seek out more information and apply systems methods in their own evaluation work.” (p.2)

The brief provides guidance for applying the three-part planning process and includes a planning worksheet. The planning worksheet is reproduced below.

<b>System Change Evaluation Planning Worksheet</b>
<b>A. What is the situation?</b>

1. Describe the situation (its boundaries, parts, and whole).	
2. Describe the dynamics of the situation's relationships (random or unknown, simple, complicated, complex or combination)	
3. Describe the diversity of the purposes or perspectives within the situation.	
<b>B. What is the system change intervention?</b>	
4. Describe the dynamics of the intervention's governance or implementation (random or unknown, simple, complicated, complex or combination).	
5. Describe the dynamics of the intervention's theories of change and action (random or unknown, simple, complicated, complex or combination).	
6. Describe the diversity and specificity of the intervention's intended outcomes.	
<b>C. What are the goals of the system change evaluation?</b>	
7. Describe the evaluation's users, purpose(s) (developmental, formative, summative, monitoring for accountability, and/or other), and methods.	

Source: Hargreaves (2010).

**Hargreaves, Margaret, Marah Moore and Beverly Parsons. 2010. "Useful Tools for Integrating Systems Concepts into System Change Evaluations." Professional Development Workshop at the American Evaluation Association Annual Conference, November 8-14, San Antonio TX. (93 slides)**

This slide presentation provides a thorough look at how to incorporate systems thinking into all stages of an evaluation. After defining key concepts, it reviews the major schools of thought that have contributed to systems thinking:

- Early cybernetics
- Late cybernetics
- General systems theory
- Systems dynamic modeling
- Complexity theory

- Soft and critical systems
- Learning systems

It then reviews how to analyze a system systematically and describe a systems change intervention. Finally, the resource outlines how to conduct each of the four stages of an evaluation using a systems change approach:

1. Design evaluation
2. Collect data
3. Make meaning of data
4. Shape practice

The presenters focus on systems generically, using two in-depth examples of a city integration initiative and an early childhood education program. Three tools are presented to support systems change evaluations: Theory of Change in Paradigms, Structures, and Conditions of Complex Systems; the 7 Cs Framework; and ZIPPER.

**Kania, John and Mark Kramer. 2013. “Embracing Emergence: How Collective Impact Addresses Complexity.” *Stanford Social Innovation Review*, Blog (January 21). Washington DC: FSG. (8 pages)**

FSG defines *Collective Impact* as “the commitment of a group of important actors from different sectors to a common agenda for solving a specific social problem,”<sup>3</sup> which it contrasts with *isolated impact*. This article outlines how the collective impact approach is relevant to contexts characterized by complexity. The collective impact concept posits that “a highly structured cross-sector coalition” (p.1) is the best mechanism to achieve impact. Such a coalition has five conditions for success: a common agenda, shared measurement, mutually reinforcing activities, continuous communication, and backbone support.

This article argues that the typical project cycle – in which an organization develops a theory of change for its project, an evaluation is then conducted, and finally the results are used to scale-up the successful aspects of the model – does not function under conditions of complexity because predetermined solutions cannot be predicted or implemented. The complex interactions of multiple actors determine most outcomes, and thus in practice many successful interventions scale-up very slowly or not at all. Tackling adaptive problems – those problems for which there are no known answers but multiple stakeholders operating in uncertain and unpredictable environments – requires “learning by the stakeholders involved in the problem, who must then change their own behavior in order to create a solution.”<sup>4</sup> The paper provides various examples in which the collective impact approach has supported the creation of “emergent solutions.” For instance, one initiative in Canada stopped using logic models with their partners and instead adopted very regular (even every two weeks) analysis of changes. Another partner used an outcome diary to regularly track changes at the individual, partner relationship and policy levels. This helped capture unexpected “emergent dynamics.” The article

---

<sup>3</sup> FSG. Winter 2011. “Collective Action.” *Stanford Social Innovation Review*. p. 36. This article explains the concept of Collective Impact.

<sup>4</sup> FSG. Winter 2011. “Collective Action.” *Stanford Social Innovation Review*. p. 39.

does not focus on a single type of system, rather drawing from examples in education and policing (in the US) and poverty reduction (in Canada).

