

MEASURING MARKET DEVELOPMENT

A HANDBOOK FOR FUNDERS AND IMPLEMENTERS OF FINANCIAL INCLUSION PROGRAMS

PATRICK SPAVEN AND KARINA BROENS NIELSEN



This work is available under the Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO) <https://creativecommons.org/licenses/by/3.0/igo/>. By using the content of this publication, you agree to be bound by the terms of this license. For attribution, translations, adaptations, and permissions, see the provisions and terms of use at <https://www.adb.org/terms-use#openaccess>.

Suggested citation: Spaven, Patrick, and Karina Broens Nielsen. 2017. "Measuring Market Development: A Handbook for Funders and Implementers of Financial Inclusion Programs." Washington, D.C.: CGAP.

All queries on rights and licenses should be addressed to CGAP Publications, 1818 H Street, NW, MSN IS7-700, Washington, DC 20433 USA; e-mail: cgap@worldbank.org.

Consultative Group to Assist the Poor/World Bank Group



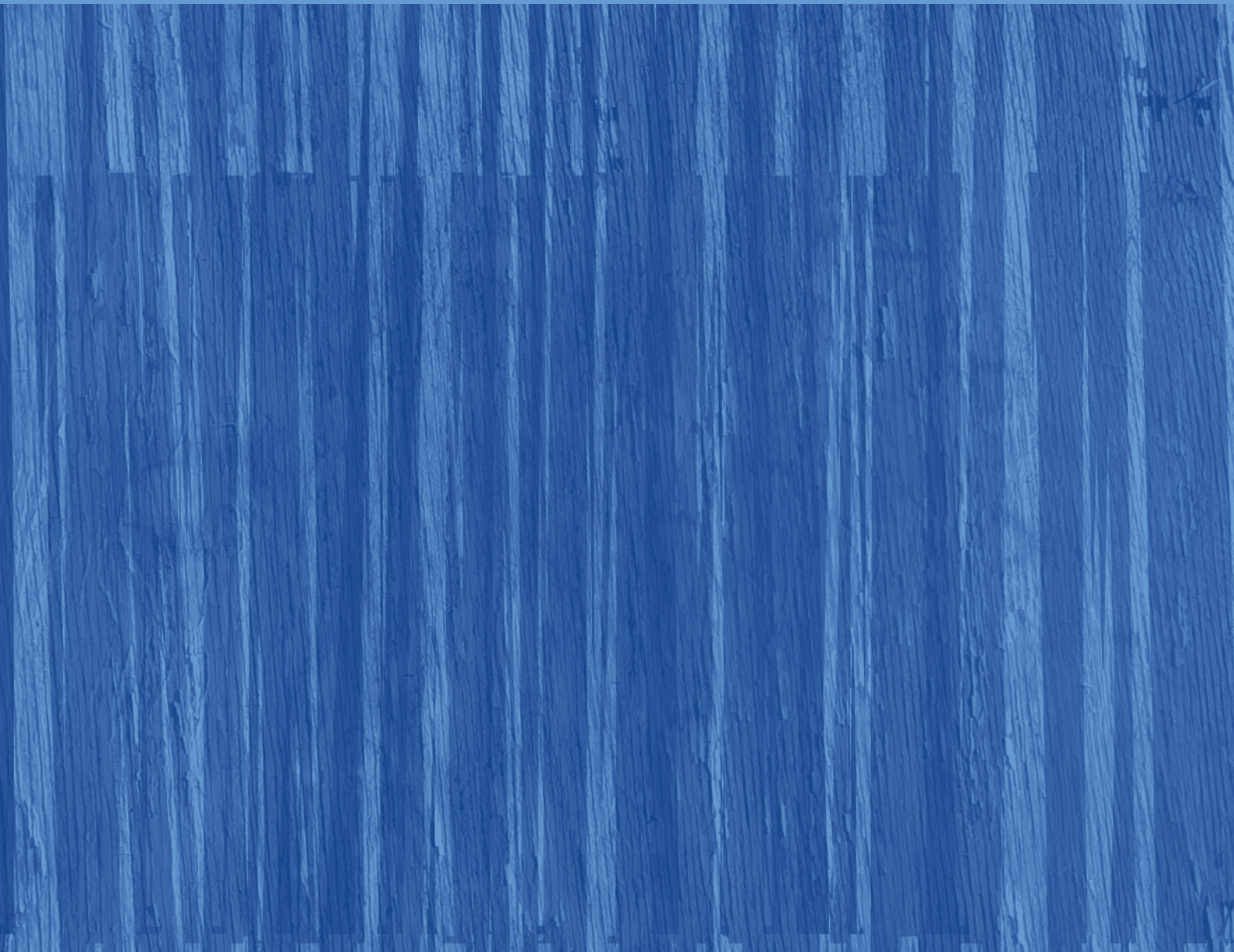
MEASURING MARKET DEVELOPMENT

A HANDBOOK FOR FUNDERS AND
IMPLEMENTERS OF FINANCIAL
INCLUSION PROGRAMS

PATRICK SPAVEN AND KARINA BROENS NIELSEN

TABLE OF CONTENTS

Executive Summary	v
What You'll Find in This Handbook	viii
Introduction	1
MODULE 1. The Systemic Approach and Measurement	3
MODULE 2. The Measurement System	9
MODULE 3. Theory of Change and Results Chains	17
MODULE 4. Measurement Questions	29
MODULE 5. Results Frameworks	37
MODULE 6. Indicators and Data Sources	49
MODULE 7. Enhanced Monitoring	63
MODULE 8. Reporting and Knowledge Sharing	69
MODULE 9. Evaluation	77
MODULE 10. Portfolio-Based Measurement	85
Glossary	93
Self-Assessment Tool for Financial Inclusion Measurement Systems	95



EXECUTIVE SUMMARY

This Handbook guides funders and their implementing partners on how to effectively monitor and measure results of financial inclusion programs that apply a systemic approach. It is primarily intended for measurement specialists who design and manage measurement systems of financial inclusion programs, oversee a portfolio of programs, or advise programs on how to measure results. The Handbook includes a self-assessment tool—a practical instrument for benchmarking and identifying gaps and areas for improvement.

FINANCIAL INCLUSION AND THE SYSTEMIC APPROACH

Governments and development partners around the world share a vision of financial inclusion where all individuals and businesses have access to, and the ability to use, a range of appropriate financial services that are provided responsibly and sustainably. Despite significant funding of financial inclusion efforts, approximately 2 billion adults remain excluded from the formal financial system in 2016 (Global Findex), and there is still substantial variation in the diversity, quality, and use of financial services.

For the financial services market to be efficient and inclusive, different market actors need to perform a variety of functions:

- Core functions of demand and supply.
- Supporting functions that shape, inform, and enable transactions between customers and providers.

- Rules and norms governing these functions.

For the financial services market to be efficient and inclusive, different market actors need to perform a variety of functions: the core functions of demand and supply and the supporting functions that shape, inform, and enable transactions between customers and providers. There must be rules and norms governing such functions. Funders need to understand the different functions in the financial services market system and address the barriers that exclude the poor. They can do this through interventions that facilitate inclusive and responsible financial services markets.

CGAP's "A Market Systems Approach to Financial Inclusion: Guidelines for Funders" (Burjorjee and Scola 2015) encourages funders to apply a systemic approach. Instead of providing piecemeal solutions through direct assistance, funders should support programs that

use a systemic approach to help market actors strengthen weak functions and develop other functions that are needed. The interventions should be designed to change the mindsets and practices of financial services providers (FSPs), clients, regulators, supervisors, and other market actors.

Two other aspects of the systemic approach stand out. First, because the financial system is complex and elements within it are dynamic and unpredictable, programs need to emphasize experimentation and learning, and be regularly adapted based on experience. Second, progress to sustainable financial inclusion at scale can be slow and not necessarily linear, so program managers need to take a long-term view toward systemic change.

IMPLICATIONS FOR MEASUREMENT

The principal implications for measurement in systemic financial inclusion programs are as follows:

1. The measurement system needs to be designed to detect and assess whether and how systemic change is happening. The system needs to include theories of change (ToCs), results chains, and results frameworks and indicators. Some outcomes and indicators will measure progress toward systemic change. Some will relate to program partners, while others go beyond partners to other actors that are affected. ToCs should depict how interventions and intermediate outcomes contribute to systemic change and how systemic change leads to a more inclusive financial system and, ultimately, to development outcomes like improved economic well-being.
2. Expected systemic change outcomes often relate to changes in capacity, practices, institutional processes and structures, and relationships between actors. These types of outcomes are often highly context-specific and nuanced. For example, indicators will need to measure degrees of change, or the quality of new processes or relationships. It may be possible to summarize these outcomes in numbers through rating scales, but it is difficult to standardize this across projects and programs. Ratings scales need to be complemented by contextualized narrative. Results frameworks that focus exclusively on quantitative targets can lead program staff to orient themselves “to chase the numbers” within target timeframes rather than to explore long-term, often less
3. Programs need to have a broad view in monitoring a complex financial system. Measurement tends to focus narrowly on interventions and intended outcomes that are defined at the beginning of a program. In systemic financial inclusion programs, it is important to broaden the scope of measurement beyond the intended results to include unexpected or unplanned outcomes and factors in the external environment that can affect progress both positively and negatively. Monitoring based exclusively on results framework indicators is inadequate for systemic financial inclusion programs and can be misleading.
4. Measurement tends to assume that predicted change is influenced by or attributable to program interventions. Conventional intervention-led measurement—the “outward” perspective—asks whether and to what extent the interventions have led to intended results. Given the complexity of financial market systems, intervention-led measurement needs to be done in conjunction with change-led measurement—the “inward” perspective. Change-led measurement asks what has changed and what has contributed to the changes. Applying these two perspectives gives a broader view of evidence of change, allows for triangulation, and mitigates dependency on just one perspective. Although it is difficult to attribute specific outcomes to funders’ interventions, this dual perspective allows funders to develop a rigorous and credible narrative about the impact of their programs and projects.
5. ToCs and results frameworks should be viewed as hypotheses and not as blueprints—results are difficult to predict. ToCs and results frameworks need to be regularly challenged, and funders and program units must be prepared and enabled to change them. In this paradigm of accountability, the work of program units is assessed against a range of factors, and they are not held to targets that may no longer be reasonable or appropriate.
6. Monitoring in systemic programs should not only ensure interventions stay on track but should also provide frequent feedback so that interventions can be adjusted as needed. Therefore, monitoring needs to be linked to regular, data-informed, evaluative reviews. These reviews are critical in ensur-

easily quantifiable, but more sustainable outcomes.

ing that the data emerging from monitoring are applied to decisions about the interventions going forward. The reviews should be owned internally, but can benefit from the participation of partners and from third-party facilitation.

7. This type of measurement needs to be closely integrated with project and program management. Because more participants are involved than in conventional measurement, support capacity is needed, for example, to understand the nature and value of evidence. Also, staff need to have incentives to be open-minded and inquisitive, and to take risks. When senior program unit management and funders show that they value measurement results, staff are more likely to see the value as well.
8. Funders need to recognize that ToCs in systemic financial inclusion programs have a longer time horizon than those in many other fields. Funders should not expect quick results at the higher levels of the ToC.¹ Instead, they should give due attention to the intervening steps. This is particularly important in setting milestones and targets that often frame the incentives for program teams and partners.
9. To understand how change happens, we need to measure impact. Although the type of evaluative monitoring described in this Handbook can give us valuable insights into program impact, there is usually a need at some point to apply externally led impact evaluation, especially for large-scale programs. It is important to focus these evaluations where significant sustained impact is likely to be detected. If an evaluation takes place during, or shortly after, the life of the program, this type of impact is likely to be found below the level of development outcomes—for example, at the level of the inclusive financial system. The impact on development outcomes needs to be evaluated strategically and over the long term, preferably in collaboration with other funders and market actors.

A BROADER MODEL OF ACCOUNTABILITY

Conventional accountability in development programs tends to center on achieving pre-defined targets. Because accountability usually passes through a chain of actors—from project staff, through to programs management, to funders, and often on to funders' governance stakeholders, it makes sense to look for standardized indicators so that key results information can be aggregated and indicators can be framed as readily measurable targets. In systemic financial inclusion programs, the scope for quantitative indicators, particularly those that can be shared with other projects and programs, is limited. And results are also less predictable because programs, teams, and partners need to be able to experiment and adapt. Despite these challenges, implementation of financial inclusion programs still need to be accountable. This calls for a broader model of accountability that measures a range of demonstrated behaviors that indicate, for example, enhanced forms of monitoring, effective risk management, and evidence-based adaptation. Funders and senior program management should create an environment that allows these behaviors. Reports to stakeholders need to reflect this broad view of accountability. Although the reporting may be more complicated than a conventional accountability model, it also has the added value of providing greater scope for learning.

Executive Summary References and Resources

- Burjorjee, Deena M., and Barbara Scola. 2015. "A Market Systems Approach to Financial Inclusion: Guidelines for Funders." Washington, D.C.: CGAP. https://www.cgap.org/sites/default/files/Consensus-Guidelines-A-Market-Systems-Approach-to%20Financial-Inclusion-Sept-2015_0.pdf
- World Bank. 2015. "The Global Findex Database 2014, Measuring Financial Inclusion around the World." Washington, D.C.: World Bank. <http://www.worldbank.org/en/programs/globalindex>

Note

1. The levels of the ToC are set out in Module 3.

WHAT YOU'LL FIND IN THIS HANDBOOK

MODULE 1 (The Systemic Approach and Implications for Measurement) introduces the systemic approach to financial inclusion programming and explains its rationale. It sets out the key principles of the approach and their implications for measurement. The module describes a broader concept of stakeholder accountability that is appropriate in the context of financial inclusion.

MODULE 2 (The Measurement System) describes the scope and components of a measurement system for programs in financial inclusion. It discusses the importance of a conducive environment for measurement and a broader conceptualization of accountability. It explains how the measurement system is mobilized through a program measurement strategy and project measurement implementation plans.

MODULE 3 (Theories of Change and Results Chains) examines the roles of the theory of change (ToC) and results chains in financial inclusion programs and projects and explains the distinction between the two terms. It identifies criteria for effective ToCs and results chains and offers a conceptual ToC module for financial inclusion programs that apply a systemic approach. It also provides examples of a program ToC and a project results chain.

Module 4 (Measurement Questions) explains how measurement questions help to sharpen the relevance of both monitoring and evaluation throughout the program and project cycle and why it is important to frame the questions at the beginning of these cycles. It then provides guidance on identifying key measurement questions.

Module 5 (Results Frameworks) describes results frameworks, how they link to ToCs and results chains, and how they are used in development programs generally. The module assesses the strengths and weaknesses of

ToCs in the context of systemic financial inclusion programs and puts forward proposals for using them effectively. It explains the importance of managing assumptions and risks in the financial inclusion domain, and presents a model risk register that complements a results framework. It presents an example of a completed program results framework and an indicator profile.

Module 6 (Indicators and Data Sources) examines the roles and limitations of indicators and targets in systemic financial inclusion programs. It proposes several indicator focus areas at different levels of the financial inclusion ToC and identifies the types of data sources that might be used. It analyzes the mix of data sourcing opportunities and the challenges that market systems for financial inclusion interventions face and how to address them.

Module 7 (Enhanced Monitoring) examines why traditional monitoring modalities often fall short of what is required for systemic financial inclusion interventions. It describes the concept of “enhanced monitoring” and points to some of its practical applications.

Module 8 (Reporting and Knowledge Sharing) describes the principal roles of reporting in systemic financial inclusion programs and how these roles can fit into a single reporting mechanism. It discusses the inter-related issues of frequency and scope or content of reports and points to their limitations for knowledge sharing with a wider range of stakeholders. It provides an example of a program annual report template from program staff to the funder.

Module 9 (External Evaluation) examines the uses of external evaluation in systemic financial inclusion programs. It addresses the concept of impact evaluation in the context of systemic financial inclusion programs and assesses different evaluation methodologies.

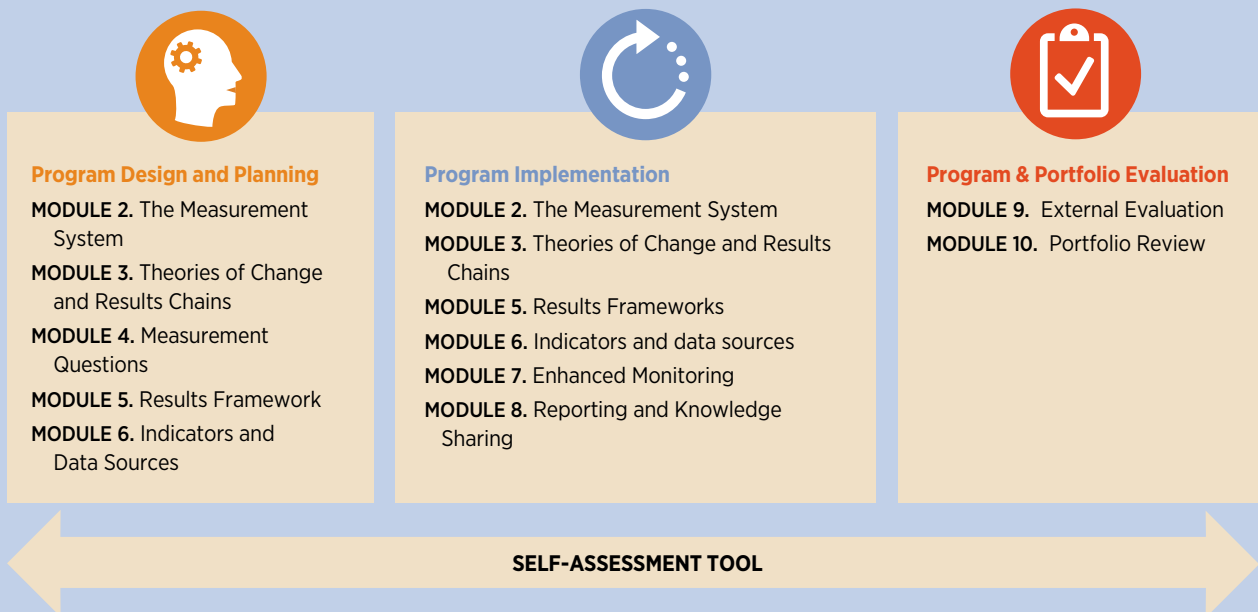
Module 10 (Portfolio-Based Measurement) explains how portfolio-based measurement, using standardized indicators and/or scoring systems, can produce insights for funders to improve their interventions and to make decisions about their overall strategy. It presents two portfolio measurement tools: portfolio dashboards (usually applied more than once a year) and portfolio reviews, which take place

at less frequent intervals. This module is particularly relevant to funders.

Annexes

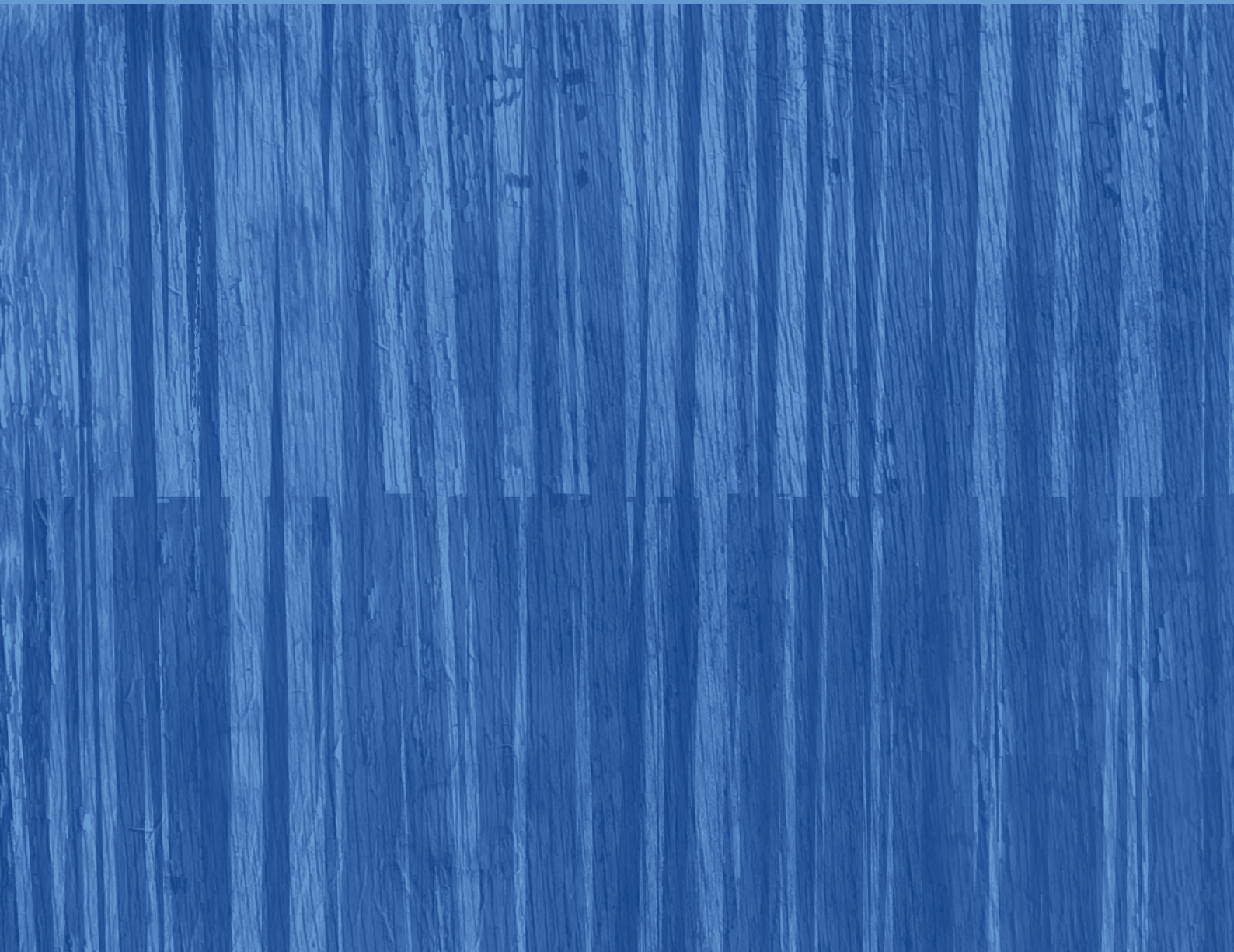
- A. Glossary
- B. Self-Assessment Tool—a practical instrument for benchmarking and identifying gaps and areas for improvement

At what stage of programming should one start applying the guidance from the handbook?



The Handbook Is for . . .

- Measurement specialists in funder organizations that have investments in the financial sector.
- Organizations—funders or program units—that want to produce their own financial inclusion measurement guidance. The Handbook is not designed to be an operational manual, but it can provide inspiration and material for such tools.
- Program and project units that construct measurement systems, strategies, and plans.



INTRODUCTION

The Market Systems Approach to financial inclusion is designed to help market actors strengthen weak functions and develop other functions that are needed.² This Handbook guides funders and their implementing partners on how to monitor and measure results of financial inclusion programs that follow a systemic approach. It was developed through collaboration with many organizations that work in financial inclusion, including CGAP members and implementing organizations such as UNCDF, USAID, IFAD, the Mastercard Foundation, FSD Africa, the Donor Committee for Enterprise Development (DCED), and the BEAM Exchange. It draws on guidance on measurement for market development, and applies it to financial inclusion. Some guidance presented here depart from mainstream measurement practices because this Handbook aims to accommodate the challenges in applying the systemic approach to a dynamic financial services market.

This Handbook is intended for measurement specialists who design and manage measurement systems of financial inclusion programs, oversee a portfolio of programs, or advise program leads on how to measure results. The intended audience includes funder staff or staff of units that implement financial inclusion programs (program units). These units could be operated by funders, but they are more typically found in organizations such as Financial Sector Deepening Trusts (FSDs), intergovern-

mental organizations, government agencies, nongovernmental organizations (NGOs), and consultancy companies. An important objective of this Handbook is to help both funders and program unit staff to align their measurement systems and agree on roles and responsibilities. It also stresses the responsibilities of both parties for effective measurement.

This Handbook is particularly useful in developing new, or conducting a comprehensive review of existing, measurement strategies and systems at the organizational and programmatic levels. It includes a Self-Assessment Tool (SAT)—a practical instrument for benchmarking and identifying gaps and areas for improvement.

This Handbook and the SAT are not alternatives to the DCED Standard for Results Mea-

FIGURE I-1. Comparing the DCED Standard and This Handbook

DCED Standard and Audit	Handbook Guidance and SAT
Describes the desired state of what a market systems-oriented measurement framework should be	Describes the journey from mainstream approaches to measurement, toward a market systems-oriented measurement framework
Focuses on the broader private sector	Focuses on financial sector
Follows an eight-part framework	Flexible and can apply sections as appropriate
Provides an audit service	Provides a Self-Assessment Tool



KEY TAKEAWAYS

- This Handbook is written for measurement specialists who design and manage the measurement systems and strategies of financial inclusion programs, oversee a portfolio of programs, or advise program leaders on measurement. It aims to promote effective measurement, thereby helping to improve program design and performance, demonstrate program effectiveness and impact, and maximize learning.
- This Handbook synthesizes relevant guidance from other fields, particularly market development, and applies it to a financial inclusion context.
- Users will need to tailor this Handbook's guidance to their own circumstances.

surement (DCED 2016) and its audit services, which are widely used by market development programs in a variety of private-sector programs. Rather, this Handbook complements the DCED Standard and other relevant guidance, such as the BEAM Exchange (BEAM Exchange 2015 and O'Sullivan 2016). Its focus on financial inclusion is a practical resource for funders and program unit staff working in this field.

Introduction References and Resources

- BEAM Exchange. 2015. "Monitoring Guidance." London: BEAM Exchange. <https://beamexchange.org/guidance/monitoring-overview/>
- DCED (Donor Committee for Enterprise Development). 2016. "DCED Monitoring and Measuring Results in Private Sector Development." Cambridge, United Kingdom: DCED. <http://www.enterprise-development.org/page/measuring-and-reporting-results>
- O'Sullivan, Fionn. 2016. "Impact Evaluations for Market Systems Programmes." London: The BEAM Exchange. <https://beamexchange.org/guidance/evaluation-guidance/>
- Springfield Centre, The. 2015. "The Operational Guide for Making Markets Work for the Poor (M4P) Approach," 2nd edition. The Springfield Centre. <http://www.enterprise-development.org/wp-content/uploads/m4pguide2015.pdf>

Note

2. See Module 1 for more detail on the systemic approach.

THE SYSTEMIC APPROACH AND MEASUREMENT



THIS MODULE . . .

- Explains the rationale for applying a systemic approach to financial inclusion.
- Sets out the key principles of the systemic approach and their implications for measurement, including a broader concept of accountability.

CONSTRAINTS TO FINANCIAL INCLUSION

Despite significant funding of financial inclusion efforts—an estimated US\$34 billion was committed as of 2015—approximately 2 billion adults remain excluded from the formal financial system (Soursourian and Dashi 2016 and Findex 2016). And there is still substantial variation in the diversity, quality, and use of financial services.

To advance financial inclusion, constraints to providing relevant financial services and preventing poor and low-income people from accessing and using financial services need to be addressed. The dearth of providers that serve low-income segments is often a symptom of underlying constraints that may be rooted in different areas of the financial system.

An inclusive financial system requires that a variety of functions be effectively performed by different market actors (see Figure M1-1):

- Core functions of demand and supply (customers and providers).