**Kessler, Adam. 2013. “Measuring Results in Challenge Funds: Practical Guidelines for Implementing the DCED Standard.” Cambridge: Donor Committee for Enterprise Development. (33 pages)**

These guidelines are oriented at practitioners seeking to comply with the DCED Standard for Results Measurement within the context of challenge funds. Challenge funds are a funding mechanism in which resources are allocated to market players. The guide “concentrates on private sector development challenge funds – often called ‘Enterprise Challenge Funds’ – which finance businesses in order to raise incomes, provide employment, and increase access to markets for the poor” (p. 2). The paper discusses several specific issues associated with sharing responsibilities for results measurement between the fund manager and the private sector grantees.

The paper also addresses the measurement of systemic change. While recognizing that the measurement of systemic change is still nascent among many challenge funds, it provides some examples of projects that are attempting to do so. For example, one project has measured systemic change using the following categories (Annex B, p. 33):

- Copying by other businesses: replication of the grantee business model that the challenge fund is supporting;
- Crowding-in: start-up of other types of businesses providing services to the grantees;
- Factor and other market systems: firms that are crowding-in, as described above, also serve others not associated with the project;
- Copying successful practice: behavior change at the end beneficiary level (e.g., smallholders);
- Business regulatory environment: supported companies join together to lobby for policy changes; and
- Innovation: grantee innovates and improves upon the ideas that the challenge fund originally financed.

**Kessler, Adam and Nabanita Sen. 2013. “Guidelines to the DCED Standard for Results Measurement: Capturing Wider Changes in the System or Market.” Cambridge: Donor Committee for Enterprise Development. (5 pages)**

This brief is for practitioners who are facilitating market systems and seeking to implement the DCED Standard for Results Measurement, an international standard synthesizing good practices in monitoring. It provides practical guidance on how to understand and implement the fifth component of the DCED Standard, “Capturing Wider Changes in the System or Market”. The guide identifies five aspects of systemic change:

1. Crowding in
2. Copying
3. Sector growth
4. Backward and forward linkages
5. Other indirect impact

The first two relate to imitation by other market actors, while the last three are focused on larger impacts created by a project's activities. The brief focused primarily on crowding in and copying, since practitioners have monitored these two aspects of systemic change more often. The guide advises on how to incorporate systemic change into results chains for each intervention and develop relevant indicators.

**Osorio-Cortes, Lucho and Marcus Jenal. 2013. "Monitoring and Measuring Change in Market Systems: Rethinking the Current Paradigm." MaFI Synthesis Report. Arlington VA: The SEEP Network. (18 pages)**

In 2012, SEEP's Market Facilitation Initiative (MaFI) coordinated a series of events to explore the application of complexity thinking to M&E in market systems facilitation programming. This paper synthesizes that work, beginning with an outline of the weaknesses of the "current M&E paradigm" (pp. 2-6):

1. "Excessive focus on 'our' direct effects on the poor". Attribution becomes more difficult when working through other market actors and seeking to influence their behaviors;
2. "Excessive focus on extraction of information for accountability to the donors", which can place inappropriate focus on outputs; and
3. "Sustainability understood as longevity of our legacy", which posits that projects are too focused on the sustainability of what they are providing, rather than on the capacity of those that are reached to develop and refine their own solutions to changing contexts in future.

The paper also posits seven principles for building a "systemic M&E framework" (pp. 7-14):

1. **Indirectness of impact:** a project focused on systemic change will primarily reach indirect beneficiaries that do not directly interact with the project, while the businesses that work directly with the project are termed "collaborators";
2. **Depth of impact:** stock and flow indicators (e.g., beneficiary incomes, # of new jobs created) are argued to be superficial indicators of change that can be achieved relatively easily by outside development intervention. Focusing on more fundamental elements of systemic change, such as leverage points, is far more helpful;
3. **Network-driven change:** the engagement of multiple actors in creating change makes it difficult to attribute impacts to a single cause;
4. **Unpredictability:** the fast-changing environment in which development programming occurs makes "flexibility, rapid learning systems and effective collaboration between facilitators, NGOs and donors" all critical;
5. **Sensitivity to external signals:** recognize that the system will change given the presence of a development program;
6. **Information deficit:** our inability to fully understand a system requires participation, learning, and flexibility; and
7. **Sustainability as adaptability:** true sustainability is argued to be the adaptability of the system.

**Osorio-Cortes, Luis, Marcus Jenal and Margie Brand. 2013. "Monitoring and Measuring Change in Market Systems: The Systemic M&E Principles in the Context of the Kenya Market Assistance Program." MaFI Case Study. Arlington VA: The SEEP Network. (40 pages)**

The SEEP Network’s Market Facilitation Initiative (MaFI) produced this paper under its Systemic M&E Initiative, which explores the application of complexity thinking to M&E in market systems facilitation programming. This publication builds on an earlier MaFI paper, “Monitoring and measuring change in market systems –rethinking the current paradigm–”, by applying the seven principles of systemic M&E to the Market Assistance Program (MAP) in Kenya. MAP focuses on facilitating change in multiple market systems.

The paper concludes that the principles are highly applicable and relevant to the Kenya MAP program, though some need to be better articulated. It also provides a framework and list of indicators used by MAP to measure systemic change:

- Behavior change
- Trust
- Loyalty
- Consumer awareness
- Business management patterns
- Participation in policy change and advocacy
- Relationships between actors
- Repeat sales
- Perceptions and preconceptions
- Knowledge

**Parsons, Beverly and Meg Hargreaves. 2009. “Evaluating Complex Systems Interventions.” Professional Development Workshop at the American Evaluation Association Annual Conference, November 9-15, Orlando FL. (93 slides)**

This presentation provides an overview of systems thinking. It focuses on three types of system dynamics, providing examples of each:

- Random (unorganized)
- Organized (simple or complicated)
- Adaptive

It emphasizes that systems have different dynamics and often multiple types simultaneously in a complex system. This affects the type of evaluation that should be conducted and the questions to be asked. These dynamics include (slides 40-41):

- Self-organizing/adaptive/organic
- Sensitivity to initial conditions
- Emergence
- Macro pattern
- Feedback
- Co-evolution
- Pattern formation and points of influence

These dynamics have various implications for evaluation, including (slides 42-43):

- Small differences can create large effects.

- The past influences but does not predict the future.
- Many points of influence exist.
- Boundaries, differences, and relationships are levers of influence toward a purpose.
- Simple rules underlie patterns.
- Pattern-based feedback and actions are iterative.
- Tensions are not resolved.
- Patterns are outcomes.

The presentation draws from curriculum development, communities of learning and family strengthening examples to demonstrate how evaluations were shaped by system dynamics.

**Patton, Michael Quinn. 2011. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford Press. (375 pp)**

Developmental evaluation is a type of utilization-focused evaluation that “guides action and adaptation in innovative initiatives facing high uncertainty” (p.36). Utilization-focused evaluation focuses on developing innovations and “achieving intended use by intended users” (p.14). This purpose contrasts with the proving/impact role of summative evaluation and the improving/fine-tuning role of formative evaluation. Early chapters describe the distinguishing features of developmental evaluation and contrast it with these other types of evaluation.

Developmental evaluation is appropriate for five purposes (Exhibit 10.1, pp.308-313):

1. Ongoing development to adapt a project, program, policy or other initiative to new conditions;
2. Adapting effective general principles to a new context;
3. Developing a rapid response in the face of a sudden major change or a crisis;
4. Preformative development of a potentially scalable innovation; and
5. Major systems change and cross-scale developmental evaluation.

Examples from many different types of systems are used to illustrate each of these purposes. Special attention is given to the relationships between development evaluation, systems thinking and complexity concepts. The role of the evaluator is to co-create social innovation by framing questions and collecting data that permit the timely recognition of patterns. While there are principles and values that can guide evaluation practice, no single inquiry framework or evaluation technique can be considered to be a “gold standard.”

**Reynolds, Martin, Kim Forss, Richard Hummelbrunner, Mita Marra and Burt Perrin. 2012. “Complexity, Systems Thinking and Evaluation - An Emerging Relationship?” *Evaluation Connections Newsletter of the European Evaluation Society (December): 7–9. (8 pages)***

This paper summarizes a panel discussion at the European Evaluation Society’s Helsinki Conference in 2012. It distinguishes between the concepts of systems thinking and complexity, and discusses both in relation to evaluation. It outlines where systems thinking and complexity converge, where they diverge, and then discusses possible ways forward.

Points of convergence between systems thinking and complexity include that they both challenge reductionist thinking and the focus on experimental evaluation approaches. Instead, both support multiple evaluation methods with an emphasis on process, co-evolution and incorporating (sometimes conflicting) perspectives. Both apply the concepts of boundaries, relationships and perspectives to understand systems and incorporate concepts of emergence and systems change.

On the other hand, systems thinking includes a wide range of evaluation concepts and tools (“craft skills”), with complexity science representing only one part of this broader range. Contemporary soft and critical systems thinking tends to consider systems to be conceptual devices rather than actual entities. From this viewpoint, complexity is more a description of the observer’s perspective than the system being observed. By focusing on choices about where boundaries are set, systems thinking “supports an explicitly ethical and political engagement with evaluation” (p.6).