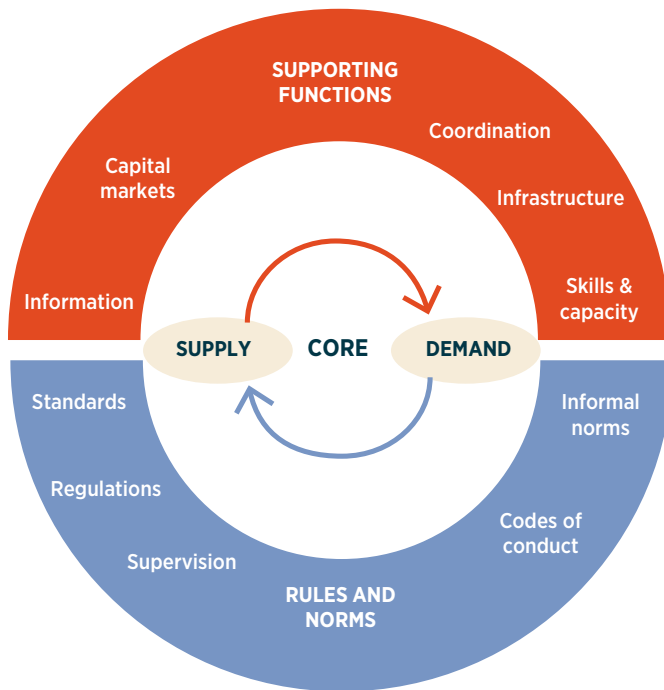
What is financial inclusion?

A state where both individuals and businesses have opportunities to access, and the **ability to use**, a range of **appropriate financial services that are responsibly and sustainably provided by formal financial institutions**.

- Supporting functions that shape, inform, and enable transactions between customers and providers.
- Rules and norms governing both core and supporting functions.

Constraints to financial inclusion often stem from knowledge gaps and other capacity weaknesses, misaligned incentives, or weak relationships and coordination among a range of market actors. Insufficient financial resources are often not the most significant problem.

FIGURE M1-1. The market system and main market functions



APPLYING A SYSTEMIC APPROACH

To promote financial inclusion, funders should first assess the different functions in the financial system and address constraints that prevent these functions from working effectively to include the poor. CGAP’s “A Market Systems Approach to Financial Inclusion: Guidelines for Funders” (Burjorjee and Scola 2015) encourages funders to use a **systemic approach** in their financial inclusion programs. Instead of being market actors that step in to fill a gap, they can apply interventions that use a systemic approach to strengthen weak functions and develop other required functions. The interventions are designed to change the mind sets and practices of financial services providers (FSPs), clients, regulators, supervisors, and other market actors, thereby altering the underlying system dynamics—a process called systemic change. Systemic change is expected to eventually lead to changes in system structure and in financial inclusion itself.

Key principles and their implications for measurement

Applying a systemic approach to financial inclusion has important implications for the way programs are designed and carried out and for how they are measured. Three key features of the financial system and its relationship with financial inclusion stand out:

What is systemic change in the context of inclusive financial markets?

A change in the underlying dynamics of a system (the financial system or an interconnected system) that leads to changes in the system’s structure, scale, sustainability, and resilience and improvements in the poor’s participation in financial services markets.

- The financial system encompasses a wide range of different actors that have interdependent roles and relationships.
- Financial markets are unpredictable and dynamic.
- Progress to sustainable financial inclusion at scale can be slow and not necessarily linear.

These features lead to the five key principles of the systemic approach and its measurement:

- Aim for systemic change
- Play a facilitative role
- Experiment, learn, and adapt
- Adopt a long-term view
- Take a different perspective on accountability

KEY PRINCIPLE #1: Aim for systemic change

The systemic approach is about sustainably changing the underlying dynamics and structure of a financial system to enable it to be more inclusive. The measurement system needs to be designed to detect and assess whether and how systemic change is happening. This calls for a distinct level in program Theories of change (ToCs) and results frameworks (and often in the project equivalents) that defines systemic change outcomes and their indicators in the specific program and project contexts. Some outcomes relate to program partners; others reach beyond program partners to a wider range of actors. ToCs should depict how interventions and intermediate outcomes contribute to systemic change and how systemic change leads to a more inclusive financial system, and ultimately to development outcomes like improved economic well-being.

A broad range of indicators is needed to measure systemic change. Expected outcomes often relate to changes in capacity, practices, institutional processes and structures, and relationships between financial system actors. Meaningful quantitative indicators are often

difficult to discern. It may be impossible to standardize indicators across projects and programs. In addition, results frameworks that focus exclusively on quantitative targets can lead program staff to orient themselves “to chase the numbers” within target timeframes rather than to explore long-term, often less easily quantifiable, but more sustainable, outcomes. Principle #1 is closely related to Principle #5.

KEY PRINCIPLE #2: Play a facilitative role

Funders and program unit staff need to think of themselves as facilitators who help to change system dynamics and enable a broad range of market actors to perform market functions more effectively, without having to provide missing functions themselves. Facilitators need to be able to step back to understand, monitor, and address the market system.

Market systems are made up of the interactions of multiple market actors that perform different functions, operate in their own interconnected systems, and are influenced by wider contextual factors beyond the funders’ control. Funders and program unit staff need to constantly monitor the overall financial system, relevant interconnected systems—such as services markets for capacity building or data—and contextual factors that affect these systems.

Measurement tends to incorporate the assumption that predicted change is influenced by or attributable to program interventions. In systemic financial inclusion programs, it is important to broaden the scope of measurement beyond the intended results to include unexpected or unplanned outcomes and factors in the external environment that can affect progress both positively and negatively.

Conventional intervention-led measurement—the “outward” perspective—asks whether and to what extent the interventions have led to intended results. Given the complexity of financial market systems, intervention-led measurement needs to be done in conjunction with change-led measurement—the “inward” perspective. Change-led measurement asks what has changed and what has contributed to the changes.

Applying these two perspectives allows for a broader view of evidence of change and triangulation. It also mitigates dependency on just one perspective. Although it is difficult to attribute specific outcomes to funders’ interventions, this dual perspective allows funders to develop a rigorous and credible narrative about the impact of their programs and projects. When evaluating and analyzing results, funders should take a broad view on how the

Programs and projects

In the Handbook, “program” is a coherent set of projects with a common focus. A program usually has a longer life than any single project within it. A project is a tightly focused set of interventions with a common work plan. ToCs apply to programs, while results chains apply to projects.

financial services market has evolved and focus on understanding how they contributed to this change.

KEY PRINCIPLE #3: Experiment, learn, and adapt

The systemic approach allows for uncertainty through its flexible design and implementation. Funders that use this approach are able to encourage experimentation and adaptation. The program design should allow for program staff to engage with a variety of market players, enter into and exit partnerships as needed, adapt strategies based on new information, and use funding opportunistically to spur innovation.

This approach has profound implications for measurement. Above all it implies that results are difficult to predict. ToCs, results chains, and results frameworks—the cornerstones of measurement—need to be viewed and managed differently from each other. Although these frameworks should be based on the best available evidence of how change happens in financial inclusion, inherent unpredictability dictates that they still should be viewed as hypotheses and not blueprints. ToCs and results frameworks need to be regularly challenged, and funders and program units must be prepared and enabled to change them as needed. This requires a new paradigm of accountability where program unit activities are assessed against a range of factors and are not held to static targets.

There are also implications for measurement processes and culture. Monitoring should not be only about ensuring interventions stay on track. Monitoring should provide frequent feedback so that interventions can be adapted when necessary. Therefore, monitoring needs to be linked to regular, data-informed, evaluative reviews. These reviews are critical in ensuring that the data emerging from monitoring are applied to decisions that shape the interventions going forward. The reviews should be owned internally and can be enriched by participation of partners and by outside facilitation.

To take root, these approaches to measurement need capacity building and incentives. Measurement of this type is integrated more closely with project and program management than conventional measurement. This means there are more participants, and many of them need capacity support—for example, to understand the nature and value of evidence. There also needs to be incentives for staff to be open-minded and inquisitive, and to take risks. Staff who see that senior unit management and funders value and use results will likely do the same.

KEY PRINCIPLE #4: Adopt a long-term view

The paths from interventions to sustainable financial inclusion at scale are long and unpredictable. Sustained impact of financial inclusion on people's lives—development outcomes—takes even longer to achieve.

Funders should not push for quick results at the highest levels of ToCs. They should focus on the intervening steps instead. How, when, and which milestones and targets are set can help to define incentives. Funders should expect an extended horizon for program interventions and recognize that, while some progress can be achieved in the lifetime of a program, in some cases, programs will be contributing to a long-term strategic vision in a country context.

To understand how change happens, we need to measure impact. Although the type of evaluative monitoring described in this Handbook can provide valuable insights into program impact, there is usually a need at some point to apply externally led impact evaluation especially regarding large-scale programs. It is important to focus these evaluations where significant sustained impact is likely to be detected. If an evaluation takes place during, or shortly after, the life of program, this type of impact is likely to be found below the level of development outcomes—for example, at the level of the inclusive financial system.³ The impact on development outcomes needs to be

evaluated strategically and over the long term, preferably in collaboration with other funders and market actors.

KEY PRINCIPLE #5: Take a different perspective on accountability

Conventional accountability in development programs tends to focus on achieving pre-defined targets. Accountability is a key feature of the relationship not just between program and funder, but also project and program on the one hand, and funder and governing stakeholders on the other. For this reason, there is a strong inclination to look for common indicators so that key performance information can be aggregated from the project level along the chain of accountability. Where common indicators are not possible or appropriate, there is still the expectation that quantitative indicators, which can be converted into more readily measurable targets, need to be used.

In systemic financial inclusion programs, there is limited scope for quantitative indicators, particularly for those that can be shared with other projects and programs. Results are also less predictable long in advance, because staff need time to experiment and adapt the program. Despite this, financial inclusion programs still need to incorporate accountability. A broader model of accountability may be needed. This broader model requires programs to be designed to demonstrate a range of behaviors, such as enhanced forms of monitoring, effective risk management, and evidence-based adaptation. It also obliges funders and senior program management to create a conducive environment for, and to engage with, these behaviors.

The scope and content of reporting to stakeholders must adequately cover the broad aspects of accountability. This approach has a more complex responsibility on reporting than that for a conventional accountability model, but it also creates an opportunity to combine accountability with learning.



KEY TAKEAWAYS

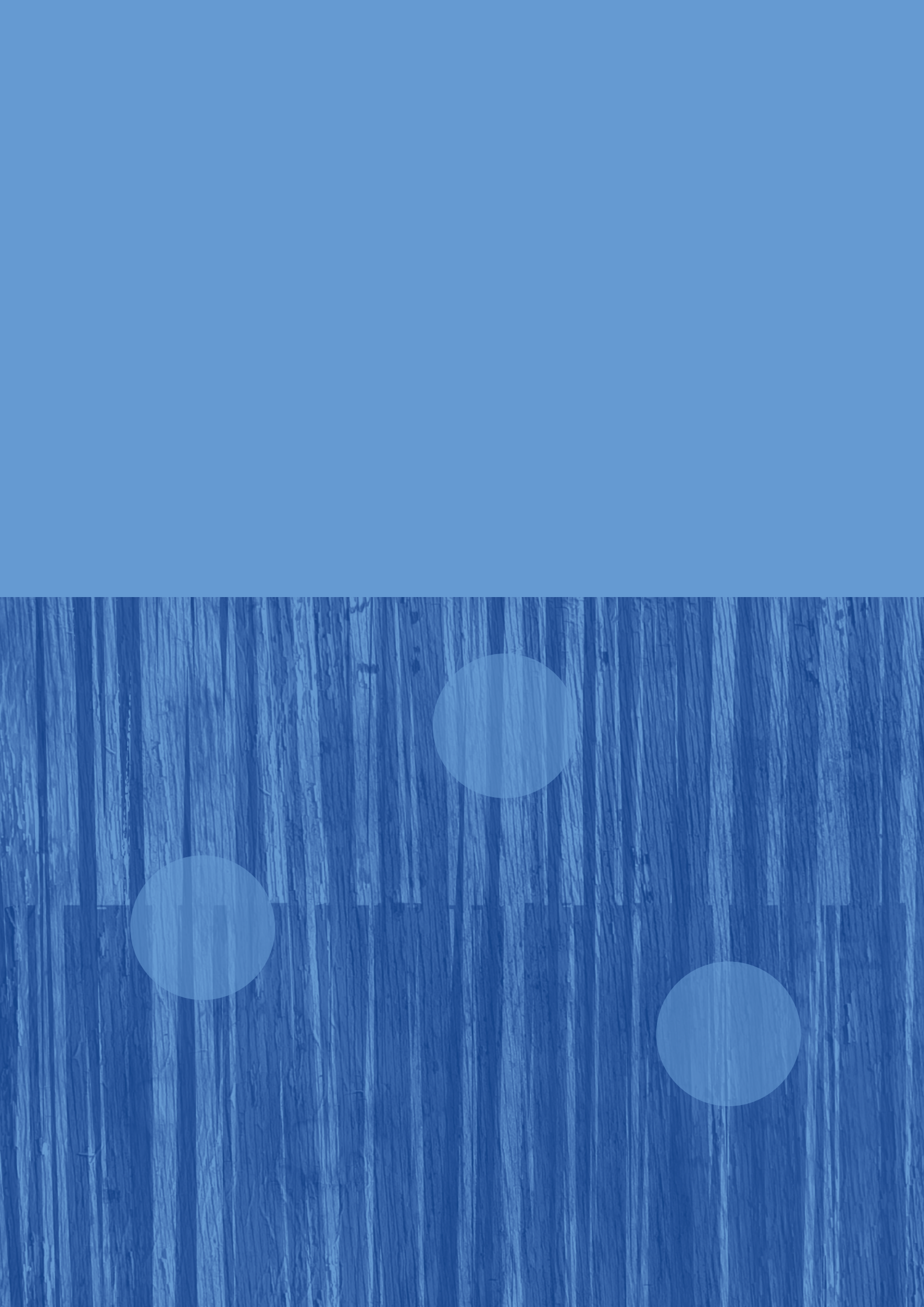
1. **Three key features of the financial system and its relationship with financial inclusion need to be kept in mind when shaping a systemic approach and how it is measured:**
 - The financial system involves a wide range of different actors with interdependent roles and relationships.
 - Financial markets are unpredictable and dynamic.
 - Progress to sustainable financial inclusion at scale can be slow and not necessarily linear.
2. **The five key principles of the systemic approach in financial inclusion programs and their most important measurement implications are as follows:**
 - ***Aim for systemic change.*** The measurement system needs to be able to detect and assess whether and how systemic change is happening. Results frameworks need a broad range of indicators, some of them qualitative.
 - ***Play a facilitative role.*** Enhance the scope of measurement beyond the intended results to include unexpected or unplanned outcomes and factors in the external environment that can affect progress both positively and negatively. Measure inward from change as well as outward from interventions.
 - ***Experiment, learn, and adapt.*** Monitoring needs to be linked to regular, data-informed, evaluative reviews. ToCs and results frameworks need to be regularly challenged, and funders and program unit staff must be prepared and enabled to change them.
 - ***Adopt a long-term view.*** Pressure to show results in a short amount of time at the highest levels of ToCs should be avoided. Instead, monitoring and impact evaluation should focus on the intervening steps in ToCs.
 - ***Take a different perspective on accountability.*** Accountability confined to the achievement of predefined targets is less appropriate in systemic programs. Use a broader model of accountability that requires programs to demonstrate a range of behaviors, such as enhanced forms of monitoring, effective risk management, and evidence-based adaptation.

Module 1. References and Resources

- Burjorjee, Deena M., and Barbara Scola. 2015 “A Market Systems Approach to Financial Inclusion: Guidelines for Funders” Washington, D.C.: CGAP. https://www.cgap.org/sites/default/files/Consensus-Guidelines-A-Market-Systems-Approach-to%20Financial-Inclusion-Sept-2015_0.pdf
- Davies, Gareth. 2016. “Getting to Scale: Lessons in Reaching Scale in Private Sector Development Programmes.” London: Adam Smith International. <https://beamexchange.org/resources/785/>
- Fowler, Ben, and Elizabeth Dunn. 2014. “Evaluating Systems and Systemic Change for Inclusive Market Development.” Washington, D.C.: USAID. http://pdf.usaid.gov/pdf_docs/PBAAC412.pdf
- Jenal, Marcus, and Mollie Liesner. 2017. “Causality and Attribution in Market Systems Development.” London, United Kingdom: The BEAM Exchange. <https://beamexchange.org/resources/950/>
- Kessler, Adam. 2014. “Assessing Systemic Change. Implementation Guidelines for the DCED Standard.” Cambridge, United Kingdom: DCED. http://www.enterprise-development.org/wp-content/uploads/Systemic_Change_DCED_Guide_August2014.pdf
- Sinha, Sunil, Johan Holmberg, and Mark Thomas. 2013 “What Works for Market Development: A Review of the Evidence.” UTV Working Paper. Stockholm: Swedish International Development Cooperation Agency. http://www.enterprise-development.org/wp-content/uploads/what_works_for_market_development-1.pdf
- Soursourian, Matthew, and Edlira Dashi. 2016. “Taking Stock: Recent Trends in International Funding.” Washington D.C.: CGAP. <http://www.cgap.org/publications/taking-stock-recent-trends-international-funding>
- Springfield Centre, The. 2015. “The Operational Guide for Making Markets Work for the Poor (M4P) Approach,” 2nd edition. The Springfield Centre. <http://www.enterprise-development.org/wp-content/uploads/m4pguide2015.pdf>
- World Bank. 2015. “The Global Findex Database 2014, Measuring Financial Inclusion Around the World.” Washington, D.C.: World Bank. <http://www.worldbank.org/en/programs/globalindex>

Note

3. The levels of the ToC are set out in Module 3.



THE MEASUREMENT SYSTEM



THIS MODULE . . .

- Describes the scope and components of a measurement system for systemic programs in financial inclusion.
- Discusses the importance of a conducive environment for measurement and a broader conceptualization of accountability.
- Explains how the measurement system is mobilized through a program measurement strategy and project measurement plans.

THE SCOPE OF A MEASUREMENT SYSTEM

An effective measurement instrument needs to be designed holistically and managed as a cohesive system, rather than as a set of stand-alone tools. The system includes several components:

- *Frameworks* define what is to be measured and how. They include ToCs and results chains, sets of measurement questions, results frameworks, and risk registers.
- *Processes* and *tools* are used to collect, process, assess, report, and use the information.

The measurement system needs to be mobilized through a measurement strategy at the program level and more detailed measurement

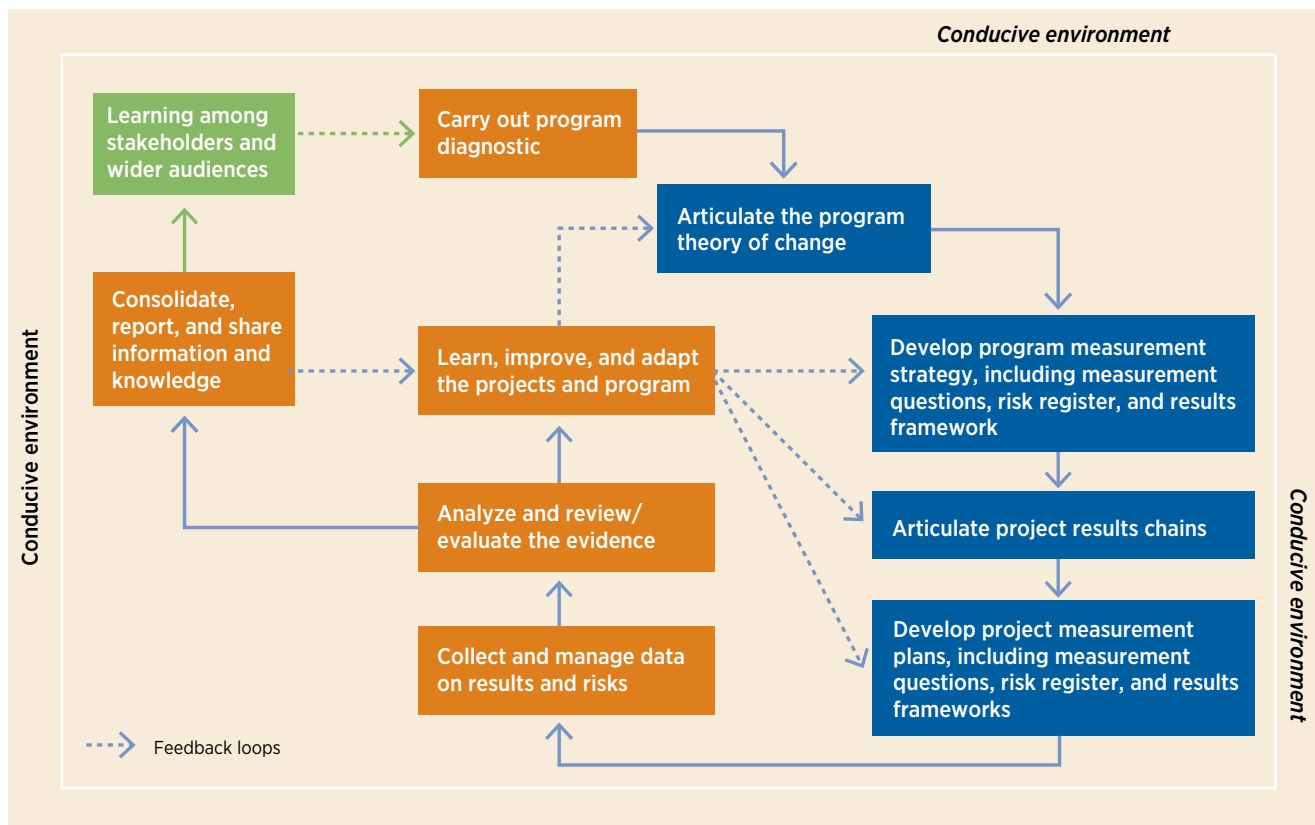
plans at the project level. These strategies and plans should also encompass complementary processes, such as measurement resourcing, role allocation, capacity building, and incentivization.

Figure M2-1 illustrates the main components of a measurement system and demonstrates the sequence in which the features are typically developed and iterated through both short- and long-term feedback loops. In the interest of simplicity, Figure M2-1 includes only the longer-term feedback loops, which are depicted by broken lines.

Although every program should have all the following components, the nature and scale of these components should be tailored to the needs of a specific program.

Diagnostic process. In systemic approaches, diagnostics aim to go beyond symptoms to

FIGURE M2-1. Measurement System Components and Cycle



identify the root causes of the problem—that is, the constraints that prevent poor and low-income people from accessing and using financial services. A diagnostic is a continuous process that not only serves program design, but also allows project management to regularly adapt interventions. From a measurement perspective, this first data-gathering effort can provide useful background information. For more information, see CGAP (2015).

Theory of change and results chains. A program ToC defines pathways from interventions to different levels of expected outcomes; it is usually expressed in the form of a diagram. It should include an explanation of the hypotheses on which progress along the pathways is based (the contribution hypotheses) and address the assumptions about external factors that support or at least do not hinder that progress. A ToC is applied at a program level; a results chain is the equivalent at a project level. A project results chain is narrower in scope and more context-specific than a program ToC, which should be broad enough to accommodate a range of projects. For more information, see Module 3.

Measurement strategies and plans. A measurement strategy provides broad and long-term guidance for measurement throughout a program or an organization. It principally covers high-level measurement matters, such as principles and standards, standardized indicators, and program-level evaluation. Project measurement plans complement the strategy. The plans provide more detailed information at the project level. They define what is to be measured, how, when, by whom, and at what cost. For more information, see the last section of this module.

Measurement questions. Measurement questions help to shed light on what stakeholders want and need to know about the program or project. They cover the whole spectrum from economy and efficiency in the use of resources to the program's impact and relevance, addressing both accountability and learning. From a learning perspective, the most important questions tend to relate to the ToC and results chain contribution hypotheses and assumptions. Program measurement questions tend not to change during the program. Project implementation tends to reveal new measurement questions at the project level,

sometimes superseding some of the original set. For more information, see Module 4.

Results frameworks. A results framework is a planning and measurement instrument that usually takes the form of a matrix. It includes a summary of intended interventions and outcomes that is directly related to the program ToC or project results chain and measurement-related information, such as indicators, baselines, targets, and means of obtaining data. The most common type of results framework is the logical framework or logframe. Project results frameworks need to dovetail with their counterparts at the program level—for example, through shared outcomes and, in some cases, standardized indicators. For more information, see Module 5.

Risk register. A program or project risk register is a framework of identified risks to interventions and outcomes, usually with ratings and information about risk management. Risk registers, particularly at the program level, may also cover risk categories such as reputation, fraud, staff, and data security. For more information, see Module 5.

Data collection. Data are collected in a variety of ways, depending on whether they are primary or secondary. Primary data are usually collected through surveys, observation, or self-monitoring. Collecting secondary data involves access to what has already been recorded, usually for a different original purpose. For more information, see Module 6.

Management of data. Once collected, data need to pass through several processes before they are ready to be analyzed and used to review, adapt, evaluate, report, and share knowledge. These processes include the transfer of data from the data collection points, quality assurance, storage, and retrieval. They involve IT applications—these do not need to be sophisticated or expensive. For more information, see the last section of this module.

Analysis. Data rarely speak for themselves. They need to be analyzed and presented in an accessible way before they can be useful for review and adaptation. Analysis should involve competent project staff and should not be done exclusively by measurement specialists. For more information, see the last section of this module.

Review. Regular internal review of the evidence emerging from monitoring and the assessment of the implications for adaptation (called enhanced monitoring in this Handbook) are essential. Although they are internal processes, reviews can include external stakeholders and can be facilitated by external staff, such as measurement consultants. In some cases—for example, at a project’s mid-point—reviews can take the form of internal formative evaluations. For more information, see Module 7.

Adaptation. The adaptation process is strongly linked to evidence-informed reviews, and it should be part of the measurement system. For more information, see Module 7.

Reporting. Reporting to stakeholders—both internal and external—plays an important role in the measurement system. Good quality and useful reporting flows from, rather than runs independently of, review processes. Well-presented, outcome-rich reporting leads stakeholders to request more such reports, which in turn create incentives to further improve the measurement system. For more information, see Module 8.

Evaluation. Internal review involves evaluative processes, but there is also a role for externally led evaluation at some point in the program or project cycle. Internal and external evaluative processes should complement each other. For more information, see Module 9.

Figure M2-1 includes a component for learning among stakeholders and wider audiences, which is outside the normal scope of a measurement system. However, in systemic programs, knowledge generated from measurement can be valuable to wider audiences.

A CONDUCIVE ENVIRONMENT FOR MEASUREMENT

Measuring market systems programs requires a range of program staff to take initiative and feel a sense of ownership. Measurement needs to be done in an environment that

- Confers legitimacy on measurement and incentives to participate in it fully.
- Provides adequate resources.
- Provides capacity building and on-going support.

Legitimacy and incentives. Measurement systems are intricately connected to the organization in which they operate, and they are influenced by its culture. Program-level measurement systems are also affected by external stakeholders, particularly funders. To increase its chances of being effective, measurement depends on several organizational factors, including a sense of ownership by program and project staff, strategic positioning of the measurement coordination unit, incentives to be open about results that are suboptimal, and explicit demand for and use of evidence by leaders.

Resources. Measurement systems require adequate financial and human resources. A program or large project would typically have a dedicated measurement specialist to manage its measurement system. Five percent of program budget is the minimum needed to fund the program or project; 10 percent or more of program budget may be required for innovative programs and projects where evidence is particularly important. In financial inclusion programs, measurement expertise combined with financial sector experience is hard to find, and these measurement specialists may command a high fee.

Capacity building and support. Measurement capacity among program staff is likely to be an ongoing challenge, especially where measurement is the work of several partners. Adequate investments in capacity building and support is needed.

AN ALTERNATIVE MODEL OF ACCOUNTABILITY

Both funders and senior management in their program implementing organizations need to create a conducive environment for measurement. The main objective is to create incentives for measurement practices that are transparent and that produce reliable data, while generating cycles of learning and adaptation. Accountability that focuses exclusively on quantitative targets can create the wrong incentives by tempting program staff to tailor their interventions “to chase the numbers” within target timeframes rather than explore longer-term, often less easily quantifiable but more sustainable, outcomes. Scale is essential and needs to be measured quantitatively, but at the right level in the results chain, and at the right time. The quantitative measurement needs to be complemented by measurement

of systemic change that is likely to lead to sustainable upscaling.

For example, program units should be accountable for the following:

- Designing projects based on thorough diagnostics and with an appropriate level of ambition.
- Putting in place and operating a proportionate and connected measurement and learning system.
- Providing credible evidence for results that cannot be expressed quantitatively.
- Managing risk effectively.
- Identifying aspects of interventions that have not worked well and making well-constructed cases for adaptation to interventions and their measurement frameworks.

Funders and senior management should do the following:

- Participate constructively in project formulation.
- Encourage, resource, and support the development of measurement and learning systems.
- Give program units appropriate levels of autonomy in their interventions.
- Legitimize and define clear protocols for adaptation, both in measurement frameworks and interventions themselves, engage constructively with proposals for adaptation, and learn from them.
- Avoid disproportionate demands for reporting.
- Have an open dialogue with program units about the data implications of their accountability to stakeholders such as executive boards, parliaments, or the general public.

This broader conceptualization of accountability will take time to achieve.

MEASUREMENT SYSTEMS ARE INTERCONNECTED

Project and program measurement systems are usually so closely related that the project system is said to be “nested” in the program system. Funders and the features of program unit measurement systems also need to be aligned, for example, in choosing intended results and reporting achievements.

Many organizations struggle to create and maintain coherent measurement systems because they can be complicated. Systems and

their dependencies need to be fully mapped so that they are understood by all stakeholders and that efforts can be made to streamline them. For example, implementing organizations often receive funding from more than one donor. If a system is not streamlined, program units may find themselves working with several different results frameworks, and reporting in different ways at different times. Stakeholders should be amenable to harmonizing these frameworks and processes as much as possible.⁴

Funders are usually preoccupied with managing their own systems and rarely have insights into how program unit systems look and perform. However, the quality of data entering funder systems from the programs depends on how effective the program measurement systems are. Funders need to provide incentives and support to program units to enable these units to build and maintain systems.

MEASUREMENT STRATEGIES AND PLANS

Organizations, programs, and projects need to plan to ensure that measurement systems are fit for purpose and effective throughout the life of the program or project. At the program

level, planning takes the form of measurement strategies that address measurement matters that transcend individual projects. These strategies chart a long-term course for monitoring and evaluation across organizations and large programs, and they paint a broad picture of the program-level measurement system, including high-level frameworks. They include strategic measurement questions and can include measurement policy if this is not set out elsewhere. This information is often contained in measurement manuals.

Measurement strategies should be complemented by more detailed and regularly updated measurement plans at the project level. Measurement plans cover all the frameworks and processes used for project monitoring and evaluation. They provide details on what is to be measured, how, by whom, when, and at what cost. They enable project managers to ensure that resources, capacity, and processes are in place to meet the project's measurement needs and stakeholder requirements. The measurement plan should include the project's key frameworks: the project results chain, set of learning questions, results framework, and risk register. Measurement strategies and plans may overlap, but they should not contradict each other. Table M2-1 provides a comparison of measurement strategies and plans.

TABLE M2-1. The Scope of Measurement Strategies and Plans

ASPECTS	STRATEGY	PLAN
Scope	Entire program, including matters that are common to all projects	Project
Principles and standards	Measurement throughout the program	Covered by strategy
Measurement stakeholders	Program stakeholders	Project stakeholders
Management and governance	Management and governance of measurement at program level; oversight of project measurement—roles and responsibilities	Roles and responsibilities for measurement in the project, including among partners
Human resources	Guide to human resourcing of measurement throughout the program, including a competency framework (relates to principles and standards)	Any measurement awareness-raising and capacity-building required for project staff and partners
Frameworks included	Program ToC	Project results chain
	Strategic measurement questions	Project measurement questions
	Program results framework, including project-level standardized indicators	Project results framework
	Program risk register, including, but not limited to, those related to ToC assumptions	Project risk register—usually related to results chain assumptions
	Reference sheets for indicators	Reference sheets for indicators
Monitoring	<ul style="list-style-type: none"> Reference sheets for standardized indicators Guidance on monitoring does not depend on project data (e.g., monitoring of program management and communications) 	<ul style="list-style-type: none"> Reference sheets for indicators Planning for baseline data collection (cross-referring to reference sheets for indicator baseline data) Guidance on monitoring beyond indicators: processes for capturing wider information about intended outcomes, unplanned results (positive and negative), and factors in the external environment that may be influencing the project
Data management	Processes for data channelling from the projects, and subsequent management of data, including quality assurance and preliminary analysis	Processes for data monitoring, including quality assurance and preliminary analysis
Internal review, adaptation, and risk management	<ul style="list-style-type: none"> Processes for conducting an internal program review, including a standard set of review questions Processes for adapting and continually managing risk 	<ul style="list-style-type: none"> Processes for doing an internal review of the project, including a standard set of review questions Processes for adapting and continually managing risk
Reporting and knowledge sharing	Processes for reporting and sharing knowledge with program stakeholders	Processes for reporting and sharing knowledge with program management and other project stakeholders
Evaluation	Initial planning for evaluation, describing the type, scope, resourcing, and timing of evaluation(s) at program level and pointers for project evaluations	Initial planning for project evaluation(s), describing the type, scope, resourcing, and timing



KEY TAKEAWAYS

- A measurement system consists of frameworks (e.g., a ToC or results chain, a results framework, and a risk register), processes (e.g., for data collection and management, review and adaptation, and evaluation), and related tools (e.g., data collection instruments, reporting templates, data processing, and storage applications).
- Measurement systems, particularly in systemic programs that demand regular and deep engagement by program staff and partners, need a conducive environment characterized, for example, by incentives to be open about results and capacity building and support. Funders and senior management need to foster a conducive environment.
- Project measurement systems have multiple linkages to the system in the governing program. Program systems often have touchpoints with the measurement systems of donors and other stakeholders, such as governments. These linkages and touchpoints need to be mapped and managed.
- Measurement systems need to be mobilized through high-level measurement strategies at the organizational or program level, and complementary, detailed, measurement plans at project level.

Module 2. References and Resources

BEAM Exchange, The. 2015. "Monitoring Guidance." London: The BEAM Exchange. <https://beamexchange.org/guidance/monitoring-overview/>

Burjorjee, Deena M., and Barbara Scola. 2015. "A Market Systems Approach to Financial Inclusion: Guidelines for Funders." Washington D.C.: CGAP. https://www.cgap.org/sites/default/files/Consensus-Guidelines-A-Market-Systems-Approach-to%20Financial-Inclusion-Sept-2015_0.pdf

DCED (Donor Committee for Enterprise Development). 2016. "DCED Monitoring and Measuring Results in Private Sector Development." Cambridge, United Kingdom: DCED. <http://www.enterprise-development.org/page/measuring-and-reporting-results>

FSD Africa and Oxford Policy Management. 2015. "Towards Impact-Oriented Measurement (IOM) Systems for the FSD Network" Nairobi: FSD Africa and Oxford Policy Management. <https://www.fsdafrica.org/knowledge-hub/documents/towards-impact-oriented-measurement-for-the-fsd-network/>

Görgens, Marelize, and Jody Zall Kusek. 2009. "Making Monitoring and Evaluation Systems Work: A Capacity Development Toolkit." Washington, D.C.: World Bank. <https://elibrary.worldbank.org/doi/abs/10.1596/978-0-8213-8186-1>

Note

4. This was one of the principles articulated in the Paris Declaration on Aid Effectiveness of 2005.



THEORY OF CHANGE AND RESULTS CHAINS



THIS MODULE . . .

- Examines the roles of ToCs and results chains.
- Offers a model conceptual ToC for financial inclusion programming.
- Illustrates a program ToC and a project results chain.

DEFINING THE THEORY OF CHANGE AND RESULTS CHAIN

For use in program measurement, the program ToC needs to be a simplified reflection of the program strategy that traces the anticipated progression from the interventions through successive levels of change to the highest outcome in focus—usually what this Handbook calls the “development outcome.” The ToC should be based on the best evidence available of what works and in what circumstances. This type of ToC is usually expressed as a diagram, but it may also include narratives. These elements should include the key assumptions about external factors on which this progression relies. It may also be accompanied by a definition of the problem or opportunity that the program is designed to address and by “contribution hypotheses,” which explain the underlying logic behind the progression from one level of the ToC to the next. Contribution hypotheses are valuable when the program is being reviewed or evaluated.

A program ToC is often developed from a broader, more conceptual ToC. This process can be managed incrementally as program design progresses. A conceptual ToC is valuable at an early stage in program formulation and planning. It helps to frame the program within the wider context of how change is understood to happen in the development domain in question. (A development domain is defined as a sector, subsector, or broad focus area, such as financial inclusion, for development interventions.) A more context-specific program ToC is progressively developed, as the diagnostics, conducted by the funder and/or the program unit, clarify the specific focus of the program and the intended outcomes at each level. It will, for instance, give more details on the components of the financial systems that the program aims to influence and on the types of intervention with which it intends to do so. At the start of program development, funders and program units should agree on the conceptual ToC they will work from. At a later

stage, after diagnostics have reached an advanced stage, both should participate in designing the program ToC.

The term *results chain* is sometimes used interchangeably with *ToC*. This Handbook uses results chain to refer to a ToC at project rather than program level.

A project results chain is rooted in the program ToC, but because it refers to a specific project, it is narrower in scope and more contextualized. The results chain defines the path from the interventions to the intended outcomes of a project, with more granularity—more steps—than the program ToC. Because it is project-specific, it can be more precise about the types of partners, the interventions, and their anticipated scale. Results chains are usually designed by the program unit, after a thorough diagnostic of the environment. Project interventions and outcomes should be easily matched with those in the ToC, although the results chain may not reach the higher levels of outcome in the program ToC because of the project's narrower scope and generally shorter timeframe.

Program ToCs and project results chains are primarily used to set the agendas for measurement—both monitoring and evaluation. But they are also valuable tools for communicating with program partners and wider audiences.

A CONCEPTUAL TOC

A conceptual ToC for the systemic approach to financial inclusion, shown in Figure M3-1a, would typically be used at the beginning of the development of a program strategy—for orientation—and as the touchstone for the later development of a program ToC. The ToC in Figure M3-1a does not prescribe the scope that programs of this nature should have. Some programs—such as single-country facilitation programs—may aim to contribute to developing all major parts of the financial market system, and therefore may have ToCs that cover the same ground as this conceptual model. Their building blocks, however, would be more specific because a program ToC needs to be evaluable—to translate readily into a results framework, the central instrument for measurement in programs and projects (see Module 5).

ToC levels

The conceptual ToC in Figure M3-1a has five levels, listed in order of highest to lowest level (for a vertical ToC):

- Development outcome
- Inclusive financial system
- Systemic change
- Intermediate outcomes
- Interventions

The levels are different in terms of timeframe (the higher the level, the longer the timeframe) and sphere of influence (the higher the level, generally the weaker and less direct the program influence). These two distinctions have clear implications for measurement; measuring outcomes will require different approaches, depending on the level.

Development outcome

In this Handbook *development outcome* refers to the highest level of the ToC. *Impact* is often found at this level, but this Handbook avoids its use in the context of ToCs and results chains for two reasons. First, impact in measurement is taken to mean the **net effect** of interventions.⁵ However, the higher levels of a ToC or results chain do not signify the net effect of the lower levels.⁶ On the contrary, the higher the level, the greater the contribution of external factors. Second, as pointed out in Module 1, the impact of the intervention in this domain should be measured at more than one level in the ToC or results chain. Most importantly, it needs to be measured at the inclusive financial system level.

The development outcome usually captures change in the well-being of poor people. In the conceptual ToC, the chosen development outcome is “improved economic well-being,” though some programs may aim for other pro-poor outcomes, including one or more relevant Sustainable Development Goals. Although it is common to specify only one outcome at this level, more than one could be included.

Depending on the status of the financial system at the beginning of the program, large-scale and sustained development outcomes may not happen for many years after programs begin. Evidence of change on a small scale should be detectable early on, but major programming decisions should not be based on this early evidence.

FIGURE M3-1a. A conceptual theory of change for financial inclusion

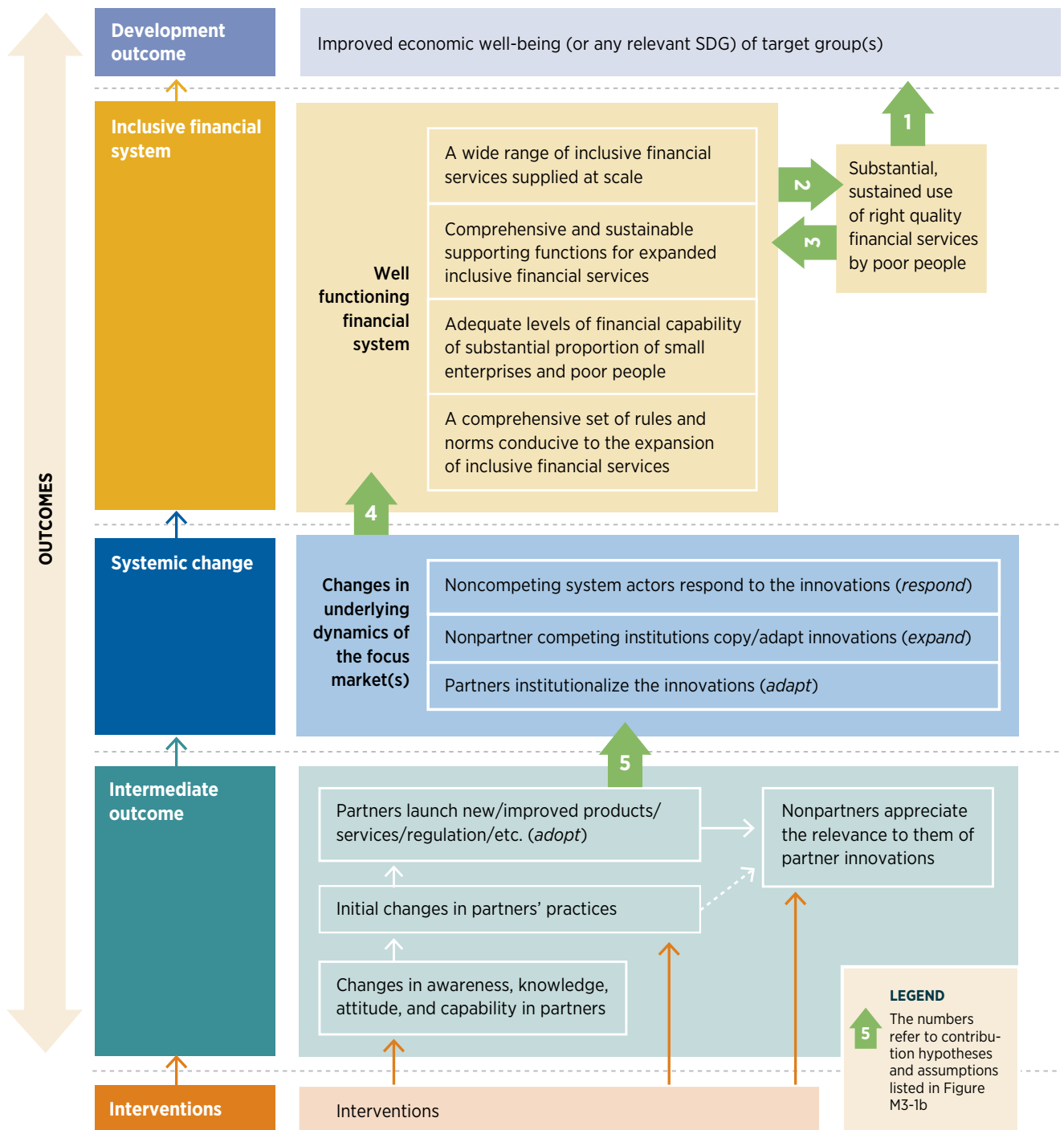


FIGURE M3-1b. Contribution Hypotheses and Assumptions

1

HYPOTHESIS #1: A fully functioning financial market system that is supported and sustained by financial inclusion has a substantial impact on livelihoods, vulnerability, and other positive development outcomes.

Assumption: Poor people have access to other basic services (affordable energy, housing, education, etc.).

2

HYPOTHESIS #2: A fully functioning market system promotes financial inclusion, enables and incentivizes access to relevant, affordable financial services for underserved people, who have the capability to take up and benefit from these services.

Assumption: Enabling social, economic, and political context (no economic crisis, no conflict, etc.).

3

HYPOTHESIS #3: Financial inclusion leads to sustained demand for relevant, affordable financial services that reinforce the incentives of other market actors and other components of the financial market system to provide these services.

Assumption: Enabling social, economic, and political context (no economic crisis, no conflict, etc.).

4

HYPOTHESIS #4: Systemic change reaches a tipping point where it becomes self-reinforcing and permanently changes the incentives, inter-relationships, and other dynamics among and between market actors. This leads to a fully functioning system.

Assumptions: Time is allowed for market development; enabling social, economic, and political context (no economic crisis, no conflict, etc.).

5

HYPOTHESIS #5: The changes generated by the program are visible and credible to other market actors who begin to invest in similar practices or respond with complementary ones; partner organizations are incentivized to continue to invest in innovation.

Assumptions: Nonpartner organizations have the incentives, capability, and resources to copy/adapt or respond; innovations supported by the program are successful.

Inclusive financial system

This level encompasses two interdependent elements: changes in the financial system (well-functioning financial system) and changes in financial inclusion (substantial, sustained use of high-quality financial services by underserved people). Financial inclusion flows from and reinforces development of the financial system: weaknesses in the system prevent advances in financial inclusion, and without substantial, sustained advances in financial inclusion, advances in the financial system are likely to unravel (e.g., market actors would stop providing supporting functions because of lack of demand, and regulators would lose focus on the bottom-of-the-pyramid segment).

Depending on the condition of the financial system at the beginning of the program, substantial systemwide advancement may be beyond the scope and timeframe of a 4–6-year program. However, significant changes in parts of the system should be evident before the end of a program.

Systemic change

This level represents tangible signs of changes in the orientation and practices of a range of

actors, and the dynamics of their engagement with markets that are likely to contribute to the development of an inclusive financial system.⁷ For instance, outcomes will include compelling a substantial number of nonpartner FSPs to provide sustained inclusive financial services—crowding-in—a crucial feature of an inclusive financial system.⁷ Similarly, while the outcome at the financial-system level might be a comprehensive set of rules and norms conducive to the expansion of inclusive financial services, at the systemic change level, a satisfactory outcome might be the early stages of the translation of policy reform into improved regulation. Significant change at this level should be detectable by the end of a 4–6-year program.

Intermediate outcomes

Intermediate outcomes primarily capture the changes directly supported by the program. These outcomes mainly happen with direct partner institutions (e.g., FSPs, support-function providers, and government agencies) or key individuals within them. They start with initial, but encouraging, changes that can be cognitive (awareness and knowledge), affective (attitudes), and behavioral

(practices). These types of change should be detectable early in the program. They indicate that partners are moving toward adopting the innovation that is expected to be completed by the end of the program.

Intermediate outcomes also capture the initial steps toward scaling up the outcomes of the program, by attracting both competing and noncompeting nonpartner market actors to contribute to the innovation in their respective ways. For example, a competing financial services provider might introduce a similar innovative financial service, and a government agency might respond to the innovation by removing restrictive regulation. Interventions designed to attract these nonpartners need to be explicitly made in the ToC, especially in a systemic approach to financial inclusion. Programs that demonstrate notable outcomes often attract other actors, but experience has shown that this demonstration effect rarely happens on its own and that program leaders need to be more intentional about it.

Interventions

This level encompasses the support given to primary program partners, such as FSPs, support function providers, and government agencies, together with the action needed to promote scaling up by organizations not regarded as primary program partners. The most frequent interventions include direct funding, technical assistance, and capacity building, and less direct action such as advocacy, the production and dissemination of knowledge products, and the facilitation of networking.

Contribution hypotheses and assumptions

A ToC should contain or be accompanied by explicit and reasonable assumptions that identify the main external factors of progression through the ToC. It is helpful to also formulate contribution hypotheses that explain the logic behind the progression from one level to the next. Figure M3-1b on page 20 provides some generic assumptions and hypotheses to illustrate how they might accompany the conceptual ToC. The numbers correspond to those inside the green arrows in the figure.

THE AAER FRAMEWORK

The conceptual ToC focuses on systemic change. One of the best-known models for conceptualizing the process of systemic change is the Springfield Centre’s Adopt, Adapt, Expand, Respond (AAER) framework (see Figure M3-2) (The Springfield Centre 2015). This Handbook interprets the framework in the context of financial inclusion programs.

The change process in financial inclusion begins with **Adopt** in the lower, left box of the model. This reflects program-supported innovation, which, despite the support, requires a considerable degree of buy-in from partners.

Systemic change encompasses the **Adapt**, **Expand**, and **Respond** phases. Adapt signifies (i) a high degree of autonomous strengthening and (ii) adaptation by the partners of the initial innovation of their modes of operating in general. Both aspects would be underpinned by partner investment of resources.

FIGURE M3-2. The AAER Framework



Adapted from The Springfield Centre

The right side of the model represents broader systemic change. Expand refers to investment by competing market actors (e.g., nonpartner FSPs or providers of supporting functions). Expand involves some degree of copying the behavior and models of the program partners.

Respond represents the actions of noncompeting actors—actors who are not competing with program partners in the market—who adjust their practices, services, or regulations in response to the innovation brought about by the program. They respond because they perceive opportunities to pursue their objectives or mandate. For example, services providers might perceive new, complementary market opportunities or government agencies might identify a need to adjust regulation.

These components of the AAER framework are embedded in the conceptual ToC. The Adopt phase, where the innovation is progressively introduced by program partners, is an intermediate outcome. The Adapt phase, involving the institutionalization of the innovation by partners, is at the systemic change level, although some programs might consider it to be an intermediate outcome if they intend to continue to support partners beyond the launch of the initial innovation. Expand and Respond are integral parts of systemic change in a ToC. They provide evidence of change in the underlying dynamics of the market system, beyond the partners.

From the conceptual ToC to a program ToC

A conceptual ToC (see Figure M3-1a on page 19) is used to broadly frame a program, before a more thorough diagnostic is performed. With this diagnostic, the funder and/or program unit—in varying combinations—can design the program and construct a more specific ToC that can help to frame projects within the program and form the platform for program-level measurement. Although the program ToC should have the same levels as the conceptual ToC, the program ToC should be more specific at each level. It should include, for example, the following:

- Aspects of a well-functioning financial system that will be targeted by the program and more details on the systemic changes necessary to improve the way the market functions.⁸

- Granularity and context at the intermediate outcome level.
- At the intervention level, the program ToC should refer to the mechanisms—such as direct funding, technical assistance—likely to be used but mostly without tying them to specific outcomes, to ensure that individual projects have the built-in flexibility necessary for their specific contexts.

The program ToC should be the following:

- Grounded
- Progressive
- Plausible
- Transparent
- Dynamic
- Simple
- Evaluable

Grounded. The ToC should spring from and address a central and clearly defined problem or opportunity. It should be based on a rigorous diagnostic process that leverages evidence from research and evaluation and different stakeholder perspectives for its construction.

Progressive. The different levels of the ToC should represent a progression. The levels should be thought of as clear building blocks that identify the intended outcomes and the proposed steps needed to achieve these outcomes. A higher level should be more than the sum of the results at the lower level; it should represent a change that builds on the lower-level results.

Plausible. The transition from one level to the next (the progression narrative) must be plausible and must reflect the challenges in the program environment and the scale of the intervention. Two key aspects of plausibility are *relevance* (are the interventions or outcomes at one level likely to contribute to the outcomes at the next level?) and *sufficiency* (are these contributions likely to be powerful enough to lead to the outcomes at the next level, given the accompanying assumptions?).

Transparent. The ToC is a theory—it is a hypothesis, not a blueprint. The underlying assumptions and hypotheses should be established and periodically tested. Assumptions that are deemed to be unreliable should be clearly identified, and may require specific activities, such as advocacy, knowledge dissemination, increased market coordination, and in some cases, additional projects, to strengthen them.

Dynamic. The ToC should be dynamic. It should be developed at a level high enough to set broad long-term parameters, but it should be flexible enough to allow for adaptation during implementation. The upper levels of the ToC, from development outcome to systemic change, should be defined with the expectation that they are unlikely to change during the program. The lower levels should be broad enough to allow for the envisaged range of projects but also be open to adaptation in light of experience. Market systems are unpredictable and inevitably need to be flexible to ensure interventions are relevant.

Simple. ToCs are often criticized for their simplicity, especially in the complex and unpredictable context of market development, but this simplicity is what makes them useful. They should enable stakeholders, both internal and external to the program, to readily understand the nature, scope, and objectives of the program and how the interventions are intended to contribute to them.

Evaluable. The building blocks of the program ToC need to be clear and evaluable. First, it should be obvious to stakeholders what the different levels and building blocks mean. Second, the outcomes should be clear enough to be interrogated through measurement. This means that translation of the program ToC to its results framework should be straightforward.

Figure M3-3 on page 24 illustrates a program ToC—excluding the contribution hypotheses and assumptions, which would be similar to those in the conceptual ToC. The program envisaged focuses on promoting pro-poor client-centered financial services with financial services providers with two critical support functions—capacity building and information services for client centricity—as the primary targets. In this example, the program design builds on diagnostics that identified a lack of client centricity among financial services providers as an obstacle to pro-poor financial market development in three countries of interest to the funder. An assessment indicates that government and financial services industry associations are likely to be receptive—and have the capacity—to adapt the regulatory frameworks and member voluntary codes. Significant investment of effort in that area is not

planned, apart from “softer” interventions, such as awareness-raising and networking.

The program is expected to lead to at least one project in each of the three countries. Each project will have a coherent set of interventions. A results chain for a possible project is illustrated in Figure M3-4.

The ToC in Figure M3-3 is an illustration; it should not be used as is. Funders or program units need to adapt this example to address their specific program contexts and diagnostic needs.