The evaluation experts on the panel suggested that, in moving forward, systems thinking can contribute to advancing evaluation practice by promoting methodological heterodoxy and shifting mainstream thinking more toward “contribution” than “attribution”. Because systems approaches are flexible, they can be used in conjunction with other evaluation traditions. For example, complexity thinking could incorporate more emphasis on boundaries, relationships and perspectives. Systems approaches also can be used to build constructive collaboration among evaluation stakeholders. In order for these advances to occur, however, the methodological debate must extend beyond the evaluation community.

**Ruffer, Tim and Elise Wach. 2013. “Review of Making Markets Work for the Poor (M4P) Evaluation Methods and Approaches.” DFID Working Paper 41. London: Department for International Development. (57 pages)**

This DFID-commissioned resource reviews the methods used by 14 evaluations of M4P/market system facilitation programs funded by donors including DFID, USAID, SDC and GIZ. Its purpose is to “help guide the design and implementation of future evaluations” (p.ii). All of the evaluated projects were focused on improving market systems, particularly in agriculture. Defining systemic change as “transformations in the structure or dynamics of a system that leads to impacts on the material conditions or behaviors of large numbers of people” (p.4), it found that only five of the evaluations evaluated systemic change in a satisfactory way. The paper identifies several challenges with measuring systemic change (p.23):

1. Factors external to a project have an increased impact on changes at the higher levels of results chains;
2. It is more difficult to distinguish between treatment and non-treatment groups when projects take a facilitation approach; and
3. It is challenging to assess the market players that are truly creating change.

The document proposes a comprehensive set of recommendations for evaluating market systems facilitation programs. The authors propose that programs’ theories of change should explicitly describe how they expect to achieve systemic change, incorporating indicators of “replication, crowding in, and wider market change” (p.24). They argue that systemic changes should be initially identified through qualitative methods, including discussion with project partners and market players. Once identified, however, quantitative methods can be assigned to assessing the impacts of the systemic changes in additional outreach or income generated.

**Sarriot, Erin, Jim Ricca, Jennifer Yourkavitch, Leo Ryan, and the Sustained Health Outcomes (SHOUT) Group. 2008. “Taking the Long View: A Practical Guide to Sustainability Planning and Measurement in Community-Oriented Health Programming.” Calverton MD: Macro International Inc. (79 pages)**

This paper describes a detailed procedure for integrating sustainability concepts into project planning and for evaluating progress toward sustainability during project implementation. The sustainability framework, tools and methods were developed to improve and evaluate the sustainability of gains made by health projects in developing countries. While the focus is on health systems, the methods are general enough to apply to other types of systems. The approach is grounded in two key assumptions:

- “Sustainability planning is most effective when approached from a ‘system perspective’”
- “Sustainability is a dynamic process”

Based on a review of the conditions that make gains in community health programs sustainable, the authors provide a “sustainability framework” consisting of six components. Each component represents an important category (domain) of conditions that support sustained gains in health outcomes (e.g., “district health office capacity” and “enabling environment” are two of the components). Each component can be represented by many different indicators. By selecting a subset of indicators under each component, the framework can be adapted to fit the local context.

Six steps for applying the sustainability framework within the context of project planning and project implementation are presented, along with suggested management and measurement tools (p.20):

1. Define the local system
2. Facilitate formulation of a long-term vision
3. Develop a plan to achieve the vision
4. Collect data on indicators selected from the sustainability framework
5. Analyze the data and present information on the results
6. Revise practice based on the results

The authors suggest using the following three tests to establish the boundaries of the local system:

1. Stakeholders can feasibly be brought together
2. Assessments can be conducted (i.e., data can be collected on the units of analysis)
3. Decisions can be made following the sustainability assessment (therefore the national government is typically not included)

They caution that the boundaries of the local system can evolve with time.

**Springfield Centre. 2014. “Good Practices in Facilitation: The System Change Framework.” Durham UK: The Springfield Centre for Business in Development. (5 pages)**

The Springfield Centre provided this resource to attendees of its training on Making Markets Work for the Poor (M4P). Currently in draft form, the resource outlines four elements of systemic change:

1. **Adopt:** A market player adopts a behavior or practice change that creates an ultimate benefit for the poor. “Adopt” is an early indication of systemic change, when a market player has demonstrated ownership of a new method and continues to follow it beyond a project’s pilot phase.
2. **Adapt:** A market player improves upon or scales-up their use of the new behavior or practice change.
3. **Expand:** Other players that are not directly targeted by the program adopt comparable behavior or practice changes as a result of the demonstration of the initial adopter(s) or competitive pressures.
4. **Respond:** The existence of the new behavior or practice elicits changes from other, dissimilar market players. This can be in terms of changing their own roles or improving their own offers.