Caveats about ToCs

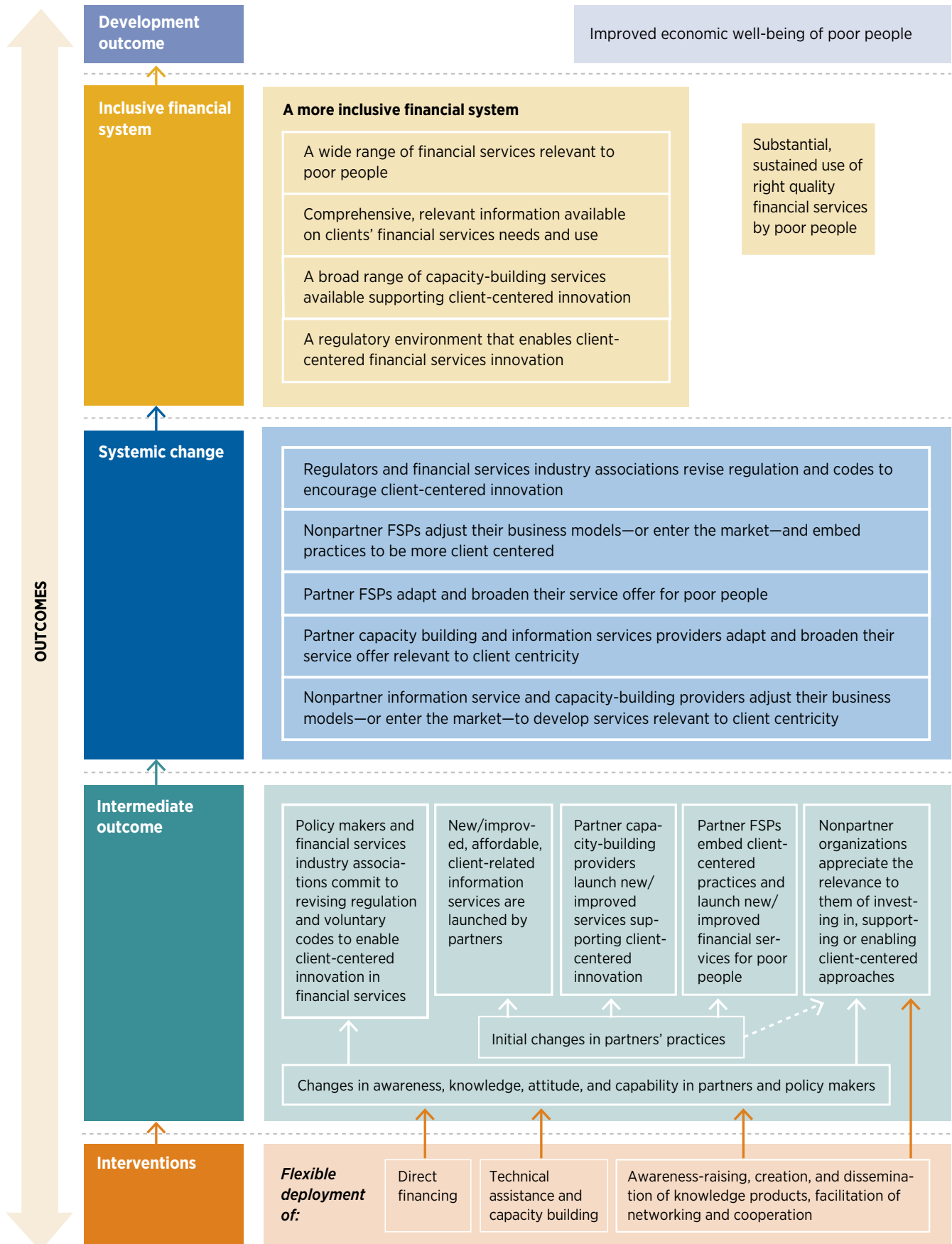
Because they are linear and deterministic (the sense that if X is achieved, then Y is bound to follow), ToCs can be challenged by the complexity and unpredictability of financial inclusion programs. Even nonlinear visual tools, such as systems maps with complex feedback loops, are simplifications of programs and their environments. The key to ensuring a ToC continues to work for a program is to make maximum use of its internal flexibility and at the same time be prepared to adapt it if the program does not unfold as predicted.



TIPS: DESIGNING A TOC

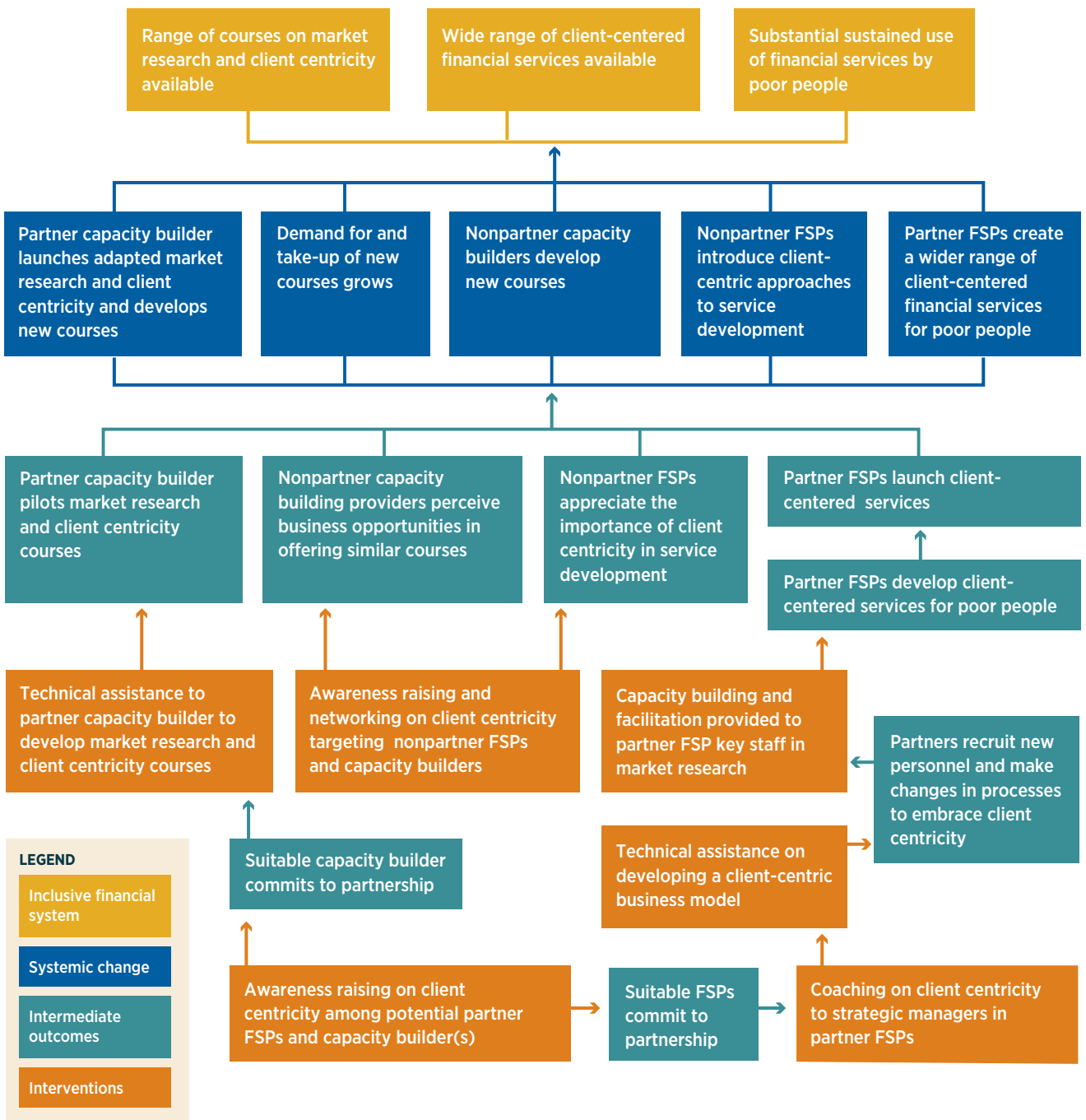
- Funders and program units should collaborate so that they both own and use the program ToC, including its iterations.
- The program ToC should be based on a thorough understanding of the market dynamics that is gained through obtaining and triangulating evidence from research and evaluation and different stakeholder perspectives.
- ToCs are not designed once and for all. Their components, including the assumptions and contribution hypotheses, should be reviewed and tested periodically as the program advances. They should be adapted, along with the interventions themselves, if evidence suggests that the hypotheses are not viable.

FIGURE M3-3. An illustration of a program theory of change



Note: FSPs = financial services providers

FIGURE M3-4. An example of a results chain for a client centricity project



Note: FSPs = financial services providers

RESULTS CHAINS

Projects need their own ToCs. This Handbook uses the term “results chains” for ToCs at the project level because there are nuances about their form and use that set them apart from the type of program ToC illustrated by Figure M3-3. The program ToC provides a coherent framework for the various projects that fall under the program. Because projects flow

from the overall program strategy, a project’s results chain should be rooted in the overall program ToC. This means that the results chain outcomes should be identifiable in the program ToC; however, they are usually narrower in scope and context-specific.

A results chain should be tightly focused on the field of influence of the project, therefore, it might not include the development outcome, or even the inclusive financial system level. It

may not include the full range of systemic change outcomes.

Interventions in the results chain are more detailed: They often refer to specific partners, and they are linked to specific outcomes. It can be helpful for results chains to be more granular than the ToC, for example, identifying more steps at both the intervention and intermediate outcome levels. These two levels often interact in the results chain, which reflects the fact that interventions and intermediate outcomes can be iterative and build on each other.

A project results chain needs to be translated into a results framework or similar measurement instrument. In a granular results chain, decisions need to be made about which steps

should be in the framework, and therefore have indicators, and which ones can be left to less formal monitoring.⁹

The program unit is likely to take the lead on developing the project results chain. The results chain is constructed once there is a thorough understanding of the current state of the financial system in the specific markets to be covered by the project. Despite the high level of specificity in the results chain at any one time, it must maintain its flexibility. The results chain needs to be reviewed and adjusted in line with adaptations in the project itself.

Figure M3-4 presents an example of a results chain for a project stemming from the program illustrated in Figure M3-3.



KEY TAKEAWAYS

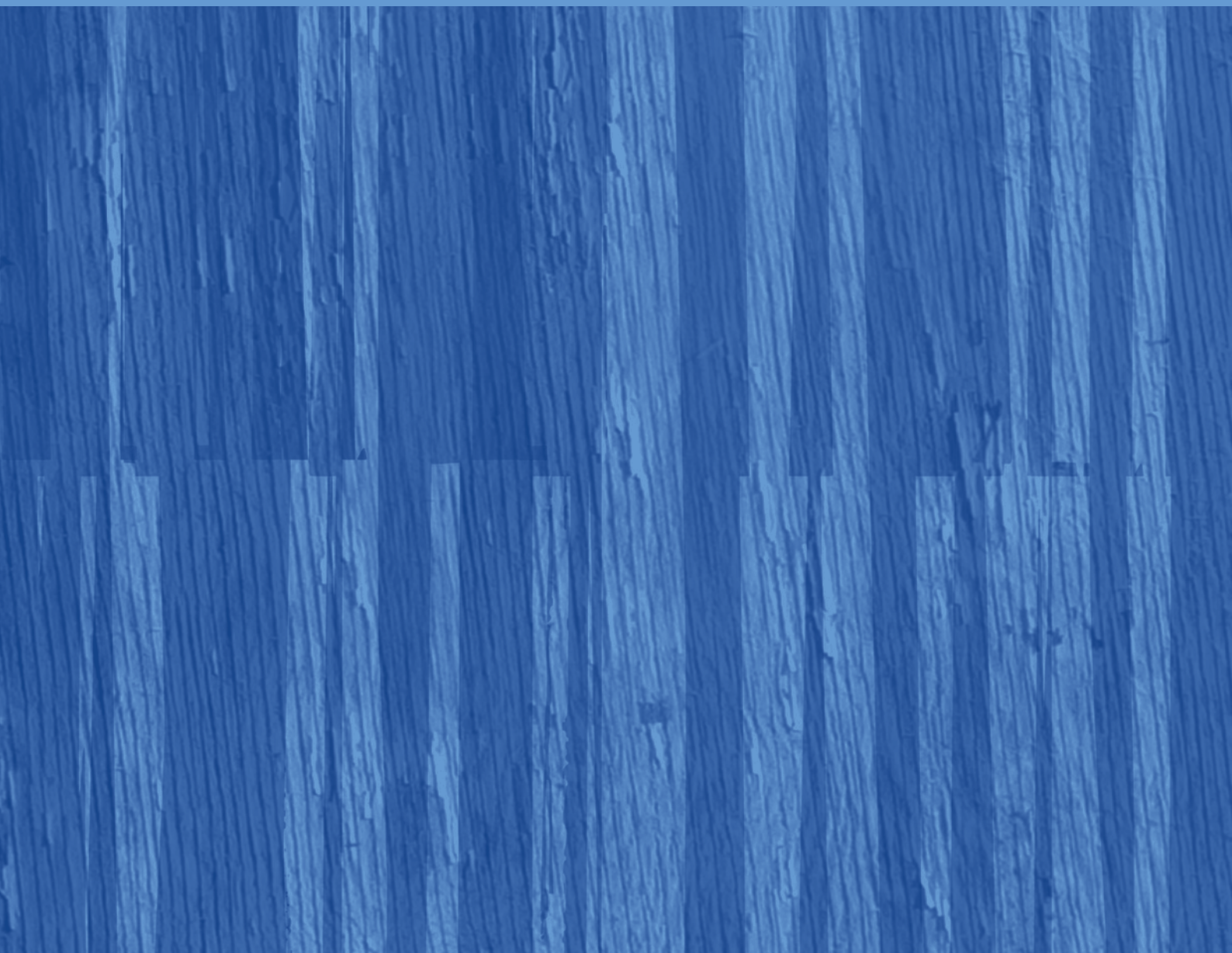
1. **A conceptual ToC is a valuable instrument at the early stage of program design and for wider communication. It needs to be complemented by a more specific program ToC and a more focused results chain that align with the program ToC for projects.**
2. **Program ToCs and project results chains have foundational roles in measurement. They lead directly to the construction of results frameworks and are the platforms for enhanced monitoring and for evaluation.**
3. **ToCs and results chains need to satisfy several criteria, particularly:**
 - The progression narrative needs to be based on the best available evidence, and there needs to be clear and reasonable assumptions and hypotheses behind this narrative.
 - The building blocks of ToCs and results chains need to be evaluable.
 - ToCs and results chains need to be understood and treated as hypotheses and not blueprints. They should be reviewed and adapted during program implementation.

Module 3. References and Resources

- Fiala, Nathan, and Markus Pilgrim. 2013. "Intervention Models and Performance Indicators of the ILO Small Enterprises Unit." Geneva: ILO. <http://www.enterprise-development.org/wp-content/uploads/ILOModels.pdf>
- Kessler, Adam, and Nabanita Sen. 2013. "Guidelines to the DCED Standard for Results Measurement: Articulating the Results Chain." Cambridge, United Kingdom: DCED. http://www.enterprise-development.org/wp-content/uploads/1_Implementation_Guidelines_Results_Chains_Apr_2015.pdf
- Nippard, Daniel, Rob Hitchins, and David Elliott. 2015. "Adopt-Adapt-Expand-Respond: A Framework for Managing and Measuring Systemic Change Processes." Briefing Paper. Durham: The Springfield Centre for Business in Development. <http://www.springfieldcentre.com/wp-content/uploads/2014/06/2014-03-Adopt-Adapt-Expand-Respond-Briefing-Paper1.pdf>
- Springfield Centre, The. 2015. "The Operational Guide for Making Markets Work for the Poor (M4P) Approach," 2nd edition. The Springfield Centre. <http://www.enterprise-development.org/wp-content/uploads/m4pguide2015.pdf>
- Stein, Danielle, and Craig Valters. 2012. "Understanding Theory of Change in International Development. A Review of Existing Knowledge." London: London School of Economics and Asia Foundation. http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/UNDERSTANDINGTHEORYOFChangeSteinValtersPN.pdf
- Vogel, Isabel. 2012. "Review of the Use of 'Theory of Change' in International Development." London: DfID. <https://www.gov.uk/government/news/dfid-research-review-of-the-use-of-theory-of-change-in-international-development>

Notes

5. In this Handbook, net effect is defined as the positive and negative, primary and secondary, medium- to long-term, intended or unintended effects of interventions.
6. *Higher* assumes a vertical ToC. ToCs can also be presented horizontally.
7. Markets at this level are not limited to the financial market. To sustainably change the way financial systems work, financial inclusion programs might have to focus also on changing how interconnected markets work (e.g., those for capacity-building or data services).
8. Program ToCs could include all aspects of the market system, with some form of differentiation—such as color-coding—to indicate which are specifically targeted by the program. This can result in a complex ToC. The alternative is to address the nontargeted aspects of the system through explicit assumptions.
9. The latter would constitute an example of enhanced monitoring, which is covered in Module 7.



MEASUREMENT QUESTIONS



THIS MODULE . . .

- Explains how measurement questions help to sharpen the relevance of both monitoring and evaluation throughout program and project cycles.
- Provides guidance on identifying key measurement questions.

WHY FRAME MEASUREMENT QUESTIONS?

Data collection is time-consuming and expensive. It is easy to be overwhelmed with data. Data collection and analysis need to focus on what is important to know. Too often, measurement is shaped by what can easily be measured rather than what stakeholders need to know. Far better an approximate answer to the right question than an exact answer to the wrong question. In fact, measurement that lacks focus not only wastes time and money but can mislead program stakeholders. It is important to be clear about what questions monitoring and evaluation need to answer.

WHY DEFINE QUESTIONS AT THE BEGINNING OF THE CYCLE?

Measurement questions are customarily identified when launching an evaluation or impact assessment. But this is too late for their optimal

use. Data that are critical for the evaluation—both at baseline and during implementation—may not have been collected. Cycles of review and adaptation throughout the program will be weaker.

Measurement questions should be identified at the start of a program or project. This is particularly important in complex environments, such as financial market systems, where the landscape for monitoring is less well-defined and needs the compass that measurement questions can provide. Identifying measurement questions early on does the following:

- **Guides monitoring.** ToCs tell us what interventions and outcomes to include in the results frameworks, but they do not say much about monitoring. Measurement questions point to where monitoring is particularly important and provides clues as to the nature and number of indicators required.
- **Drives the regular use of data for learning and decision-making.** Monitoring is not just

about collecting data. Enhanced monitoring should probe and test the hypotheses on which the program is founded. It should ask why and why not questions and identify what the data and analysis mean for the program going forward or for future programs. Posing these questions at the beginning of the cycle helps to ensure that program stakeholders regularly apply this type of thinking to the data that are collected and use them effectively.

- **Supports evaluation** by ensuring that relevant baseline data are collected and that monitoring continues to collect the types of results information that eventually will be needed.
- **Shapes research agendas.** Not all questions of interest to stakeholders can be adequately answered through measurement within a program or project. Defining these questions early in a program or project can help funders and other stakeholders decide whether and where to invest in complementary research, such as impact evaluation, to more fully answer the questions.
- **Promotes consensus around accountability and learning priorities.** Framing measurement questions can help funders, program units, and other stakeholders see eye-to-eye on measurement. Funders tend to ask questions about both the impact of programs and the value for money. Program units will be more interested in how effective their interventions are in the near term so that they can adjust their focus and strategies if necessary. Other stakeholders may be interested in the effectiveness of the program or project to make decisions about their own interventions. A collaborative approach to defining measurement questions and applying them promotes clarity and balance in the measurement system, and realistic expectations of what should be measured and when. Having constructive dialogue around measurement questions at the beginning of the program cycle will help to ensure that measurement is adequately resourced.

IDENTIFYING THE QUESTIONS

Measurement questions are usually a mix of generic and ToC-specific questions. The generic questions are likely to be relevant for every program or project, particularly those related to effectiveness and efficiency as

defined in the OECD-DAC framework (OECD 1991).¹⁰ Appendix A for this module provides examples of generic questions and indicates whether they are more likely to be answered through monitoring or evaluation, or both. These generic questions are organized along the different levels of the ToC as presented in Module 3.

At the intervention level, they include questions on the type of resources provided and the timing, scope, and relevance of the activities. At the various outcome levels, they cover achievement of expected outcomes and changes other than those that were planned. Questions also cover the overall relevance and coherence of the program and project: Was the program the right thing to do? Could it have been done better, or by other parties? Is this type of program still needed?

Program measurement questions are broad and high level. Measurement questions for projects are more detailed and focused. Program measurement questions are usually determined before projects are formulated, and to some extent they shape the measurement questions for those projects. At the same time, some program questions may come from those defined at project level, particularly where more than one project independently identifies the same or similar measurement question.

Questions address specific program or project priorities. For example, it may be important to track or evaluate cross-cutting dimensions of your program or project. Some stakeholders may be particularly interested in gender equality and women's empowerment. Measurement specialists and other stakeholders for mature programs may be more interested in longer-term outcomes and impact, while those of start-up projects will want to know if the intervention approach is working. The set of measurement questions defined for each program or project should reflect these specific interests.

The program ToC or project results chain will be the main reference frame for specific measurement questions. Questions can be asked about any level, and about relationships between the levels. However, some areas of the ToC may be more important than others in a specific context and time. The following are some examples:

- It is important to agree on the highest level at which to assess impact. For example: What has been the impact of the program on development of the market system or economic well-being?

- Some linkages may need more interrogation than others. For example: Did the regulatory improvements achieved through the program lead to expanded provision of appropriate financial services or increased take-up of those services?
- The relative effectiveness of different types of intervention may be of interest.¹¹ For example: Is stimulating the growth of a market for training providers a more effective use of program resources than direct training interventions? Is financial education provided by FSPs more sustainable than financial education provided by NGOs?

Outward and inward questions

In measurement, interventions tend to be at the center of the program environment. The need for accountability reinforces this. Most indicators and targets will be about what the interventions deliver and what effects they have had.

However, this tendency carries risks. It can lead to “intervention bias”—a tendency to overemphasize the influence of the intervention and underemphasize or ignore the contribution of other factors. It also reduces the chances that stakeholders will understand the dynamics of the market system: what is influencing what, and therefore, where should the program lend its support to be most effective at any time. This is particularly the case in financial systems because of their complex and changing dynamics.

Measurement efforts in financial inclusion need to ask questions about change from two perspectives. What were the effects of the interventions? This perspective provides an “outward” view the intervention. The other perspective is that of changes identified in areas of interest to the program or project stakeholders, without initial reference to the

interventions. How have these changes come about? What are the main factors that have contributed to them? This perspective provides an “inward” view from the changes in toward the interventions. Inward approaches should be based on a comprehensive understanding of the forces at work in the program environment.

Figures M4-1a and M4-1b show how the combined outward and inward approach to measurement questions for an intervention aiming at improving rules and norms is used. The example shows that inward questions are restricted to the middle zone of the ToC. At the development outcome level, in the context of whole or large sections of populations, there are so many potential factors at work that in most cases it would be unrealistic for a typical program measurement strategy to apply inward questions such as “What led to improvements in economic well-being?” This requires evaluative research on a different scale. On the other hand, at the lowest levels of the ToC, within the “program envelope,” contribution is more transparent and more easily mapped, so inward approaches are less relevant.

APPLYING MEASUREMENT QUESTIONS

Measurement questions are meant to steer measurement planning. They are a key first step in planning and shaping monitoring and evaluation because they clarify the questions they should help to answer. Together with the results framework, measurement questions determine the measurement strategy, which encompasses all aspects of the measurement system.

It is important to periodically review measurement questions and how to answer them, and to adjust the measurement plan where necessary.

FIGURE M4-1a. Outward and inward measurement questions based on the conceptual ToC

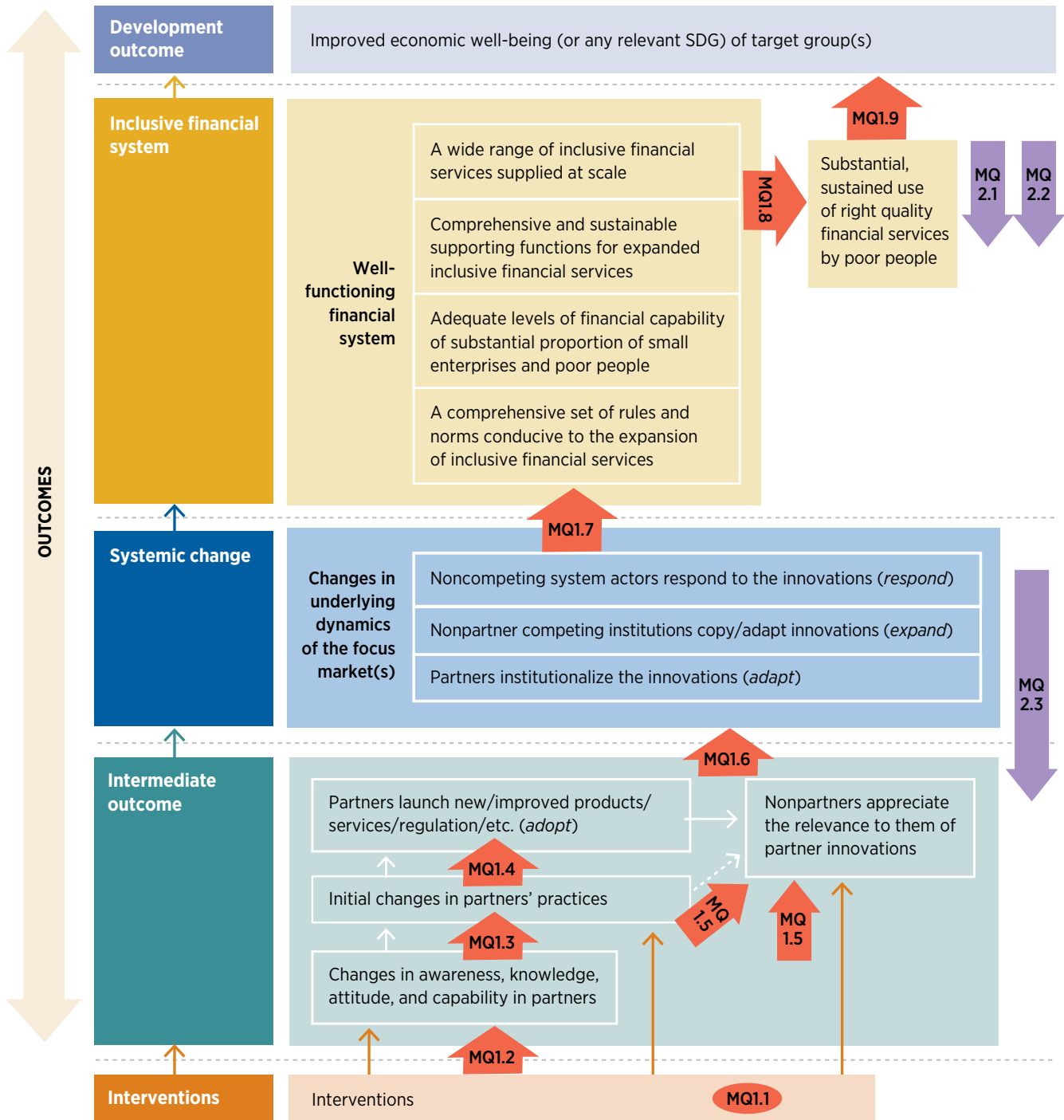


FIGURE M4-1b. Outward and inward measurement questions based on the conceptual ToC**OUTWARD** from interventions to financial inclusion
(and in some cases to development outcomes)

- MQ 1.1** Have the interventions been delivered efficiently, to the intended beneficiaries, and of sufficient scale and quality?
- MQ 1.2** To what extent and in what ways have the partners' awareness, attitudes, knowledge, and capacity been affected by the program interventions?
- MQ 1.3** To what extent and in what ways have the partners' practices been affected by the program interventions?
- MQ 1.4** Have the changes in the partners' practices supported by the program enabled the partners to produce new/improved products/services/regulation/etc. of the right quality?
- MQ 1.5** To what extent have program interventions and/or exposure to partner innovations promoted an appreciation among non-partners of the relevance for them of the innovations?
- MQ 1.6** To what extent and how have the intermediate outcomes of the program—including unplanned outcomes—led to partners adapting their innovations and nonpartners expanding, and responding to, them?
- MQ 1.7** To what extent and how has the early systemic change progressed and matured in the form of a well-functioning financial system?
- MQ 1.8** To what extent and how has the development of the financial system promoted financial inclusion (substantial, sustained use of right quality financial services by poor people)?
- MQ 1.9** To what extent and how has financial inclusion promoted development outcomes (such as improved economic well-being)?

INWARD from financial inclusion to systemic change

- MQ 2.1.** To what extent has there been substantial, sustained use of right-quality financial services by poor people; what have been the drivers and obstacles?
- MQ 2.2** To what extent and in what forms has there been development of a well-functioning financial system and what have been the drivers and obstacles?
- MQ 2.3** To what extent and in what forms has there been systemic change and what have been the drivers and obstacles?

**KEY TAKEAWAYS**

- Measurement questions, which are identified at the beginning of the program or project cycle, set the scope and focus of both monitoring and evaluation, to ensure measurement captures useful information for both learning and accountability.
- Measurement in systemic interventions needs to avoid intervention-centricity, by asking questions about change from two perspectives: starting from the intervention (outward questions—what are the effects of the intervention?) and starting from the identified change (inward questions—how has this change come about?).
- Measurement questions usually include generic questions, especially those related to OECD-DAC criteria, and complement these with specific questions relating to the program ToC or project results chain.
- As with other aspects of measurement, it is important to periodically review and revise measurement questions.

Module 4. References and Resources

IFAD (International Fund for Agricultural Development). 2002. "Deciding What to Monitor and Evaluate. Managing for Impact in Rural Development: A Guide for Project M&E." Rome: IFAD. <https://www.ifad.org/documents/10180/b916757f-3eff-45a6-80ca-3f3f53cd0e8f>

OECD (Organization for Economic Co-operation and Development). 1991. "The DAC Principles for the Evaluation of Development Assistance." Paris: OECD. <http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>

Preskill, Hallie, and Nathalie Jones. 2009. "A Practical Guide for Engaging Stakeholders in Developing Evaluation Questions" Princeton, N.J.: Robert Wood Johnson Foundation. <http://www.rwjf.org/en/library/research/2009/12/a-practical-guide-for-engaging-stakeholders-in-developing-evalua.html>

Notes

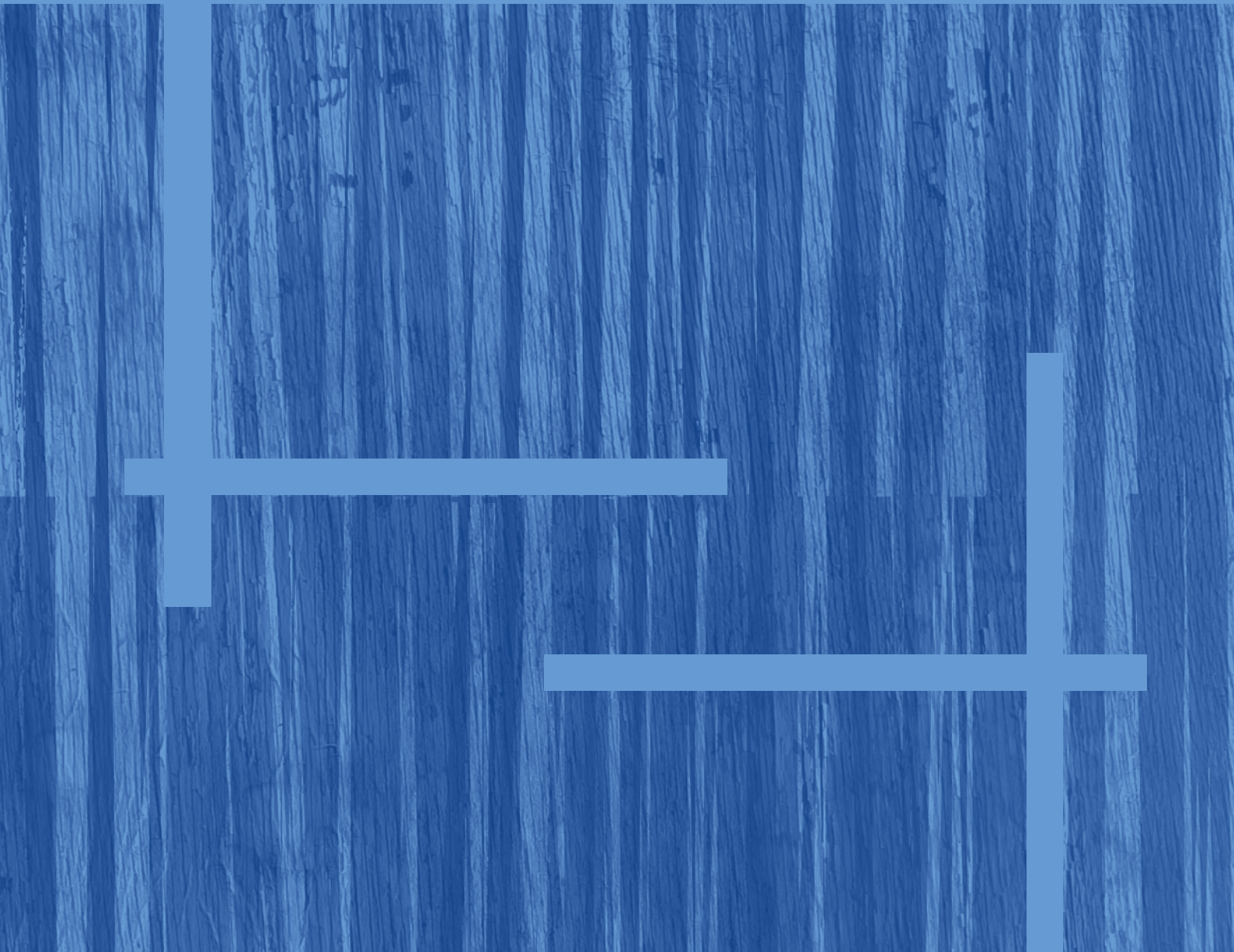
10. The broad OECD-DAC evaluation criteria—relevance, effectiveness, efficiency, impact, and sustainability—are likely to be appropriate for program-level evaluation, but impact and sustainability may not be relevant to all project evaluations or to monitoring, which tend to have shorter time horizons.
11. Comparisons between different types or areas of intervention may not yield firm conclusions, but they can uncover issues that need further exploration.

APPENDIX M4-A. Generic measurement questions

FOCUS OF QUESTIONS	FOCUS OF MEASUREMENT	EXAMPLES OF MEASUREMENT QUESTIONS
Interventions	Project monitoring	<ul style="list-style-type: none"> • Are these the right types and quantity of financial, human, and other resources? • Are the delivery partnerships functioning well?^a • To what extent are the interventions being carried out as planned in terms of timing and specification? If not, are there good reasons for deviation? What can be learned? • Are the intervention processes likely to generate ownership among the target institutions and groups? • To what extent are the right customer segments and organizations being reached and on a sufficient scale? • What is the quality of the interventions in terms e.g., of their relevance and usability for their intended target institutions and groups? • Are the interventions being delivered at the right cost? • If interventions are below expectations, why? • What needs to be done to improve or adapt the interventions going forward?
Intermediate outcomes	Project and program monitoring and evaluation	<ul style="list-style-type: none"> • To what extent is expected progress being made with planned outcomes (i.e., those defined by the results framework)? To what extent have the planned outcomes been achieved? • What significant changes have taken place in the program environment—other than the planned outcomes—to which the intervention may have contributed? • What would have happened if there had been no interventions? (counterfactual question) • To what extent and in what ways are program partners progressing beyond the program outcomes in contributing to market systems development? • How sustainable do the positive outcomes (planned and unplanned) appear to be? • How and why did these outcomes happen? Why did expected progress or outcomes not happen? • In what ways did the intervention contribute to what happened? In what ways did other factors contribute? • Are the assumptions about external factors affecting the interventions proving to be reliable? Have the assumptions stood up in practice? What significant changes have happened in the program environment that may have affected the program or may in the future? • What should the projects do—or have done—differently to be more effective?
Systemic change	Program and project monitoring and evaluation	<ul style="list-style-type: none"> • What is the evidence of systemic change? <ul style="list-style-type: none"> – To what extent and how have partners institutionalized the innovations? – To what extent and how have nonpartner competing actors copied or adapted the innovations? – To what extent and how have nonpartner noncompeting actors responded to the innovations? • To what extent has the program or project interventions contributed to the changes? What other factors have influenced and constrained systemic change? • What should the projects do—or have done—differently to be more effective?
Inclusive financial systems	Program monitoring and evaluation	<ul style="list-style-type: none"> • What changes happened at the financial-systems level—including the status of financial inclusion—beyond the intermediate and systemic change outcomes? • What was the contribution of the lower-level outcomes and the interventions? What were the contributions of other factors? • What does the status of financial systems outcomes tell us about where and how we should target interventions going forward?
Development outcomes	Program evaluation ^b	<ul style="list-style-type: none"> • What development outcomes in the program environment might have been influenced by financial inclusion? • What were the contributions of financial inclusion to those outcomes? • What do the findings about the relationship between financial inclusion and development outcomes tell us about where and how we should target interventions going forward?
Relevance and coherence	Project and program evaluation	<ul style="list-style-type: none"> • Was the program or project the right thing to do? Did it have any significant negative effects? • Were the parties involved the right ones to have undertaken it? Could it have been done better by other parties? • Is this type of program or project still needed? If so, which parties are best placed to take it forward?

a. A delivery partner is an organization or person that contributes to the interventions with, e.g., funding, expertise, logistics, or actual implementation.

b. It may not be possible to answer questions at this level through every program evaluation.



RESULTS FRAMEWORKS



THIS MODULE . . .

- Describes results frameworks and how they link to ToCs and results chains.
- Assesses the strengths and weaknesses of results frameworks.
- Puts forward proposals for the effective use of results frameworks.
- Explains the importance of managing assumptions and risks, and presents a model risk register that complements a results framework.
- Presents an example of a completed program results framework and an indicator profile.

WHAT ARE RESULTS FRAMEWORKS?

A results framework summarizes key measurement information. The most common form of a results framework is the logical framework—or logframe—which has been widely used in development planning and measurement for over 40 years. The terminology, precise format, and scope of results frameworks vary to some extent, but they have principal purposes and core components in common. They are used at both program and project levels.¹²

Results frameworks were originally introduced to bring systematic and logical thinking into formulating and planning programs. Although results frameworks still do this, ToCs and results chains have largely taken over this

function because they present a more intuitive way to identify the key contributions and results of a program or project and to align and link them logically.

Results frameworks link the definition of results with their measurement strategies by identifying indicators and means of obtaining data. Results frameworks at the program level are used mainly for accountability to funders and for overseeing measurement at the program level. Project results frameworks are mainly used for accountability to program management and for managing measurement in the project. The results framework is not in itself a comprehensive measurement strategy or plan; it is a core component of these instruments.

RESULTS FRAMEWORK COMPONENTS

A results framework is usually constructed as a matrix and invariably includes the following two components:

1. **A summary of planned interventions and their intended outcomes directly relating to the ToC or results chain.** The left-hand column of the matrix is often referred to as the “narrative summary.” For many years, logical frameworks used standard terminology and number of levels for the narrative summary: a single “goal,” a single “purpose,” and a limited number of outputs and activities. This is no longer the case; users tailor the narrative summary to their needs. Complex, long-term programs and projects in financial inclusion, in particular, need more levels and the flexibility to have as many results at each level as require measurement.
2. **Measurement variables, such as indicators, baselines, projections (often known as milestones and targets), and data sources or “means of verification.”** Most results frameworks have similar approaches to these measurement variables. A few basic rules should be followed.
 - Indicator development is often an iterative process and requires time.
 - Measurement variables should be aligned with the intervention or outcome to which they relate.
 - Baselines should be captured as early as possible, once indicators are agreed on.
 - Milestone projections do not have to be expressed in the same terms as in the end-line target. For example, a mid-term projection for an indicator like “number of new financial services” could be a stage in product development rather than several new services (which is likely to be zero).
 - Projections should not be firmly defined until enough is known to make an informed prediction.

Most results frameworks also include a column for assumptions or risks relating to interventions and outcomes, taking their lead from the ToC or results chain. This can be used to monitor assumptions, which is an essential aspect of enhanced monitoring (see Module 7). A more practical tool for comprehensively managing risks is a risk register. Risk registers are discussed in this module, and an example is included in Appendix M5-C.

With projects using systemic approaches, it is important to monitor beyond the results framework indicators. A column can be included in the project-level results framework for pointers to where monitoring beyond the indicators might focus. This would not be relevant at the program level, where this type of monitoring does not take place.

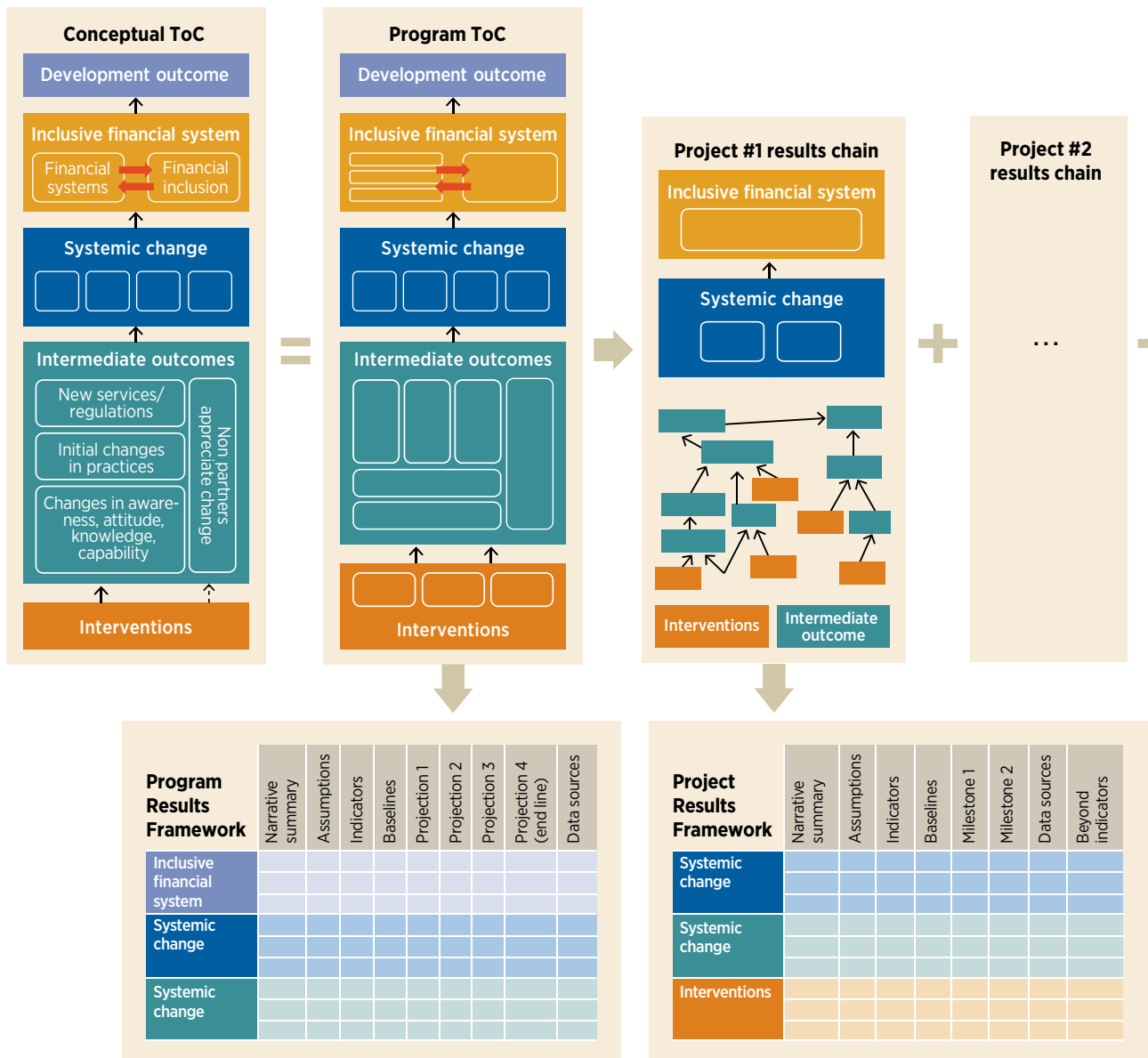
The results framework should reflect the program ToC or project results chain. Figure M5-1 shows the recommended relationships among ToCs, results chains, and their respective results frameworks. As noted in Module 3, project results chains may include multiple steps, particularly for interventions and intermediate outcomes. The user may need to select fewer key steps when translating the chain into the framework.

Example of a program results framework

Appendix M5-A contains an illustration of a program results framework relating to the program ToC in Module 3. Like the ToC, it is not put forward as a model, but simply as an illustration of what a program-level results framework could look like. Points to note about the example include the following:

- In this program, the highest level to be monitored systematically to assess contribution is the inclusive financial system. No indicators are included at the development outcome level. Data on changes in development outcomes will be reviewed in the planned end-term evaluation and principally to inform recommendations for future programming.
- Program-influenced change in an inclusive financial system is not expected to be significant until late in the program, so projections are not included until the final year when they will be assessed by the end-term evaluation. Similarly, projections for systemic change are not included for the first four years of the program. Any evidence of systemic change in earlier years will be captured through monitoring and will be reported (see Module 8).
- No intermediate outcome projections are made until the end of Year 3. This is because the intermediate outcomes in the program results framework example are focused on those expected to be achieved at or near the end of the projects, the first of which will be approaching completion at this point. This does not mean that evidence of intermediate outcomes will not be captured and reported in earlier years.

FIGURE M5-1. Relationship between ToCs, results chains, and results frameworks



- At the intervention level, because there is little or no scope for meaningful common indicators across projects, performance is assessed using a rating scale at each level with several set criteria. These assessments are made during annual reviews and contribute to a progress “dashboard.”
- In the example, only one indicator is given for each result for the sake of brevity. Outcomes often have two or more indicators.
- No actual milestone data or projections are given in the example.

Indicator profiles

A results framework usually contains only a summary of the information necessary for effective measurement. This information needs

to be elaborated to ensure that the data needs of indicators are unambiguous and that the data are collected reliably and on time. This is usually done in one of two ways. One approach is to produce an expanded version of the results framework, to provide the necessary guidance for meeting the data requirements of the indicators. This often ends up being a cumbersome document. An alternative that promotes ownership by the persons who are monitoring the indicator is to create a separate microplan or “profile” for each indicator. These should be developed as part of the measurement strategy (for a program) or plan (for a project) (see Module 2).

The expanded results framework or the indicator profile should include the following:

- Additional information about the indicators to ensure they are clear and unambiguous.
- Sufficient detail about data source(s) and the instruments (e.g., surveys) and processes for collecting the data.
- Responsibilities and timelines for collecting, analyzing, and reporting the data.
- Risks to data collection and any mitigating steps that need to be taken.
- Costs, particularly if surveys are involved.

Appendix M5-B provides an example of an indicator profile that complements the results framework for a client-centered innovation program.

STRENGTHS AND WEAKNESSES OF RESULTS FRAMEWORKS

Using and managing results frameworks requires appreciation of their benefits, an understanding of their limitations, and the mitigation of the risks associated with them. Results frameworks have been criticized for their linear, intervention-centric, and restricted view of outcomes. Results frameworks can also turn into binding and inflexible contracts that restrict measurement and create obstacles to adaptation. These risks are a particular concern for measurement in systemic programs because of their complex, dynamic, and unpredictable environments.

Appreciating the benefits

Results frameworks define, in one instrument, the expected results of a program or project, their hypothesized relationships, and key fea-

tures of how the results should be measured. Bringing this information together and aligning it can facilitate dialogue and understanding in the late stages of planning, and go on to serve as a reference for measurement during and after an intervention. The fact that results frameworks with similar formats are so widely used also means that stakeholders can readily interpret what they see across different projects, programs, and organizations.

Understanding the limitations and mitigating the risks

Be realistic about the intervention's field of influence. ToCs in financial inclusion present hypotheses about how change happens up to and including the development outcome level. Because results frameworks are primarily measurement instruments, it can be misleading to include indicators at levels beyond what is likely to be the measurable influence of the interventions.

Adopt a mixed basket of indicators to measure progress beyond financial services access and use. Results frameworks in financial inclusion programs and projects typically emphasize “snapshots” of inclusion indicators that are linked to access and use. Results frameworks should complement these types of indicators with others that capture evidence of systemic change and financial system development. (See Module 6 for more on indicators and data sources.) This will also help to make it less tempting to take shortcuts with interventions to maximize inclusion results (e.g., number of new accounts) that may look promising at first, but turn out to be unsustainable.

Why are results frameworks criticized?

There are two main concerns about results frameworks:

1. Like results chains and most ToCs, results frameworks are, by definition, intervention-centric, linear, and deterministic. They present a simplified view of a project or program environment, which implies that the intervention is the principal contributor to the changes and that cause and effect are unidirectional. It confines external factors to the assumptions column. These issues are more acute in the results framework because they are often used as a type of contract between funder and program manager.
2. Indicators often present only a partial view of the results in the framework, particularly if they are exclusively quantitative. This can overshadow less concrete, but nevertheless important, aspects of the outcomes in question.

However, these concerns need not outweigh their benefits as long as the concerns are recognized and managed.

Monitor results beyond indicators. Indicators tend to become the principal, often the sole, focus of attention for monitoring. This may limit the scope of monitoring and reduce the potential for learning and adapting within the program. Monitoring and evaluation needs to be open to aspects of intended results that are not captured by indicators and unplanned outcomes beyond the results framework. See Module 7 for information on how this can work in practice.

Monitor and manage assumptions. Although assumptions are usually included in results frameworks, they are often not diligently monitored and managed. Completing the assumptions column of a results framework is often viewed as a compliance exercise. It is important to monitor aspects of the program environment that are not in the intervention's direct field of influence, but that could affect the intervention and its outcomes. Although it is not possible to monitor everything in the program environment, an important place to start is the set of identified risks. The list of assumptions or risks should be regularly reviewed and updated as necessary.

Do periodic reality checks and adjust results frameworks, as needed. The narrative summary of a results framework—like the ToC or results chain—needs to be thought of as a sequence of hypotheses that are designed to be tested through measurement processes. This has three main implications:

- Measurement processes need to test if these hypotheses hold in practice.
- The same processes need to consider explanations for outcomes beyond the effects of the interventions (see modules 4 and 8).
- The stakeholders for the results framework need to be open to changes that reflect the “discovered reality” of the program and its environment.

A broader model of accountability

Results, indicators, and targets must be kept realistic and up-to-date. If they are not, and they remain in the contract between funder and program unit or between program units and project teams, they can impede adaptation and innovation. In the broader model of accountability, funders give program units appropriate levels of autonomy to design and adapt their interventions as required, while the units put in place comprehensive measurement and learning systems to account for their overarching objectives.

This is not to say that program units should change the content of results frameworks at will. There should be clear processes and criteria for change. Funders should fully consider the rationales for changes, not least because of the mutual understanding and learning that this can bring.

RISK MANAGEMENT

Risks affect programs and projects from start to finish. The success of interventions and the realization of intended outcomes depend on internal and external factors, which inevitably carry risks. Although risks are encountered mostly at the project level, program management needs to oversee how the projects identify and manage risks.

Risks to interventions include (*i*) internal risks (e.g., to the proper functioning of the project, such as funding or availability of competent personnel) and (*ii*) exogenous risks from outside actors and factors (such as obstacles created by governments, cultural barriers, threats to security, or macroeconomic volatility that can delay or frustrate interventions).

Risks to outcomes are mostly exogenous—and more so at higher levels in the results chain. All outcomes depend on external actors and factors. The launch of a new product by a partner financial services providers or a new regulation by a partner government agency, for example, depends on a high degree of autonomous action by the partner and support from the program and project. Outcomes also depend on a favorable operating environment. The higher up the results chain, the greater the dependency on external actors and factors to achieve those outcomes.

All programs and projects need to manage risks systematically. Most results frameworks include a column for assumptions or risks relating to interventions and outcomes. However, a more practical vehicle for risk management is a risk register. A risk register comprises risks that need to be managed. These risks are identified by assessing both the likelihood of the risk materializing and the impact on the project or program if it does. The major risks to interventions should be identified as early as possible when the project is being formulated. Risks that are assessed to be significant but manageable should be included in the risk register. Risks identified as significant and unmanageable—killer risks—should prompt some degree of project redesign or a decision to change the nature of the program or project altogether to avoid the risk. For example, it may

be decided to not run a project in a country where it is unlikely to make sufficient gains because of security or political concerns. Lower-level risks should be included in the register if they may require action, for example, creating a contingency plan to mitigate the risks before they have a chance to materialize. An example of a risk register is included in Appendix M5-C. The example is populated with some intervention and outcome risks that might arise in the

client-centered innovation program envisioned in the results framework in Appendix M5-A.

All significant assumptions and risks at every level—activity, output, or outcome—need to be monitored. This is part of enhanced monitoring, which is described in Module 7. However, there are limits to the risks to outcomes that can be practically managed through mitigation. Risk registers may not extend above the intermediate outcome level.



KEY TAKEAWAYS

1. **Results frameworks have an important role in presenting core measurement information for both programs and projects in a convenient and easy-to-understand format. They are mostly used for accountability and managing measurement.**
2. **It is important to understand the limitations of results frameworks, and to actively manage the risks associated with them. For example**
 - Be realistic about the intervention’s influence when agreeing on the highest level of outcome that attracts indicators in the results framework.
 - Track a basket of indicators—both qualitative and quantitative—to measure progress in market systems development.
 - Extend measurement beyond indicators in results frameworks, which may be only the “tip of the iceberg.”
 - Do not make the results frameworks the only element in accountability.
 - Regularly review and, when necessary, revise results frameworks to avoid misaligned incentives and wasting resources used to collect irrelevant data.
3. **A results framework is an integral part of any measurement strategy or plan, but it needs to be complemented by more detailed tools for managing indicators, such as the “indicator profile” presented in Appendix M5-B.**
4. **Risks to planned interventions and outcomes are inherent in systemic financial inclusion programs and projects. Actively monitor and manage key risks (using a risk register).**

Module 5. References and Resources

- Bakewell, Oliver, and Anne Garbutt. 2005. “The Use and Abuse of the Logical Framework Approach.” Stockholm: SIDA. <http://www.mande.co.uk/docs/LFA%20Review%20final.doc>
- Glaister, Leslie, and Nathalie Holvoet. 2013. “Feeling for the Intangible: A Framework for Donor’s Monitoring and Evaluation of Capacity Development Interventions.” Antwerp: Institute of Development Policy and Management, University of Antwerp. <https://www.uantwerpen.be/images/uantwerpen/container2143/files/Publications/WP/2013/05-Leslie-Holvoet.pdf>
- Itad (Information, Training and Development). 2012. “GEMS Results Measurement Handbook.” Brighton, United Kingdom: Itad. <http://itad.com/reports/gems-results-measurement-handbook/>
- Jensen, Greta. 2013. “The Logical Framework Approach.” London: Bond. <https://www.bond.org.uk/data/files/resources/49/The-logical-framework-approach-How-To-guide-December-2013.pdf>

- Wanitphon, Phitcha, and Alexandra Miehlsbradt. 2015. “Developing a Program Specific Monitoring and Results Measurement Manual.” Cambridge, United Kingdom: DCED. http://www.enterpris-development.org/wp-content/uploads/RMCase7_MRM_Manualb.pdf
- World Bank. 2012. “Designing a Results Framework for Achieving Results: A How-To Guide.” Washington, D.C.: World Bank. http://siteresources.worldbank.org/EXTEVACAPDEV/Resources/designing_results_framework.pdf

Note

12. In this Handbook, the term “program” is used to designate a coherent set of projects that together help to influence common high-level outcomes. A project is a tightly focused set of interventions with a common work plan. A program usually has a longer life than any single project within it.