While “adopt” is always the first element to occur, the other three will not always occur sequentially. In some cases they may happen simultaneously. The elements may occur as a result of a project’s post-pilot facilitation or independently, depending on the nature of the market system and market players. Sample indicators are presented for each element.

**Stern, Elliot, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies and Barbara Befani. 2012. “Broadening the Range of Designs and Methods for Impact Evaluations.” DFID Working Paper 38. London: Department for International Development. (24 pages)**

This DFID-commissioned study examines fundamental issues in impact evaluation (IE) for the purpose of assessing the potential for using non-experimental evaluation designs for development programs. They argue for a wider range of evaluation designs and more attention to contribution analysis (as opposed to attribution analysis) in order “to deal with contemporary interventions that are often complex, multi-dimensional, indirectly delivered, multi-partnered, long-term and sustainable” (p. 5). Chapter 4 describes the four key types of evaluation questions and lists suitable evaluation designs for each question (see Table 4.2, p. 48):

1. To what extent can a specific (net) impact be attributed to the intervention?
2. Has the intervention made a difference?
3. How has the intervention made a difference?
4. Can this be expected to work elsewhere?

Complexity issues are addressed in Chapter 5, based on a review of complexity literature and a set of DFID programs. The following complexity-related program attributes are identified, along with their evaluation challenges and approaches for addressing the challenges (see Table 5.5, pp. 60-61):

- Overlap of multiple interventions with similar aims
- Multiple and diverse activities and projects within a single program
- Locally customized non-standard programs, often in diverse contexts
- Program impacts are likely to occur over the long term
- Working in areas of limited understanding/experience
- Working in areas of high risk or uncertainty
- Stated impacts are difficult to measure, possibly intangible and/or reflect composite goals
- Programs working “indirectly” through “agents” and often at multiple levels and stages

Design questions related to selecting the unit of analysis and the sequencing of evaluation for long-term programs emerged at several points in the review. Finally, the authors propose a framework for quality assurance

in IE consisting of three types of standards: 1) process, 2) technical, and 3) normative. The process and technical standards address threats to the validity of the IE.

**TANGO International. 2013. “Summary of the Expert Consultation on Measuring Resilience.” Summary of expert consultation held in Rome, February 19-21. Tucson AZ: Tango International. (23 slides)**

This presentation summarizes an expert consultation on the topic of measuring resilience in food security and nutrition programs. The consultation defined resilience as “[t]he ability of countries, communities and households to anticipate, adapt to and /or recover from the effects of potentially hazardous occurrences (natural disasters, economic instability, conflict) in a manner that protects livelihoods, accelerates and sustains recovery, and supports economic and social development” (slide 5). Held in February 2013, the consultation was attended by donors, NGOs, foundations, universities and research institutes.

The paper outlines considerations for monitoring resiliency:

- Recognize the importance of context. Measure the resilience of a specific individual or group to a specific shock or stress. Recognize that the context changes with time.
- Use panel data from the same households over time.
- Understand structural and transitory thresholds and tipping points.
- Consider technical capacity. Resilience is complex and thus often so are the necessary measures. Align the selected methods with available technical capacity.
- Use locally and culturally relevant metrics.
- Consider aspects beyond the individual. These can include “formal/informal governance and institutional processes and systems enhance/limit individual and household resilience; policies, knowledge/information management, laws, programming” (slide 8).
- Consider interrelationships between individual, household, community, region and how they influence each other.
- Consider aspirations and motivations, which influence risk-taking capacity and behaviors at individual, household and community levels.
- Consider natural resource and ecosystem health, as they greatly affect household livelihoods.

The presentation outlines current efforts and approaches to measure resilience, as well as areas where little is currently being done. It presents an analytical framework for measuring resilience with indicator areas defined for the following four stages:

1. Baseline well-being and basic conditions measures
2. Disturbance measures (shocks/stresses)
3. Resilience response measures
4. End-line well-being and basic conditions measures

The presentation closes by listing the next steps to be facilitated through the Food Security Information Network over the short-term, medium-term and long-term. These include identifying the contributors to resilience, the contexts in which they are applicable, and for which shocks, and ultimately developing standard resilience indicators that are valid and reliable.



**U.S. Agency for International Development**

1300 Pennsylvania Avenue, NW

Washington, DC 20523

Tel: (202) 712-0000

Fax: (202) 216-3524

**[www.usaid.gov](http://www.usaid.gov)**