APPENDIX M5-A. Example of a partially completed results framework for a client-centered innovation program in financial inclusion

NARRATIVE SUMMARY	CRITICAL ASSUMPTIONS (behind achievement of the outcome at the same level)	INDICATORS	BASELINES AND PROJECTIONS					DATA SOURCES FOR INDICATORS
			Baseline	Projection Year 3	Projection Year 4	Projection Year 5	Projection Year 6	
Level: Development outcome								
Improved economic well-being of poor people		The development outcome level will not be systematically monitored in this program						
Level: Inclusive financial system								
An inclusive financial system	Targeted populations have adequate access to relevant services	# available financial services in each country, in each category (i.e., savings, credit, insurance, and payments) that meet criteria for relevance						Program surveys of FSPs' service range
	Targeted populations have sufficient information about services and financial capability	# suppliers of relevant information services available in each country						Program survey of information service suppliers Assessment of services by FSPs
	Political commitment to create an enabling environment is maintained	% of client-centric FSPs in each country whose information needs are met by affordable information services						Program surveys of FSPs
		# suppliers of relevant central bank services in each country						Program survey of capacity-building services Assessment of services by FSPs
		% of FSPs in each country who perceive (i) statutory and (ii) voluntary regulation to be enabling client centricity in service development						Program surveys of FSPs
		# poor people who report frequency and perceived value, using different types of financial services						Where available, financial inclusion insights surveys supplemented by program-initiated financial diaries by target group samples
Level: Systemic change								
Regulators and financial services industry associations revise regulation and codes to encourage client-centered innovation	Political will is maintained Regulators and industry associations have the capacity to revise regulation and codes	% regulators and relevant industry associations undertaking relevant regulation and code revision						Program surveys of regulators and industry associations

NARRATIVE SUMMARY	CRITICAL ASSUMPTIONS (behind achievement of the outcome at the same level)	INDICATORS	BASELINES AND PROJECTIONS					DATA SOURCES FOR INDICATORS
			Projection Baseline	Projection Year 3	Projection Year 4	Projection Year 5	Projection Year 6	
<i>Level: Systemic change, continued</i>								
Nonpartner FSPs adjust their business models, or enter the market, and embed practices to be more client centered	Nonpartner FSPs have the incentives, capability, and resources to copy or adapt	# and % of targeted nonpartner FSPs adjusting their business models, or entering the market, and embedding practices to be more client centered						Triangulation of informal surveys, networking information, media references
Partner FSPs adapt and broaden their services offer for poor people	Services launched with program support are successful in the markets	% partner FSPs adapting and/or broadening their services offer after disengagement from direct program support						Program follow-up surveys of partners
Partner capacity building and information services providers adapt and broaden their services offer relevant to client centrality	Services launched with program support are successful in the markets	% of partner (i) capacity building and (ii) information services providers adapting and/or broadening their services offer after disengagement from direct program support						Program follow-up surveys of partners
Nonpartner information services and capacity-building providers adjust their business models, or enter the market, to develop services relevant to client centrality	Nonpartner providers have the incentives, capability, and resources to develop the services	# and % of targeted nonpartner information services and capacity-building providers adjusting their business models or entering the market to provide services relevant to client centrality						Triangulation of informal surveys, networking information, media references
<i>Level: Intermediate outcomes</i>								
Policy makers and financial services industry associations commit to revising regulation and voluntary codes to enable client-centered innovation in financial services	Interest at the political level is maintained	Evidence of commitment to revising regulation in each country						Collection of qualitative data through engagement with political processes by projects
New and/or improved, affordable, client-related information services are launched by partners	Political regimes continue to permit data gathering	# of new and/or improved, relevant, information services launched and assessed as relevant and affordable by FSPs						Collection of quantitative and qualitative data by projects through engagement with partner capacity building and information services providers. Assessment by a panel of FSPs (partners and nonpartners).

APPENDIX M5-A. continued

NARRATIVE SUMMARY	CRITICAL ASSUMPTIONS (behind achievement of the outcome at the same level)	INDICATORS	BASELINES AND PROJECTIONS					DATA SOURCES FOR INDICATORS
			Projection Baseline	Projection Year 3	Projection Year 4	Projection Year 5	Projection Year 6	
Level: Intermediate outcomes, continued								
Partner capacity building providers launch new and/or improved, affordable services supporting client-centered innovation	Commitment sustained by capacity-building providers throughout the services development process despite lack of FSP demand	# of new and/or improved, relevant, capacity-building services launched and assessed as relevant and affordable by FSPs						
Partner FSPs embed client-centered practices and launch new and/or improved financial services for poor people	Buy-in by FSP management and governors Commitment sustained by FSPs throughout the service development process despite lack of early-stage market data	# new and/or improved services launched by partner FSPs that have been developed using client-centric approaches and aimed at poor people						Collection of quantitative and qualitative data by projects through engagement with partner FSPs
Nonpartner organizations appreciate the relevance to them of investing in, supporting, or enabling client-centered approaches	Receptiveness of nonpartner organizations to change Visible and credible partner progress	% of nonpartner organizations (in each category and in each market) engaged by the program or directly by FSPs that indicate (through their survey responses) that they appreciate the relevance of client-centered approaches						Project surveys of nonpartner organizations
Level: Interventions			Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Yr 6
Program successfully implemented		% of projects on track or better in achieving results						Annual program reviews of projects

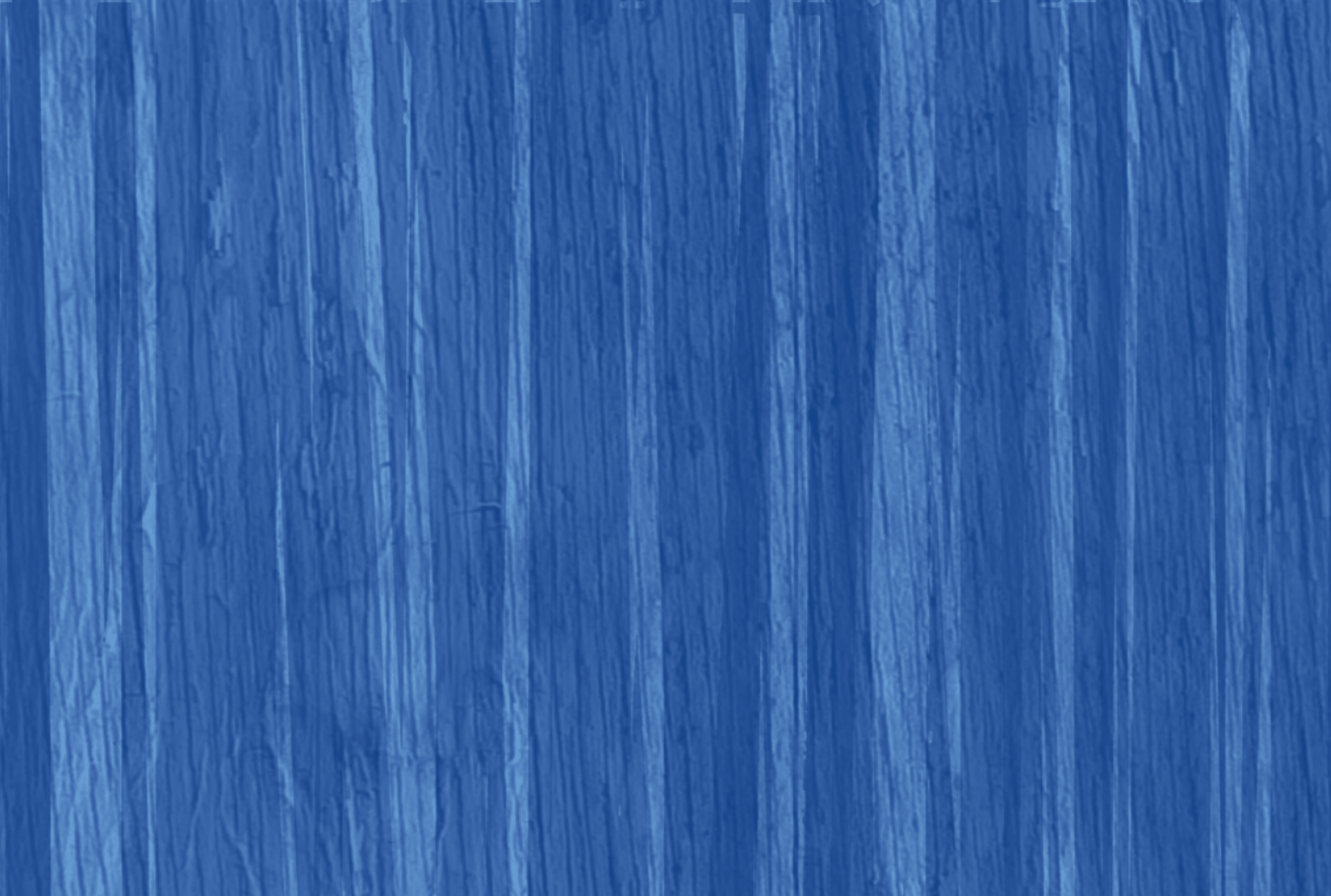
APPENDIX M5-B. Example of an indicator profile for a client-centered innovation program

INDICATOR TITLE	% of partner (i) capacity building and (ii) information services providers adapting and/or broadening their services offers after disengagement from direct program support
Outcome or intervention to which it relates	Systemic change outcome: Partner capacity building and information services providers adapt and broaden their services offers relevant to client centricity
Indicator owner (person reporting on this indicator)	Program measurement team leader
Indicator architecture, including definition of terms	A partner is an organization that received support from a project worth at least US\$100,000. To be included in the count, a partner would have invested over US\$50,000 (equivalent) in services development to reconfigure one or more of the services created during the project and/or to add a new service to its range.
Detailed description of the means of collecting data for the indicator	Semi-structured interviews will be conducted in person with the partner CEOs or heads of services development. They will be asked to show or at least describe their current business plans.
Schedule for collecting data for the indicator, including baseline	Baseline data will be collected by project teams on a rolling basis within two months of disengaging with partners. Progress data will be collected in Quarter 4 of program Year 5 for partners who have disengaged and in Year 6 during the summative evaluation scheduled for quarters 3 and 4.
Schedule for reporting the indicator	End Year 5; in line with summative evaluation reporting
Person(s) collecting data	In Year 5 data will be collected by projects (project managers will be accountable for ensuring this is done). In Year 6, the program measurement team leader needs to ensure this is done during the evaluation.
Anticipated costs of each round of data collection	Baseline data costs are subsumed in project management. Year 5 survey costs estimated to be US\$2,000 per project (included in project budgets). Year 6 survey costs estimated to be US\$9,000 (included in evaluation budget).
Baseline (copy and paste from results framework) with explanation if it is zero or N/A	To be determined. (Baseline data to be collected within two months of disengaging with partners.)
Projections (copy and paste from results framework)	To be determined in Year 4 (depends on progress in the projects)
Issues (e.g., risks to reliable data collection and strategies for mitigation or contingency)	Partners may be reluctant to reveal their business plans in detail. They should be assured about strict confidentiality. If this is not sufficient, a judgment will be made as to whether the information available is sufficient to include or exclude them from the count.

APPENDIX M5-C. Program risk register partial example

RISK CATEGORY	RISK FOCUS	RISK DESCRIPTION	INITIAL RISK RATING	PLANNED MITIGATION	RESIDUAL RISK RATING	CONTINGENCY
Interventions	Projects perform in line with expectations	Insufficient number of committed partners found to develop momentum in the projects	Likely: M Impact: H	Decision to launch a project will be based on an extensive diagnostic of partner commitment. Partners will be required to invest in services development before receiving extensive support.	Likely: L/M Impact: H	Decision to cease support and redeploy funding to another market to be made after 18 months
		Projects do not provide adequate reporting of their progress	Likely: H Impact: M	M&E capacity support from program in project inception period	Likely: M Impact: M	More intense oversight of those projects that still give rise to concern
Outcomes	Partner FSPs embed client-centered practices and launch new and/or improved financial services for poor people	Commitment is not sustained by some FSPs throughout the service development process despite lack of market data in the early stages	Likely: M Impact: M	Plan for early soft launch of a service to demonstrate potential	Likely: L/M Impact: M	Stop funding support to the partner(s) in question and look to move funding to more promising partners
	Regulators and financial services industry associations revise regulation and codes to encourage client-centered innovation	Regulators and industry associations do not have the capacity to revise regulation and codes	Likely: M Impact: H	Network regulators and/or financial services industry associations with counterparts in at least one other country who have successfully revised regulation and/or codes	Likely: M Impact: H	Consider co-funding technical support in phase 2
Resourcing—financial, human, material						
Fiduciary						
Reputation						
Data security						
Security and health of personnel						

Note: The cells above have been left blank deliberately because this is only a partially completed register for illustration only.



INDICATORS AND DATA SOURCES



THIS MODULE . . .

- Examines the roles and limitations of indicators and targets (projections).
- Proposes focus areas, where indicators are needed at different levels of a conceptual ToC.
- Categorizes types of data sources.
- Analyzes data sourcing opportunities and challenges in systemic financial inclusion projects.

THE ROLE OF INDICATORS

Indicators make interventions and expected outcomes measurable. If indicator measurement is needed more than once over time, or across interventions, and consistent data collection is feasible, indicators also enable comparisons. They also can be used as platforms for projections.¹³ Indicators and projections help funders and program managers monitor planned performance; they serve as the principal focus for evaluative processes.

Because indicators tend to be tangible. They often drive measurement systems, thereby overshadowing or crowding out less mechanistic means of capturing progress and achievement, which are also needed in complex intervention environments. These less mechanistic alternatives are described in modules 7 and 9.

Certain features of systemic approaches have implications for indicators and their use in measurement systems:

- Because systemic financial inclusion programs have long ToCs, indicators need to be identified at all measurable levels so that progress can be effectively monitored.
- Market systems development, particularly at lower levels in the ToC, implies changes in the awareness of key actors and the practices and shape of institutions. These changes need to be expressed qualitatively, through descriptors.
- Indicators need to be regularly reviewed to ensure they are still relevant and have kept pace with adaptations in the project or program. Projections should be thought of as hypotheses and should be reviewed regularly and revised, if necessary.

INDICATORS IN PRACTICE

Indicators need to represent, as closely as possible, the results that the interventions are expected to contribute to. At lower levels in project results chain, they need to be context-specific. At higher levels, there is more scope for standardizing indicators across projects in a program or even across different funder programs, as long as there is sufficient convergence around what needs to be measured.

With these qualifications in mind, tables M6-1 through M6-4 propose focus areas for indicators, using the outcomes in the Module 3 conceptual ToC and typical data sources. These are illustrations and are not intended to be an exhaustive list.

1. Development outcomes

There are several possible development outcomes, depending on the focus of the program. Improved economic well-being is usually a central objective of financial inclusion programs, although other aspects of well-being, such as empowerment, health, and education, can also be explicit either as a direct or indirect result of financial inclusion. Economic well-being can take the form of increased income, increased personal consumption (generally or relating to, e.g., education or nutrition) and reduced vulnerability to shocks and variation in income. If programs and projects include development outcomes in their ToCs and results chains, specific aspects of well-being—and their target groups—should be incorporated.

TABLE M6-1. Development outcomes—Indicator focus areas and data sources

INDICATOR FOCUS AREAS (<i>illustrative</i>)	TYPICAL DATA SOURCES
Development outcome: Improved economic well-being	
<ul style="list-style-type: none"> • Levels of household and/or enterprise asset ownership • Patterns of income-generation trials • Patterns of employment income • Consumption patterns • Degree of consumption smoothing in response to a negative shock or to variable earnings • Degree of protection of savings from external pressure and temptation • Capacity for long-term financial planning and risk management 	<ul style="list-style-type: none"> • National and international surveys and statistics • Program-led sample surveys of clients • Client longitudinal surveys, diaries, etc., for controlled trials

2. Inclusive financial system

The inclusive financial system level contains outcomes in terms of both financial system development and financial inclusion. Measuring progress in financial inclusion requires both quantitative and qualitative evidence of the use of financial services by underserved people. Table M6-2 proposes areas of focus for indicators and typical data sources.

Measuring financial system development requires an explicit description of what is expected to change. As discussed in Module 3, the program ToC should build on the conceptual ToC and provide more details on the targeted components of the financial system. Changes can be expected at various levels:

- Supply (e.g., a significant increase in the scale of reach and diversity of financial services provided).
- Demand (e.g., customers are more able to engage constructively with the services on offer).
- Supporting functions (e.g., increased quality, quantity, or sustainability of certain supporting functions such as market coordination, payment infrastructure, information, or capacity building services).
- Rules and norms (e.g., improved regulatory environment).

TABLE M6-2. Inclusive financial system—Indicator focus areas and data sources

INDICATOR FOCUS AREAS (<i>illustrative</i>)	TYPICAL DATA SOURCES
Financial inclusion (substantial, sustained use of right-quality financial services by poor people)	
<ul style="list-style-type: none"> • Number and % of users, with various breakdowns: <ul style="list-style-type: none"> – Clients segments (types of people or households, enterprises) – Types of financial services (savings, credit, insurance, payments, etc.) – Duration – Regularity (e.g., more frequent use of savings accounts and digital payments than insurance and credit) • Attitudes of users and nonusers 	<ul style="list-style-type: none"> • Third-party supply- and demand-side surveys (e.g., Finscope, Findex, Financial Inclusion Insight Survey) and national statistics • Program-led sample surveys of FSPs, clients, and nonclients
Financial system development (well-functioning financial system)	
<p>Supply side</p> <ul style="list-style-type: none"> • Number of services providers, in specified categories (e.g., type, scale, length of time in market) or specified types of services • Proximity of specified types of services to potential customers • Extent of choice available to specified categories of customers • Costs of specified types of services • Responses of providers to market opportunities and setbacks • Degree of customer centricity of providers' business models • Degree of interoperability of points of service • Number and/or % of specified types of services meeting given quality criteria • Client satisfaction with specified types of services 	<ul style="list-style-type: none"> • Third-party supply- and demand-side surveys (Finscope, Findex, Financial Inclusion Insight Survey, etc.) and national statistics • Program-led sample surveys of FSPs, clients, and nonclients
<p>Demand side</p> <ul style="list-style-type: none"> • Customer understanding of how to access and use specified types of services • Customer awareness of the range of services • Loan default rates among specified categories of customers 	
<p>Supporting functions</p> <ul style="list-style-type: none"> • Infrastructure • Access to and use of independent points of sale by targeted customer segments; number and type of FSPs using independent infrastructure—and cost of use; volumes of payments transactions due to interoperability; degree of competition in provision of infrastructure • Sustainability of the infrastructure, user (both FSP and customer), satisfaction with infrastructure • Other supporting functions' (information, capacity building, market coordination, product development, capital markets, etc.) relevance and comprehensiveness, access and use by FSPs or clients, user (FSP and customer) satisfaction, affordability (where relevant) 	
<p>Rules and norms</p> <ul style="list-style-type: none"> • Comprehensiveness and relevance of policy, regulation, and normative frameworks • Responsiveness of guardians of policy, regulation, and normative frameworks to perceived new needs • Quality of supervision frameworks • Extent of enforcement • Extent of buy-in among services providers regarding voluntary industry codes • Extent to which FSPs and other market actors face regulatory barriers to entry or have regulatory incentives to enter the market • Perceptions of policy, regulation, and normative frameworks among stakeholders affected • Inclusiveness of policy-making and regulatory processes (e.g., are stakeholders consulted?) • Extent and depth of cooperation and collaboration among market actors 	<ul style="list-style-type: none"> • Public record documents relating to policy, regulation, and normative frameworks • Program-led surveys of rule-setters • Rule-setters' own monitoring • Third-party or program-led surveys of stakeholders (e.g., FSPs, micro and small enterprises) • Third-party demand-side surveys

3. Systemic change

The systemic-change level captures changes in the dynamics of the market systems in focus, particularly the behavior of the actors beyond the direct influence of the program interventions. (See Table M6-3.)

TABLE M6-3. Systemic change—Indicator focus areas and data sources

INDICATOR FOCUS AREAS (<i>illustrative</i>)	TYPICAL DATA SOURCES
Partners institutionalize the innovations that were fostered by the program (<i>adapt</i>)	
<ul style="list-style-type: none"> • Extent and scope of partners' continuous improvement of policies and practices • Partners' use of customer feedback and other data in decision-making • Partners' use of diagnostics and other approaches to scaling up and developing new services • Partners' governance stakeholders' approaches to innovation and risk • Partners' engagement with risk identification, monitoring, and management • Partners' responses to shocks • Partners' investment in human capacity building • Partners' networking for knowledge and ideas 	<ul style="list-style-type: none"> • Structured and semi-structured interviews with principal partner interlocutors • Partner qualitative self-monitoring (e.g., with journals) • Partner meeting reports, MOUs, budgets, and other records that capture decisions and changes • Structured organizational assessment • Program staff or third-party observation
Nonpartner competing actors copy or adapt partner innovations (<i>expand</i>)	
<p><i>Extent to which nonpartners . . .</i></p> <ul style="list-style-type: none"> • Develop or adopt similar practices • Acquire technology similar to that of partners' • Develop similar services • Engage more actively with target segments or with actions in relation to target segments of population 	<ul style="list-style-type: none"> • Third-party industry-level surveys • Program-led interviews with nonpartner institutions
Nonpartner noncompeting system actors respond to partner innovations (<i>respond</i>)	
<ul style="list-style-type: none"> • Extent to which noncompeting market actors adjust or develop new services and/or regulations, etc., in response to innovation by partners and expansion by nonpartners 	<ul style="list-style-type: none"> • Third-party industry-level surveys • Program-led interviews with nonpartner institutions

4. Intermediate outcomes

Intermediate outcomes mostly relate to the partner institutions, such as FSPs, infrastructure providers, industry bodies, and government agencies. Inasmuch as there have been interventions to stimulate crowding-in and scaling-up of the program innovation, intermediate outcomes should extend to those actors as well.

- Among partners, this level of the ToC captures (*i*) initial changes triggered by the interventions (changes in mindset and capability, commitments to change direction); (*ii*) initial changes in practices, processes, and structures that build on the new mindset or capability; and (*iii*) the adoption of these practices by partners, illustrated by upgrading an existing product, service, or regulation or developing a new one. Changes in awareness, attitude, knowledge, and capability can be measured directly through questionnaires and testing, but a more practical and reliable approach is to look for evidence of these changes in actual behaviors. Table M6-4 suggests focus areas for indicators to track these changes.
- Intermediate outcomes also capture initial outcomes beyond partners, that is, the initial steps to the crowding-in and/or scaling up phase. At this level, outcomes are expected to show that nonpartners appreciate the relevance of the change at the partner level to their own strategy and/or operations.

TABLE M6-4. Intermediate outcomes—Indicator focus areas and data sources

INDICATOR FOCUS AREAS (for illustration only)	TYPICAL DATA SOURCES
Changes in awareness, knowledge, attitude, and capability in partners	
<ul style="list-style-type: none"> • Partner staff awareness, knowledge, and attitudes • Confidence expressed by partner staff in their own and their organization's capability • Application of new knowledge and skills • Extent to which discussion in partners' governance forums reflects interest in engaging with greater financial inclusion • Extent, scope, and nature of partner engagement in dialogue with other actors about financial inclusion • Public declarations of intent by partners 	<ul style="list-style-type: none"> • Structured and semi-structured interviews with principal partner interlocutors • KAP surveys of wider groups of partner staff • Partner governance and management meeting reports, capturing of dialogue • Program staff observation • Public media output
Initial changes in partners' practices	
<ul style="list-style-type: none"> • Create budgets for developing services for underserved segments • Adjust strategies and/or structures to focus on developing services for underserved segments • Obtain and analyze data on underserved segments • Engage with internal capacity assessments related to developing new services • Procure external services aimed at advances in financial inclusion 	<ul style="list-style-type: none"> • Structured and semi-structured interviews with principal partner interlocutors • KAP surveys of wider groups of partner staff • Partner meeting reports, MOUs, budgets, and other records that capture decisions and changes • Partner qualitative self-monitoring (e.g., with journals) • Program staff or third-party observation • Public media output
Partners launch new, improved products, services, regulation, etc. (adopt)	
<p>Supply: New or improved inclusive financial services provided by partner institutions</p> <ul style="list-style-type: none"> • Number, type, and specific features (e.g., cost, transparency, flexibility) of partners' services targeting underserved segments • Accessibility of partners' services by underserved segments • Efforts made by partners to develop customer awareness and understanding of new services • Take-up of new partners' services by targeted segments • Client satisfaction with new partner services <p>Supporting functions: New or improved services provided by the partner institutions</p> <ul style="list-style-type: none"> • Types and specific features (e.g., scope, robustness, ease of use, customer affordability, transparency) • Customer reach • Affordability • Use by targeted clients and user satisfaction <p>Rules and norms: New or improved rule or norm or process by partner institutions</p> <ul style="list-style-type: none"> • Types and specific features (e.g., scope, transparency, relevance to underserved segments, and the providers of services targeting them) of new or improved partner-generated policy, regulation, and normative frameworks • Capacity of actors to apply, supervise, or enforce the policy, 	<ul style="list-style-type: none"> • Structured and semi-structured interviews with principal partner interlocutors • Partner qualitative self-monitoring (e.g., with journals) • Partner meeting reports, MOUs, budgets, and other records that capture decisions and changes • Partner business data (for private-sector partners) • Public record of changes in policy, regulation, and normative frameworks (for program working on rules and norms) • Client, user, stakeholder surveys
Nonpartner institutions appreciate the relevance to them of partner innovations	
<ul style="list-style-type: none"> • Engagement by nonpartner institutions with partners' business models for underserved segments • Competing nonpartner institutions identify the case for copying or adapting the innovations • Noncompeting market actors identify the case for modifying their own practices in response to the innovations 	<ul style="list-style-type: none"> • Third-party, industry-level surveys • Program-led interviews with nonpartner institutions

Note: KAP = Knowledge, attitude, and practice; MOUs = memoranda of understanding

DATA SOURCES

Sourcing data for measuring financial inclusion programs presents few conceptual but many practical difficulties. Data are needed in three stages: before the program or project begins, during the program or project, and after it has been completed.

Before the interventions begin. Data are needed before interventions begin, both to diagnose market barriers and opportunities and to establish baselines for future measurement of change. Data might be needed on any of the following:

- Actual and potential client perceptions of financial services: access, use, affordability, risks, cultural fit, capacity, and information deficits.
- Current status and potential for advancing financial services provision: number and types of providers and of services, extent and nature of outreach and use of services, types of customers, services provider knowledge of (and attitudes toward) the low-income market segment, types of services, providers' strategies, and openness to innovation, capacity, and information deficits.
- Current status and potential for developing supporting infrastructure and services: number and types of providers and of services, extent and nature of use, types of services, providers' strategies and openness to innovation, capacity and information deficits, and availability of technology.
- Current structure and application of policy, regulation, and other rules and norms. Attitudes and capacity and information deficits among policy makers, regulators, and other rule or norm shapers and guardians.

Data collection at this stage should be influenced by a measurement strategy, including the program ToC, if one exists. (See Module 2 for a description of measurement strategies.) Diagnostics at such an early stage (program or project formulation) are not strictly measurement activities; however, data collected at this point can be used for measurement baselines if data are relevant and up-to-date—and complemented by indicator-specific baseline data, once the results framework has been created.

During the program or project. Data are needed during the program or project to monitor intervention performance and critical aspects of the environment. Data collection

is guided by the measurement plan,¹⁴ including the measurement questions that apply to monitoring (see Module 3). Data for systemic financial inclusion programs may be needed on the following:

- Indicators in the program or project results framework, at the intervention, intermediate, and systemic change levels. Initially, baseline data may be needed, particularly for outcomes, and this should be collected as early as possible in the intervention, if not before. Baseline data are the main references for setting targets and for assessing change.
- Aspects of planned results not adequately captured by indicators. These data are often qualitative.
- Unplanned or unexpected outcomes, both positive and negative.
- Reasons why planned interventions and outcomes have or have not been achieved, including the contributions of actors outside the interventions.¹⁵
- Pre-identified assumptions and risks and other relevant dynamics beyond the field of direct influence.

At the end of or after the program or project.

Data are needed at the end of, or after, the program or project, for the summative and impact evaluation, which is usually conducted externally (see Module 8). At this stage, it is often necessary to collect additional data to fill gaps in the monitoring data, to complement it through triangulation and to assess contribution. Data collection is normally directed by a detailed evaluation framework, led by the broad measurement questions set at the beginning of the program cycle, and further specified through judgment criteria, indicators, and other means of guiding data collection. Data at this stage might be needed on the following:

- Results framework indicators at systemic change outcome level—if not already or adequately collected—and at the financial system and, where relevant, development outcome levels.
- Perspectives on planned results not captured by indicators, for triangulation and deeper understanding. These data will be mainly qualitative.
- Unplanned or unexpected outcomes, both positive and negative.
- How and why outcomes have or have not been achieved, including the role of actors outside the intervention.

Criteria for data collection. Even if there are appropriate data sources, choosing the right data to collect and on the right scale can be challenging. Organizations can end up drowning in data and/or being faced with incomplete or poor-quality data when it is too late to improve data quality.

Successful data sourcing and management is often guided by use cases—that is, the specific potential use of the data. Relevance, intention, and incentive to use are important criteria for data collection. However, it is also important to be aware of other strengths and limitations of data sources. The following is a suggested checklist of criteria for data sources.

Practicability

The data source exists and is accessible. Access is affordable; the cost of access is proportionate to expected value.

Reliability

Trustworthy—data are sufficiently complete, representative, and free from unintentional bias or deliberate manipulation.

Consistent—data in the same categories can be collected every time they are needed.

Precision

The data sets are sufficiently granular (e.g., data are collected by target groups and locations of interest).

Timeliness

Data are available when needed and are sufficiently up to date.

Ethics

Data are collected ethically and sustainably, ensuring that human or institutional sources will be prepared to provide data in the future.

Data sources for each level of the ToC

The higher the level of the ToC, the more programs are likely to rely on secondary or shared data, although this should be triangulated where possible with program-generated data. At the lower levels of the ToC, interventions need to generate their own context-specific primary data. Data triangulation should be pursued when possible, especially when posing how and why questions. For market systems programs and projects, assessing changes in knowledge, attitude, capacity, and behavior (intermediate outcomes) is essential to understand what, how, and why the underlying dynamics and incentives are changing. Hence, primary—often qualitative—data are essential. Tables M6-1 through M6-4 illustrate how multiple data sources (qualitative and

quantitative as well as primary and secondary) can be used at different levels of the ToC.

DEMAND- AND SUPPLY-SIDE DATA SOURCES

When focusing on the intermediate outcome, systemic change, and financial system levels of the ToC, it is important to distinguish between supply- and demand-side data collected by third parties. Demand-side data originate from actual or potential users of financial services: individuals, households, enterprises. Supply-side data originate from services providers. Both vary in several ways that are important for financial inclusion program use, including frequency of collection and how representative they are.

These differences need to be known before deciding to choose them as data sources for indicators. Comparing the attributes of different data sources to understand their strengths and limitations can help guide you to the sources and indicators that are most relevant.

In recent years, data sources for financial inclusion have become richer and more complicated to navigate. There are numerous data sources on the demand side. Although having ample data to measure progress is an asset, each of these sources treats data in different ways, which can lead to confusing or misleading results. Depending on what data source you use, you can arrive at significantly different conclusions about financial inclusion. See Table M6-5 for the best-known demand-side data sources; more details are provided in Appendix M6-A. On the supply side, sources vary in scope and in survey attributes as shown in Table M6-6.

TIPS: COLLABORATE TO SOURCE DATA



Where there are no relevant data, it may be possible to collaborate with other funders, program units, and country stakeholders. This is not only important for funders' and program units' measurement needs, but a well-functioning financial system also relies on collecting, processing, and interpreting data. Funders can facilitate market development by collaborating with each other to develop sustainable data sources at the industry or country level. Collaboration with other funders can help to develop local capacity, ownership, and a robust financial market data infrastructure.

TABLE M6-5. Demand-side data sources for financial inclusion

DEMAND-SIDE DATA SOURCES	DATA DESCRIPTION	QUANTITATIVE OR QUALITATIVE	LAUNCH AND FREQUENCY	INDIVIDUAL/HOUSEHOLD/FIRM	NUMBER OF COUNTRIES	SAMPLE SIZE	SUB-NATIONAL DATA	EXAMPLE INDICATORS OF "ACCESS" AND OTHERS
Global Index	Global trends and cross-country comparison	Qual	2012; 3 years	Individual	148	1,000+	No	<ul style="list-style-type: none"> • % holding an account at a formal FI • % obtaining a loan from an FI within the past year
FinScope	Individuals' financial mgmt; perceptions regarding financial services; formal and informal	Quant + Qual	2002; 4 years	Individual; some household conclusions	17	1,000–21,000	Regional	<ul style="list-style-type: none"> • % operating an account with an FI • # of adults holding an account with a SACCO
FinAccess/ Access to Financial Services Surveys	Individuals' financial mgmt; perceptions regarding financial services; formal and informal	Quant + Qual	2006 in Kenya; 3–4 years; varies by country	Individual; household	Kenya, Nigeria, others	8,250–20,850	Regional	<ul style="list-style-type: none"> • % formally included • % owning a mobile phone • % access by region, gender, education
Financial Inclusion Insight Surveys	Individual perception and behavior regarding formal digital financial services	Quant + Qual	2013–2015; yearly or twice/year	Individual	10 + add. countries likely to come soon	3,000–45,000	Regional	<ul style="list-style-type: none"> • % with a bank account • Average distance to closest banking facility • % with a mobile money acct
Financial Inclusion Tracker Surveys	Trends in households' financial behavior; trends in poverty levels of mobile money users	Quant + Qual	2012–2015; yearly + quarterly	Household (panel)	3 (Tanzania, Pakistan, Uganda)	3,000–5,000	Regional	<ul style="list-style-type: none"> • % of HHs owning or with access to a mobile phone • % of HHs with active SIM • % of mobile money accts/region
WB Financial Capability and Consumer Protection Surveys	Assessment of people's understanding of financial concepts	Qual	2008; varies	Individual	14	1,000–2,000	Urban/rural	<ul style="list-style-type: none"> • % who correctly answer financial knowledge questions • % who budget • % with formal credit
IFC Enterprise Finance Gap Database	Estimation of the number of MSMEs in the world and the degree of financial inclusion	Quant + Qual	2010; as data becomes available	Firm (panel) data for select countries	177	Varies, 1,000	No	<ul style="list-style-type: none"> • % of MSMEs with access to credit • % constrained by access to credit
Living Standards Measurement Study	Household welfare and behavior	Quant (limited) Qual	1985; varies	Household (panel)	38	800–36,000	Regional	<ul style="list-style-type: none"> • Average monthly expenditure by type and region • % of HHs operating nonfarm enterprises

Note: FI = financial institution; SACCO = Savings and credit cooperative organization

TABLE M6-6. Supply-side data source to measure financial inclusion—examples and features

SUPPLY-SIDE DATA SOURCES	INFORMATION SOURCE	FREQUENCY	NUMBER OF COUNTRIES	ANALYTICAL TOOLS	GEOSPATIAL SUBNATIONAL	EXAMPLE INDICATORS OF “ACCESS” AND OTHERS
IMF Financial Access Surveys	Central banks in 189 countries	Yearly, 2004–2012	189 jurisdictions	No	No	<ul style="list-style-type: none"> • Bank branches per 1000km² • ATMs per 100,000 adults
MIX Maps	MFIs; increasingly other financial services providers and regulators	Quarterly; depends on country and providers	15	Tableau visualization; benchmarking	Yes	<ul style="list-style-type: none"> • # of financial institutions, by type or region • Locations per 100,000 adults • Locations per 1000km²
fspmaps.com	Georeferencing providers in-country	1 round; sustainability plans under discussion	7	Analytical geospatial online tool	Yes	<ul style="list-style-type: none"> • # of poor people living within 5km of a financial access point (urban/rural divide) • Locations of financial access points
GSMA Mobile Money Adoption Survey	Mobile money providers; telcos	Since 2011, yearly	208 operators in 83 countries	Benchmarking	No	<ul style="list-style-type: none"> • # of registered MM users • # of active MM agents • Value and volume of transactions
Microinsurance Centre Landscape Studies	Landscape studies by the Munich Re Foundation, IADB, MFW4A, and GIZ	Launched in 2016; yearly	102	Benchmarking; online visualization	No	<ul style="list-style-type: none"> • # and % of people covered by different types of insurance
WB Remittance Prices Worldwide	Surveys; certified national and regional databases	Launched 2016; 2011 onwards available online	32 sending; 89 receiving	Benchmarking	No	<ul style="list-style-type: none"> • Average cost of sending or receiving remittances from specific countries • Cost of sending and receiving by institution type
World Bank’s Global Payment	Central banks	2 years (2008, 2010)	139	Benchmarking	No	<ul style="list-style-type: none"> • Volume of transactions • # of countries with consumer protection legislation
Mftransparency.org	Microloan providers	Launched 2006 yearly and as data become available (2008–2013)	26	Benchmarking and online graphing tool	No	<ul style="list-style-type: none"> • # of transparent products • % of products with one or more fees • APR by institution type

Note: MM = mobile money



KEY TAKEAWAYS

- Indicators can be used to evaluate outcomes and other subjects of interest. If measured more than once over time, or across interventions, indicators enable comparisons.
- Because they are tangible, indicators may be used to drive measurement systems. Although indicators may seem to offer a practical solution to the measurement needs of hard-pressed stakeholders, they can overshadow or crowd out less concrete means of capturing progress and achievement that may be more appropriate in complex intervention environments.
- Indicators need to represent the results that the intervention is expected to bring about or contribute to. At the systemic change level and below, they need to be intervention-specific. Qualitative indicators may be more appropriate than quantitative.
- There is some scope for standardizing indicators at higher levels in the ToC.
- Data are needed throughout the project or program lifecycle. This requirement can create technical, logistical, and budgetary challenges. It is important to focus on use cases—on relevance and a clear intention to use—when considering data collection and to apply a range of other criteria for data fitness.
- Different levels of the ToC or results chain—and therefore different stages in the program or project cycle, respectively—require different data sources and instruments. Qualitative data are more relevant at the lower levels of the ToC, while quantitative data feature more at higher levels where there is a greater focus on clients.
- Collaboration is needed to ensure consistent, good quality, data sourcing at the national and international levels for financial inclusion.

Module 6. References and Resources

- Church, Cheyanne, and Mark M. Rogers. 2006. “Designing for Results: Integrating Monitoring and Evaluation in Conflict Transformation Programs.” Washington D.C.: Search for Common Ground. http://www.dmeforpeace.org/sites/default/files/SFCG_Designing%20for%20Results_Ch4.pdf
- Kessler, Adam, and Jim Tanburn. 2014. “Why Evaluations Fail: The Importance of Good Monitoring.” Cambridge, United Kingdom: DCED. http://www.enterprise-development.org/wp-content/uploads/Why_Evaluations_Fail_Aug2014.pdf
- Sen, Nabanita, and Adam Kessler. 2015. “Guidelines to the DCED Standard for Results Measurement: Defining Indicators of Change.” Cambridge, United Kingdom: DCED. http://www.enterprise-development.org/wp-content/uploads/2_Implementation_Guidelines_Indicators_May_2015.pdf

Demand-side Data Sources

- FinAccess. <http://fsdkenya.org/publication/finaccess2016/>
- Financial Inclusion Insight Survey. <http://www.intermedia.org/data-home/>
- Financial Inclusion Tracker Survey. <http://www.audiencescapes.org/fits>
- FinScope. <http://www.finscope.co.za/pages/default.aspx>
- IFC Enterprise Development Database. <http://financegap.smefinanceforum.org/>
- Living Standards Measurement Survey. <http://iresearch.worldbank.org/lsmssurveyFinder.htm>
- World Bank Financial Capability and Consumer Protection Survey. <http://responsiblefinance.worldbank.org/surveys/users-of-financial-services>
- World Bank, Global Findex. 2015. “The Global Findex Database 2014, Measuring Financial Inclusion Around the World.” <http://www.worldbank.org/en/programs/globalindex>
- World Bank. Financial Capability and Consumer Protection Survey. <http://responsiblefinance.worldbank.org/surveys/users-of-financial-services>

Supply-side Data Sources

- Bank for International Settlements (BIS) Payment Systems Statistics. http://www.bis.org/list/cpss/tid_57/index.htm
- Finclusion Lab. MIX. <http://finclusionlab.org/>
- Fspmaps.com. <http://fspmaps.com/>
- GSMA Mobile Money Adoption Survey. <http://www.gsma.com/mobilefordevelopment/programmes/mobile-money-for-the-unbanked>
- International Monetary Fund. Financial Inclusion Access Survey. <http://fas.imf.org/>
- Mftransparency.org. <http://www.mftransparency.org/microfinance-pricing/>
- Microinsurance Centre Landscape studies. <http://www.microinsurancecentre.org/landscape-studies.html>
- World Bank. Global Payment Survey. <http://www.worldbank.org/en/topic/paymentsystems/remittances/brief/gpss>
- World Bank. Global Survey on Consumer Protection and Financial Literacy. <http://responsiblefinance.worldbank.org/surveys/providers-of-financial-services>
- World Bank. Remittance Prices Worldwide. <http://remittanceprices.worldbank.org/>

Notes

13. An indicator should be neutral and framed as, e.g., “the number of FSPs who launch new inclusive products,” rather than “x new firms launch new products” or “an increase of y% in the number of FSPs who launch new inclusive products.”
14. Measurement strategy is linked to the ToC; measurement plans are linked to the results chains and are more detailed. See Module 2.
15. There will be limits to how rigorously this can be done outside the scope of external evaluations.

APPENDIX M6-A. Selected demand and supply data sources

There is no single best source for data on financial inclusion. To find out what data you need, start by asking yourself what you want to know. You also need to ensure that the data are reliable and valid.

DEMAND-SIDE DATA SOURCES

1. **Global Findex** is the only global demand-side data source that allows for global and regional cross-country analysis. It includes data from 148 countries and collects information on 506 indicators from at least 1,000 individuals over 15 years old within each country. The sample is nationally representative and randomly selected. Since the survey is a module added to the Gallup World Poll, it combines information about sociodemographic conditions and access to or use of financial services. Global Findex is mainly used to analyze global trends and to make cross-country comparisons using key financial inclusion indicators, such as the number of adults with access to formal bank accounts. A drawback is that the data are not subnationally representative, which means that they are less useful for in-country policy makers because there is not sufficient granularity. Also, the definition of formal financial services used in this source is based on people's perception of whether their provider is a formal financial institution, which is not necessarily aligned with the regulatory and supervisory framework of a country. The sample is randomized at the individual level, which allows users to aggregate the data by individual characteristics, such as income and gender, but this also makes the data incompatible with household-level surveys.
2. **The FinScope Survey** is a globally recognized demand-side data source that allows for financial inclusion indicators to be measured at a sub-national level. It was launched in 2002 and is trademarked and owned by FinMark Trust. FinScope is a nationally representative survey that explains how individuals manage their financial lives. It also provides insight into attitudes and perceptions about financial products and services. Sample sizes vary widely across countries; surveys have included responses from 1,000 to 21,000 individuals. The unit of sampling is at the individual level, but the survey does enable some conclusions at the household level. Eighteen countries have conducted or are in the process of conducting FinScope surveys. Often industry players will pay some of the cost of the survey and help tailor the questionnaire to meet multiple stakeholders' needs. FinScope data are not comparable across countries on all indicators.
3. **FinAccess and Access to Financial Services** surveys are similar to FinScope surveys, but they are not conducted by FinMark Trust. To a large extent, FinAccess in Kenya and Access to Financial Services in Nigeria (and the like) follow the same principles as FinScope, but they are not conducted by nor owned by FinMark Trust. As with FinScope, these surveys are designed through industry consultation, which means they have the potential to meet many needs and answer many questions. They have some of the same drawbacks as FinScope: they are not designed for cross-country comparison. Because these surveys are commissioned and carried out by various entities, data quality may be inconsistent. In addition, across both FinScope and FinAccess surveys, there is no standard definition of "financially included," so the meaning behind this term varies according to the definitions used by local stakeholders. For instance, some countries might include payment cards or mobile wallets that are not linked to an account, while others may not.
4. **Financial Inclusion Tracker Surveys (FITS)** are a nationally representative panel survey designed to collect trend data about households' financial behavior over time. The Bill & Melinda Gates Foundation's Financial Services for the Poor team in partnership with InterMedia designed these surveys to run over a three-year period in three countries. The sample size is 3,000 households each in Uganda and Tanzania and 5,000 households in Pakistan. The survey will measure the same households throughout the entire period. Data from the survey represent collective behavior and use patterns for all members of a particular household. The data are used to estimate trends in poverty levels of mobile money users. This focus on households, while a useful perspective, can also be a shortcoming because it is not as helpful for analysis at the individual level. Furthermore, the survey has been carried out in only three countries.
5. **Financial Inclusion Insight (FII) Surveys** were launched by the Bill & Melinda Gates Foundation in partnership with InterMedia. FII data first became available in 2013. Unlike FITS, FII surveys are not panel surveys because they do not track the same household over time. They focus more on measuring individual perception and behavior, making them comparable to FinScope, FinAccess, and the like. However, their intentional focus on mobile money and digital financial services sets these surveys apart. The surveys are typically conducted

annually, with commitment from the Gates Foundation to fund the surveys in eight countries (Kenya, Tanzania, Uganda, Nigeria, India, Pakistan, Bangladesh, and Indonesia) over three years. CGAP has also engaged InterMedia to conduct FII surveys in Rwanda and Ghana (released in 2015) and Myanmar and Côte d'Ivoire (data collected 2016–2017). Furthermore, UNCDF's Mobile Money for the Poor program has hired InterMedia to conduct FII surveys in Benin and Senegal. The sample size is typically high to allow for subnational representation (even more so than, e.g., FinScope). The surveys include welfare measures based on the Grameen Progress Out of Poverty index, which is unique to FII surveys. Because FII surveys mainly focus on insights into digital financial services, they do not capture many indicators around access to and use of nondigital financial services.

SUPPLY-SIDE DATA SOURCES

- 1. The IMF Financial Access Survey (FAS)** is the most comprehensive global supply-side data on financial inclusion. In addition to providing policy makers and researchers with annual geographic and demographic data on access to basic consumer financial services worldwide, FAS is one of the main data sources for the G20 Basic Set of Financial Inclusion Indicators endorsed by the G20 leaders at the Los Cabos Summit in June 2012 as well as the more comprehensive G20 Financial Inclusion Indicators. The FAS database contains annual data for 189 jurisdictions, including all G20 economies, covering a 10-year period (2004–2014). Countries manage their data and metadata. FAS has the same functionality as Findex's cross-country comparison advantage for demand-side data, but for supply-side data. The IMF FAS is not subnationally representative, and the data depend on countries' ability to capture data from FSPs. Moreover, FAS includes data on prudentially regulated FSPs only.
- 2. GSMA Mobile Money Adoption Survey.** In 2011, Mobile Money for the Unbanked (MMU) initiated a global adoption survey to give managers of mobile money deployments better insights into the performance of their services relative to each other. Subsequently, GSMA created its Global Annual Mobile Money Adoption Survey—its annual results are published as part of the MMU Annual State of the Industry report. In the 2015 survey, 107 mobile money providers from 67 countries participated in the survey. While the database itself is not public, MMU publishes an analysis of the aggregated results. The survey offers a snapshot of the mobile money industry every year and provides benchmark data to mobile money services providers.
- 3. World Bank's Global Payments Systems Survey (GPSS).** GPSS is a comprehensive survey that is conducted in 139 countries. It provides information on the status of national payments and securities settlement systems worldwide. It covers all aspects of national payments systems—from the legal and regulatory environment, to infrastructure, technological and business model innovations, international remittances, and oversight framework—and combines quantitative and qualitative measures of payments system development. GPSS is expected to guide reform efforts in the payments system arena both nationally and globally. The 2008, 2010, and 2012 surveys provide a snapshot of the payments and securities settlement systems in both advanced and emerging economies. Data from the fourth and latest iteration of the survey were expected in late 2016.
- 4. MIX's Finclusion-Lab.** The MIX Market is the premier source of public information on microfinance institutions (MFIs) and their financial and social performance. MIX offers a suite of popular analysis reports at the global, regional, and country levels, including global analyses of key issues for the sector. MIX has expanded its focus through geospatial mapping efforts to aggregate and visualize data for FSPs, including and beyond MFIs. To date (2016), 23 countries have subnational data visualized through the portal. MIX's move to visualize geospatial subnational supply-side data through publicly available geospatial maps has enriched the supply-side data landscape. This will be a challenge because frequent data collection can be expensive and/or ad hoc, depending on when data become available.
- 5. Fspmaps.com** is a website funded by the Bill & Melinda Gates Foundation in partnership with Spatial Development International. Fspmaps.com provides analytical tools to answer several financial access questions. As with MIX, the website leverages geospatial information for financial inclusion tracking and analysis. It hosts comprehensive geospatially referenced financial access point data and high-resolution population data, including poverty densities and other demographic attributes. Its analytical tools can be used to obtain detailed information about where people—including poor people—live in relation to financial services access points. Underserved areas can be better detected this way. Another analytical tool allows you to drop a pin on a map and calculate population served with mobile coverage but without adequate financial access. The website also allows users to import private datasets. Fspmaps.com currently hosts data for Tanzania, Uganda, Nigeria, Kenya, Bangladesh, and India. For most countries, the data are snapshots in time and have not been collected over time.



ENHANCED MONITORING



THIS MODULE . . .

- Examines why traditional monitoring falls short of what is required for systemic financial inclusion programs.
- Describes the concept of “enhanced monitoring” and points to some of its practical applications.

This module will be useful to staff of program units; funders will also benefit from enhanced monitoring to promote a supportive environment. Enhanced monitoring has implications for resourcing and capacity building in measurement, and for how accountability is interpreted.

THE NEED FOR ENHANCED MONITORING

Traditionally, monitoring is used to track performances and to capture data for evaluations. Monitoring tends to be thought of as a narrow, technical process that is conducted in the margins of program and project management. This model is being challenged in many development fields, but practice has been slow to adapt to the change in thinking.

In systemic financial inclusion programs and projects, it is difficult to predict the opti-

mum implementation strategies and the precise nature and timing of intermediate outcomes. For example, certain FSPs or other institutional partners chosen at the outset of a project may turn out to be unprepared or not committed to stay the course in developing new financial services, support functions, or rules. There are several reasons for this, including change of leadership or policies. In these circumstances, staff need to adapt implementation strategies and expected intermediate outcomes to the new and changing circumstances.

The monitoring process is used to map the terrain and support the interventions’ navigation through it. The main challenge lies in the unpredictability of results. Traditional monitoring is inadequate and can yield misleading information because it is usually confined to predefined indicators and relies on structured surveys conducted at infrequent intervals for data on outcomes. The monitor-

ing process needs to have a broad scope, to be conducted more frequently, and to be integrated with evaluative internal review and adaptation processes:

- Outcomes should be regularly monitored, at least as far as the systemic change level in the ToC or results chains, through indicators, and more broadly, it should capture the narrative behind the indicators.
- Monitoring should be designed to be broad enough to capture unplanned or unexpected results and aspects of the environment that are affecting the program or project. These environmental concerns may have been identified as assumptions or risks or may be unexpected.
- Monitoring should be designed to have a broad perspective about what is driving results. It should not be intervention-centric; it should include both an inward and outward perspective (see Module 4).
- Regular evaluative reviews should be conducted. Reviews should focus on outcomes that are based on evidence. They should include partners, where possible and appropriate.
- Agreed processes and criteria need to be in place so that the project or program (and its measurement frameworks) can be adapted based on the reviews.

ENHANCED MONITORING IN DETAIL

Three characteristics of systemic financial inclusion programs and projects call for enhancements to monitoring and adjustments to its relationship with external evaluation. These characteristics and their implications for monitoring are set out in tables M7-1 through M7-3.

TABLE M7-1. Implications for monitoring—Complex environment

INTERVENTIONS OPERATE IN A COMPLEX ENVIRONMENT	IMPLICATIONS FOR MONITORING
<p>The complexity and unpredictability of financial markets mean that</p> <ul style="list-style-type: none"> • It is difficult to predefine results, indicators, and targets with confidence • Assumptions and risks relating to the external environment play a major role in intervention management 	<ul style="list-style-type: none"> • Results monitoring needs to be flexible. Predefined outcomes, indicators, and targets are important, but they need to be reviewed and adjusted or replaced if they turn out to be inappropriate. This is easier if they are not locked into an inflexible results framework. • Results monitoring needs to take a broad approach that extends beyond predefined indicators. Results-focused mission reports and logs or journals kept by partners and project staff are the most common tools for this type of monitoring. See Appendix M7-A for an example of a monitoring log. • The intervention impacts its external environment, but the environment also affects the intervention and its outcomes. Aspects of the external environment need to be monitored. This is chiefly done through identifying and monitoring assumptions behind, or risks to, the progression narrative of the ToC or results chain. However, it is not always possible to identify important external factors in advance, so the intervention's active set of assumptions and risks needs to be regularly reviewed and adjusted.

TABLE M7-2. Implications for monitoring—Change cannot always be quantified

CHANGE CANNOT ALWAYS BE QUANTIFIED	IMPLICATIONS FOR MONITORING
<p>Change at lower levels in the ToC and results chains tends to be about attitude, behavior, organizational processes, and structure. These changes can be nuanced and often refer to very small numbers of people and organizations.</p>	<ol style="list-style-type: none"> 1. The monitoring of intermediate outcomes and systemic change exclusively through quantitative indicators is likely to be inadequate and sometimes misleading. Useful data at these levels are often qualitative and highly context-specific. They often take the form of narrative descriptions, especially where there are few units of change (e.g., organizations). The two types of indicators need to be appropriately balanced. 2. Qualitative data pose challenges to the aggregation of results across programs or portfolios and comparative assessments between them. Unlike standardized quantitative indicators, there is no way to automatically aggregate and compare the results of interventions if they are expressed qualitatively. However, in some cases, these types of results can be categorized to enable comparisons through scoring systems (see Module 9).

TABLE M7-3. Implications for monitoring—Dynamic environment

INTERVENTIONS OPERATE IN A DYNAMIC ENVIRONMENT	IMPLICATIONS FOR MONITORING
<p>The intervention environment is dynamic, particularly where technology is involved, and our understanding of it needs to keep pace. Interventions need to enable rapid response to remain relevant.</p>	<ol style="list-style-type: none"> 1. Data need to be collected regularly and in a timely fashion. This is easier to do if the data are collected by the partners themselves or project staff on the ground, rather than by remotely located staff during field visits. Relying on an infrequent outcome survey is not enough. Data need to be regularly reviewed, and their implications for the intervention assessed. Reviews should be conducted with both internal and external key stakeholders, where appropriate. 2. Reviews should make use of data previously collected through monitoring, but can also be used to generate data themselves in real time by asking measurement questions during the review (e.g., about significant changes that participants have observed or experienced). Answers to these questions should be triangulated with data already recorded or should be followed up for corroboration. 3. Do not assume that changes identified can be attributed to the interventions. Program and project staff should try to identify plausible drivers of the change in question and examine the evidence to help determine which drivers may have been the most influential. This “inward” perspective in monitoring and internal review can realistically be done robustly only up to a certain level of the results chain—in some cases, no higher than the systemic change outcome level. The demands on data collection and analysis at higher levels may be too great. 4. Processes for making necessary adaptations to both the interventions and the measurement plans should be systematized. For example, this could be a mandatory item in review agendas.

HOW TO MAKE IT WORK

Enhanced monitoring needs legitimacy, time, resources, guidance, and support. Funders and senior program management need to create an enabling environment for enhanced monitoring.

The proposed enhancements to monitoring can be demanding. The following are some tips on tailoring processes and tools to fit the program or project and to manage measurement capacity of a unit’s staff.

Two overarching instruments promote effective enhanced monitoring: measurement questions and a project measurement plan. As Module 4 demonstrates, clear measurement questions at the beginning of the intervention cycle will set the agenda for both monitoring and evaluation. For example, questions about whether assumptions in the ToC and results chains have turned out to be reliable, and whether they were the right ones, are relevant for both monitoring and ex-post evaluation. To develop a project measurement plan, the details of what is to be monitored, how, when, by whom, and at what cost need to be worked out. The plan provides a realistic context and helps to systemize the approach (see Module 2). However, it is also important not to “over-engineer” the data collection tools and processes. If the processes are cumbersome and rigid, they will be seen as a bureaucratic chore.

Beyond processes, staff capacity is crucial. The organization should enable staff to identify evidence of change in a complex intervention. The measurement plan should be used to point to the types of data to look for and explain how they should be recorded. At the same time, staff should be encouraged and empowered to create and own a narrative that captures their experiences, observations, and partner feedback.

Enhanced monitoring, enabled by commitment from management and effective processes, should lead to improved implementation and results. It is likely to happen only if funders allow for flexibility in results frameworks and program implementation. Changing course, after due process, needs to be an acceptable option and encouraged.

Transitioning from traditional to enhanced monitoring

Transitioning from traditional to enhanced monitoring can be a big change for program unit or organization staff. With enhanced monitoring, more people—often most project team members, partners, and in some cases, their clients—are involved in collecting and reporting relevant data. Implementing enhanced monitoring requires substantial change to processes, appropriate expertise in design and coordination, and more financial resources than that of traditional monitoring. Staff may

find it difficult to move directly from traditional indicator-based monitoring to the new paradigm of enhanced monitoring. This section addresses managing this transition.

INTEGRATING MONITORING WITH EVALUATION

In an enhanced model, monitoring takes on a new relationship with evaluation. Enhanced monitoring is not a substitute for evaluation. It cannot achieve the scope or rigor of a well-designed and well-conducted evaluation. There is usually a place for evaluation in a project and invariably in a program.

Monitoring and evaluation can be integrated in three principal ways:

- Where evaluations are planned—whether internally or externally led—monitoring can support evaluations by ensuring that rele-

vant data have been collected and that a degree of analysis and interpretation has taken place. Identifying measurement questions at the beginning of the program or project can help.

- External evaluators can add value to measurement planning at the start of a program or project cycle. They can help to identify measurement questions, construct the results framework, identify areas for monitoring beyond indicators, build capacity, and steer baseline research. In some cases, external evaluators are retained to work as real-time or developmental evaluators.
- Evaluations should inform future monitoring in the same or similar programs by highlighting what needs to be monitored and how.

Evaluation and its relationship with monitoring is explored further in Module 9.



TIPS: TRANSITIONING TO ENHANCED MONITORING

Enhancing monitoring does not have to be implemented in one fell swoop. Instead of taking one big step, teams can move forward incrementally.

- The first step can be to improve reviews, so that they can be conducted frequently; include all necessary stakeholders; are supplied with adequate, well-presented data; and lead to actionable conclusions.
- Another step is to foster a sense of ownership of data collection. Ensuring that the staff who collect data are involved in interpreting the significance of the data provides an incentive for them to perform this function more effectively.
- Encouraging staff to broaden their field of vision in looking for evidence of change may be difficult at first, but the effort is rewarded with a richer engagement with data. For the project team, this may mean having closer and more regular contact with stakeholders in the field (partners, clients, and other market actors).
- Piloting enhanced monitoring in one or two projects and demonstrating how it can lead to improved implementation can drive buy-in, because staff will be able to see the positive effects of their efforts.



KEY TAKEAWAYS

- It is widely recognized that “traditional” monitoring, which is based exclusively on results framework indicators, is inadequate and can be inappropriate for programs that apply a systemic approach to financial inclusion and other sectors.
- Indicator monitoring needs to be complemented by a wider range of data about planned and unplanned results.
- The external environment will affect the intervention and its outcomes both positively and negatively, predictably and unpredictably. These external factors need to be monitored mainly through the lens of assumptions and risks.
- Systemic financial inclusion programs need to be adaptable. Data collected through monitoring need to be regularly reviewed, and the conclusions of the reviews should lead, through due process, to adaptations to both the program or project itself and its results framework. During the review process, it is important to be clear about what is driving the identified changes: the extent to which the changes were driven by the intervention and other players and factors.
- Monitoring and external evaluation have different roles, but there is a middle ground where monitoring staff and evaluation staff need to collaborate. Measurement strategies (Module 2) need to identify how the roles are different and what aspects call for collaboration.
- Monitoring enhancements amount to substantial organizational change. Staff involved in the enhanced monitoring should have incentives and support.

Module 7. References and Resources

- Dunn, Elizabeth, et al. 2014. “Monitoring Facilitation Activities: Report to USAID Uganda.” Washington, D.C.: USAID. http://pdf.usaid.gov/pdf_docs/PBAAC328.pdf
- Hummelbrunner, Richard, and Heather Britt. 2014. “Synchronizing Monitoring with the Pace of Change in Complexity.” Washington, D.C.: USAID. <file:///C:/Users/patri/Downloads/Pace%20of%20Change%20in%20Complexity%202014-09-25%20FINAL.pdf>
- Perrin, Burt. 2012. “Linking Monitoring and Evaluation to Impact Evaluation.” New York: Rockefeller Foundation. <https://www.interaction.org/sites/default/files/Linking%20Monitoring%20and%20Evaluation%20to%20Impact%20Evaluation.pdf>
- Ripley, Matthew, and Daniel Nippard. 2014. “Making Sense of Messiness: Monitoring and Measuring Change in Market Systems: A Practitioner’s Perspective.” Durham, United Kingdom: Samarth-NMDP and The Springfield Centre. <http://www.enterprise-development.org/wp-content/uploads/Samarth-Bangkok-5Mar14.pdf>
- Samarth-NMDP. 2015. “Results Measurement System: User Manual.” Lalitpur, Nepal: Samarth-NMDP. <http://samarth-nepal.com/resource/results-measurement-system-user-manual>
- USAID (United States Agency for International Development). 2016. “Program Cycle Guidance. Complexity-Aware Monitoring.” Washington, D.C.: USAID. <https://usaidlearninglab.org/library/complexity-aware-monitoring-discussion-note-brief>

APPENDIX M7-A. Example of a project monitoring log

The log is designed to capture observations—first hand or from reliable secondary sources—about:

1. Project interventions.
2. Changes in the project environment (outcomes) that appear to be influenced by project interventions, whether planned or unplanned.
3. Other (contextual) factors in the project environment that have had, or are likely to have, a bearing on the project and its outcomes.

Ref #	Short account of what has been observed (provide hyperlink or other reference to any sources with fuller information). Explain what happened, to or with whom, and when. If it is an outcome, explain why you believe it was influenced by a project intervention and which one. Note also if any action is needed to corroborate the link with an intervention.	How, when, and where was this information captured?	Relevance of the observation: <ul style="list-style-type: none"> • If it relates to an intervention or outcome (and, where relevant, indicator) in the results framework, specify which. • If it is about a contextual factor, indicate which pre-identified assumption or risk (if any) it relates to. (If the observation implies a heightened risk factor, it should be followed up with risk management.)
1	Example: We received feedback from Paul FF that the coaching he received so far is not practical. He finds it to be too theoretical. He is not asking for the coach to be replaced but for us to discuss alternative approaches with her. A recording of the call can be accessed at. . . . (Paul was aware that the call was being recorded.)	Phone call from Paul on 25.04.16	Intervention: “Coaching on client centricity to strategic managers in partner FSPs.” Indicator: Satisfaction rating.
2	Example: We learned that one of our FSP partners, PPP Bank, has signed a contract with TTT Insight to conduct qualitative research among savings groups in three counties in western Kenya. This is its first venture into this type of research and is likely to have been triggered by the participation of its marketing director in our Nairobi workshop on client research in January 2016. (We will need to follow up with the manager to confirm this.)	We were informed by TTT Insight in a meeting at its office on 04.05.16.	Intermediate outcome: Partners recruit new staff and make changes in processes to embrace client centricity.
3	Example: We have become aware that Miriam LL, the board member of HHH Bank most committed to promoting client centricity in the bank, has left. This could represent a risk to board-level commitment to the continued process.	HHH Bank website seen 11.05.16	Intermediate outcome (assumption): Commitment sustained by FSP management and governance throughout the services development process.

REPORTING AND KNOWLEDGE SHARING



THIS MODULE . . .

- Describes the principal roles of reporting in systemic financial inclusion programs and how reporting and knowledge sharing can be captured in one report.
- Discusses the inter-related issues of frequency and scope and/or content of reporting.
- Provides an example of a program annual report template.

REPORTING ROLES

Reporting is a critical element to achieving both accountability and learning. It also tends to focus on results expressed through indicators and targets. Reporting usually follows a chain that begins with projects and moves through to programs, funders, and finally to funders' governance stakeholders. To streamline the performance information passing along this chain, funders often want programs to use common indicators so that key information can be aggregated.

With financial inclusion, as with other interventions in complex systems, there is limited scope for common indicators, particularly at the lower levels of the ToC and results chains. This makes it difficult to aggregate information gathered along the chain, and alternative approaches to accountability reporting may be

needed. Reporting needs to be carefully crafted around key stakeholders' agreed vision of accountability. This might include, for example, the obligation to demonstrate through reporting the effective management of risk and of adaptation and the credibility of qualitative results. The key to successful reporting is engaging with stakeholders to determine their needs and expectations for reporting.

Reporting of systemic approaches to financial inclusion programming needs to be strongly knowledge-oriented. Measurement in these programs is designed to inform funders and program units about whether the program is on track and whether the track is the right one—and when and how to change course if it appears not to be. With systemic approaches, funders and other stakeholders have a strong interest in this broader program perspective, which can be gained through reporting.

Integrating reporting into internal review

Module 2, Figure M2-1, demonstrates how reporting and knowledge sharing should flow directly from data analysis, review, and evaluation. Reporting consumes a lot of time and other resources. Reporting that is integrated with review and evaluation is both more efficient and more effective than when it is a detached process designed solely for accountability. Review and evaluation can transform data into knowledge, which can be harnessed in reporting.

Knowledge-oriented reporting can lead to learning for all program stakeholders up and down the reporting chain. Meaningful reporting to direct stakeholders can stimulate interest in the program, promote more constructive engagement, and trigger valuable feedback that can be used to improve and adapt the program. Furthermore, well-presented, outcome-rich reporting can generate demand for more such information, thus creating incentives to further improve the measurement system.

Broader information and knowledge sharing

Figure M2-1 demonstrates that review and evaluation can produce knowledge that is of

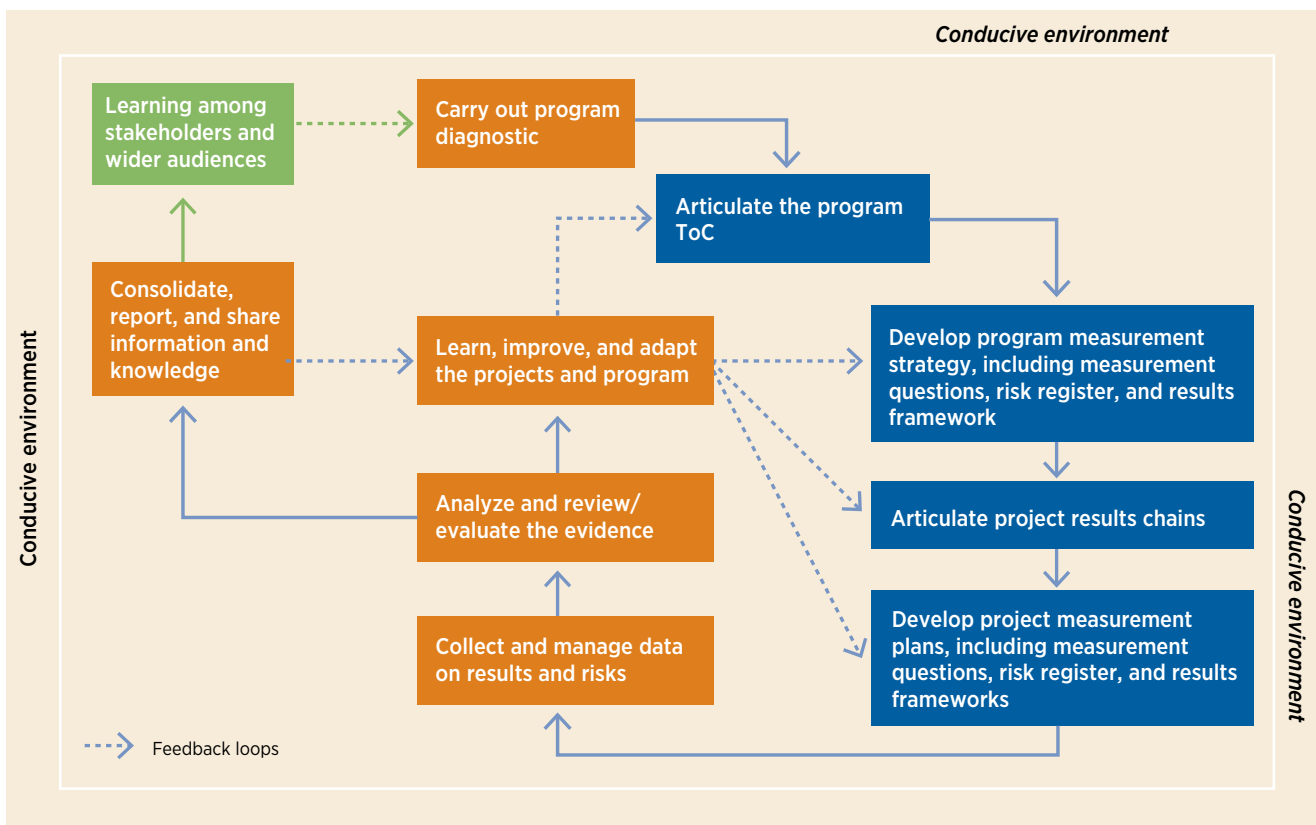
interest to broad audiences, such as managers and other staff of institutions that have similar or complementary objectives, beyond a program’s immediate stakeholders. Formal reports often are not the most engaging media for wider audiences. Audiences may be more receptive to relevant narratives in other formats, such as blogs, live broadcasting, webinars, or face-to-face workshops.

Frequency of reporting

Frequency of reporting depends primarily on how useful the reports are to recipients—the level and the nature of the interventions. Most systemic financial inclusion programs span a long period, and substantive change at the intermediate level does not usually happen quickly or often. Many funders do not have the time to make sense of detailed quarterly or even semi-annual reports. An exception would be if the program is operating in a high-risk environment, such as a fragile or conflict-affected state, in which case the funder would have an interest and a duty to be kept informed about the risk landscape.

Project staff usually report to program unit staff more frequently than program staff report to funders because the capacity for en-

FIGURE M2-1 (repeated from Module 2). Measurement System Components and Cycle



gagement and granularity of interest between the former is usually greater than that of the latter. Under average risk conditions, detailed project reports should be generated for program stakeholders once a year; less detailed reports may be submitted mid-year. Apart from the value to the recipients, semi-annual reporting should create incentives for project staff to conduct in-depth reviews, but only if the reviews are integrated with reporting—that is, timed to add value to the reports.

Reporting from program staff to funders could follow the same timetable as reporting of projects to program unit staff. Preferably, the timing of the reporting will allow for project reporting to inform program reporting. However, funders that do not have the capacity to peruse two reports a year should not ask for them. Funders may find it more useful to engage with program staff and project staff between annual reports through face-to-face or virtual meetings that involve a less formal dialogue about what is working, what isn't, and what lessons can be drawn.

THE SCOPE AND CONTENT OF REPORTS

Module 4 shows how the different information and knowledge needs and interests of a program's stakeholders can be expressed through measurement questions. These questions not only help to shape the measurement conducted during a program, they also indicate the scope and focus of reporting to these stakeholders.

The core of a project or program report invariably focuses on the results framework. This applies irrespective of the frequency of the report, although reports that are more frequent than annual usually focus mainly or exclusively on interventions rather than outcomes. The interventions and outcomes in the results framework, nuanced and calibrated by indicators and projections, are what the project or program is intended to achieve.

Reporting against projections takes place only when projections are scheduled. In systemic financial inclusion programs where significant change takes time, projections are usually not scheduled every year. Between projections, progress is still reported against the interventions and outcomes in the results framework. These reports are guided by, but are not necessarily limited in scope to, the indicators.

At the program level, indicators are often broad and lack context. Information from the projects in program reports should illustrate and expand on the basic indicator data. The results framework is the principal perspective on intended results, but reporting, especially for systemic programs, is much broader. Reporting should be complemented by a broad narrative that provides a full picture of what has happened in the program domain—both intended and unintended changes—and what have been the main drivers of and inhibitors of these changes—whether in the program's sphere of influence or outside of it. Although reporting does not require the full rigor of an evaluation, it should include credible evidence of the program's contribution to change. In constructing this narrative, it is important to refer to the assumptions about the external environment—whether they have proved to be reliable, or whether risks have materialized and how the program responded.

Reporting is not just about the past; it should also look to the future. It should include information about lessons learned that have a bearing on the current program and concrete proposals for adaptation, including revisions to measurement and risk frameworks, based on these lessons.

It is difficult—even risky—to prescribe a particular format for reporting. The form and content of reports should be designed around stakeholder needs and interests. Some contexts demand more structure than others. Nevertheless, an example of a template for a program-to-funder annual report is given in Appendix M8-A, with these reservations in mind.



KEY TAKEAWAYS

1. Reporting is fundamental to accountability, and it can also address stakeholders' learning needs and appetites. It is essential to identify what these needs are and to design reports to meet these needs. Measurement questions can be important guide posts for this process.
2. Reporting should be integrated into review cycles.
3. Significant changes do not happen quickly or frequently in systemic approach programs, and the length of reporting cycles should take this into account.
4. The interventions and outcomes in the results framework, which are nuanced and calibrated by indicators and projections, should be the focus of the reports. In addition, there should be reporting beyond the results framework, including reporting on unplanned outcomes and relevant external developments.
5. Reports should present a plausible narrative of how the program has contributed to change.
6. Reports should describe what lessons were learned and how they will be applied going forward.
7. Information and knowledge brought together for reporting can serve wider audiences, although they usually need to be reformulated and disseminated in more appropriate formats.

Module 8. References and Resources

- Forum Syd. 2014. "Manual on Reporting." Stockholm: Forum Syd. <http://www.forumsyd.org/PageFiles/5098/Manual%20on%20Reporting.pdf?epslanguage=sv>
- Jones, Harry, and Enrique Mendizabal. 2010. "Strengthening Learning from Research and Evaluation: Going with the Grain." London: ODI. <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/6327.pdf>
- New Philanthropy Capital, et al. 2012. "Principles of Good Impact Reporting." New Philanthropy Capital. <http://www.thinknpc.org/publications/the-principles-of-good-impact-reporting-2/>
- Schacter, Mark. 2010. "The Art of the Performance Dashboard." Ottawa: Mark Schacter Consulting. http://docs.wixstatic.com/ugd/dadb01_67024a8b0ac5243ee0d4e8f668ca1e4a.pdf

APPENDIX M8-A. Example of a program-to-funder annual report template

This example assumes a coherent program similar to the example in modules 3 and 5, with an overall results framework that focuses on outcomes. In this example, it is assumed that the number of projects in the program is small enough—say under five—to enable digestible summary reporting on each project. For programs with many projects, the project narratives may need to be integrated into a large and more granular summary section, although the project ratings could be retained.

Even with just a few projects, some funders may prefer more granularity than is proposed in this template, for example, differentiated narrative reporting on each program outcome in the results framework or separate reporting on certain cross-cutting themes, such as gender equity or sustainability.

Risks and/or assumptions from a ToC perspective are covered in this template. However, comprehensive risk management reports, including adjustments to the risk register, are normally made more frequently.

Reporting should be accompanied by guidance on terminology, rating criteria for projects, and so forth. The template should specify the word-count range for each subsection.

A. PROGRAM NARRATIVE

Program staff may prefer to complete this section after completing sections B and C.

Most significant areas of progress and challenge

- With particular reference to the program ToC—including the defined assumptions and/or risks and contribution hypotheses—summarize progress and challenges with the program over the past year.
- Identify the principal factors in program implementation that have helped and hindered progress.
- Describe the most significant developments in the program's external environment and the effects these have had on the program, both positive and negative.

Program-facing lessons, implications, and adaptations

Describe the main lessons learned in the past 12 months about internal and external factors, their implications for the program, and the principal adaptations to the program (*i*) that have already been made in the period and (*ii*) are proposed to be made in the next 12 months. Adaptations should include those to program design, resourcing, and management and to the ToC, risk register, learning questions, results framework, and other aspects of measurement.

Wider insights

Describe any lessons or other insights that you have drawn from your experience with the program and its environment in the past year that may have wider application. Suggest which stakeholders, groups, or institutions might be interested and how they might be engaged.

B. PROJECT PERFORMANCE

This section is for reporting on project performance.

Project 1. [title]

Summary narrative of the project and its progress in the past year.

RATING

Provide an overall rating for the project as whole. Choose from one of the following categories:

Progress significantly above expectations	Progress with all the project's interventions can be shown to be satisfactory and, for at least one, is above expectations. The lessons from this have been captured for wider dissemination.
Progress broadly in line with expectations	Progress with all the project's interventions can be shown to be satisfactory.
Progress moderately below expectations	Progress with at least one of the project's interventions cannot be shown to be satisfactory. Steps are being taken by the project to address this.
Progress substantially below expectations	The project is seriously underperforming. This has been escalated to the program management, which is addressing the issue with project management. A further report will be issued before the mid-point of the reporting year.

Rating	<input type="text"/>
--------	----------------------

Rating narrative

Summarize the reasons for the rating.

Project 2, etc. [title]—Repeat for each project

C. PROGRAM OUTCOME PROJECTIONS

This section is for reporting on outcome projections at all levels in the program results framework. Reporting should be completed only where indicator projections were scheduled in the reporting year.

Intermediate Outcome 1—[title]

1. Indicator data against projections

Indicator 1	xxxx	Baseline (date)	Interim projection (date)	Interim projection (date)	End-line projection
	Planned				
	Actual				
Indicator 2 etc.	xxxx	Baseline (date)	Interim projection (date)	Interim projection (date)	End-line projection
	Planned				
	Actual				

2. Narrative

For each indicator whose projections were scheduled for this year, provide a short commentary on “actual” data, explaining the difference, if any, between the projected and actual result.

Repeat for each Intermediate Outcome

Systemic Change Outcome 1—[title]

1. Indicator data against projections

Indicator 1	xxxx	Baseline (date)	Interim projection (date)	Interim projection (date)	End-line projection
	Planned				
	Actual				
Indicator 2 etc.	xxxx	Baseline (date)	Interim projection (date)	Interim projection (date)	End-line projection
	Planned				
	Actual				

2. Narrative

For each indicator whose projections were scheduled for this year, provide a short commentary on “actual” data, explaining the difference, if any, between the projected and actual result.

Repeat for each Systemic Change Outcome

Inclusive Financial System Outcome 1—[title]

1. Indicator data against projections

Indicator 1	xxxx	Baseline (date)	Interim projection (date)	Interim projection (date)	End-line projection
	Planned				
	Actual				
Indicator 2 etc.	xxxx	Baseline (date)	Interim projection (date)	Interim projection (date)	End-line projection
	Planned				
	Actual				

2. Narrative

For each indicator whose projections were scheduled for this year, provide a short commentary on “actual” data, explaining the difference, if any, between the projected and actual result.

Repeat for each Inclusive Financial System Outcome

D. UNPLANNED OUTCOMES

Describe any unplanned outcomes, positive or negative, for which you have evidence. Refer to specific projects where relevant. Cite the nature of the evidence. Describe the contributions of the program to these outcomes.



EVALUATION



THIS MODULE . . .

- Examines the uses of external evaluation in systemic financial inclusion programs.
- Explores the concept of impact evaluation in the context of these programs.
- Proposes the most appropriate evaluation designs and methodologies.

TIMING AND PURPOSES OF EXTERNAL EVALUATION

Generally, the conventional timing and purposes of evaluations are as follows:

- Mid-term evaluations are usually conducted to determine what can be improved and adapted—How can we do things better? What do we need to do differently?
- End-term evaluations are usually conducted for accountability purposes and to take advantage of strategic learning—Did the program or project achieve what it was intended to? Was it value for money? What lessons should we learn and share with others for future interventions?
- Ex-post evaluations are mainly conducted to capture strategic learning—What differ-

ence did the program or project make? Do the changes appear to be sustainable? What are the implications for replicating this type of program or project?

This basic perspective provides a broad overview of the types of evaluation frequently found in a wide variety of settings. However, evaluating systemic financial inclusion programs requires a certain amount of repositioning.

External evaluation is expensive—monetarily, and perhaps more importantly, regarding the time and effort spent by everyone involved, including partners. To maximize efficiency, funders and program unit staff need to clearly understand what accountability, attribution, and impact mean in their programs. They also need to reconsider the evaluation and monitoring relationship.

Uses of evaluation

For formative learning and improvement

Because financial inclusion programs usually take place in uncertain and dynamic environments, the need for learning and adaptation, both at operational and strategic levels, is more pronounced than in relatively well-understood development interventions, such as some in the health sector. Internal enhanced monitoring and self-evaluation in financial inclusion programs and projects can and should address areas that are traditionally addressed by external evaluation, particularly formative evaluation. Internal staff are likely to understand the complexities of the program environment better than external evaluators, and learning and its application are more likely to flow from an evaluative process in which staff are deeply involved. Furthermore, because of its cost in time and money, only one formative external evaluation is likely to be conducted in the life of a program or project. Because enhanced monitoring takes place more frequently, it provides shorter feedback loops.

External evaluators are still able to contribute to enhanced monitoring through a collaborative model of measurement in which internal reviewers and external evaluators work together. Although most evaluators will not have the same depth of knowledge of the intervention and its environment that internal staff have, they can offer an outsiders' perspective by asking questions and identifying patterns in the responses. External evaluators can contribute to learning and improvement through internal reviews and monitoring/evaluation hybrids such as developmental and real-time evaluation. Appendix M9-A explains the different types of evaluations and their relevance for funders using a systemic approach.

Regarding ToCs, which were presented in Module 3, collaboration between internal monitoring and external evaluators is most relevant where the focus is on the intermediate outcome level. Intermediate outcomes are often qualitative and are more open to interpretation than results at other levels. Their assessment benefits from a collaborative approach. At this level, formative learning is a particularly important use of evaluation and can be effectively leveraged through a collaborative model.

For strategic learning

A more pronounced and independent role for external evaluators is often appropriate for strategic learning at program, portfolio, or organization levels, to help answer a broad

range of measurement questions about multiple interventions. Examples are as follows:

- Is it possible for funder interventions to have a significant impact on the development of market systems that promote financial inclusion? If so, how? Under what circumstances?
- Does financial inclusion have a significant impact on poverty reduction and welfare? If so, with which types of financial services and channels? In what circumstances? How?
- Should the funder continue with a given line of intervention or invest its resources in other ways?

These types of evaluations are major undertakings that need to be planned strategically and allocated considerable amounts of time and other resources.

For accountability

Conventional use of evaluation for accountability calls for a high degree of detachment—which can be provided more credibly by outsiders. It often focuses on validating the achievement of predefined results.

However, in systemic programs, predefined results are only a part of a wider mosaic of change that needs to be pieced together by an evaluation. Attribution can be a contested concept. Accountability in systemic programs and projects needs to have a broad perspective (see Module 2). It should complement the requirement to meet specific targets.

External evaluation plays an important role in a redefined accountability model. It can help to put together a broader set of evidence and bring an outside perspective to judgments about it. It contributes to learning in ways that enhanced monitoring cannot easily do. However, funders should not rely exclusively on evaluations. They should engage with programs enough to know whether other key management processes are in place. They should encourage and support enhanced monitoring that will contribute to the required body of evidence.

APPROACHES TO IMPACT EVALUATION IN SYSTEMIC FINANCIAL INCLUSION PROGRAMS

“Impact” is among the most commonly used words in measurement, and yet it is interpreted in several different ways. OECD-DAC defines impact in a broad, inclusive manner as “positive and negative, primary and secondary

long-term effects produced by a development intervention, directly or indirectly, intended or unintended.” This definition distinguishes impact from the narrower concept of effectiveness in measurement by also referencing unintended and negative effects. Effectiveness is normally confined to outcomes that are positive and intended. Although the definition does not prescribe a specific design for impact evaluation, development stakeholders tend to use the term exclusively in reference to evaluations that use a control group or counterfactual and randomized designs.

The principal factors for approaches to impact evaluation in systemic financial inclusion programs include the breadth and complexity of the financial system, and the time it takes for change at this level to have a large-scale, sustainable impact on financial inclusion and, in turn, on development outcomes. Stakeholders of programs that promote change in more than one part of the financial system at the same time are likely to struggle to control for factors external to the interventions. Randomization is not an option, and even nonrandomized quasi-experimental designs may be difficult to apply because of the lack of an appropriate control group or because statistical approaches are hampered by the small number of units (e.g., partner institutions) at the intermediate outcome level.

It is important to establish how change happens in a systemic approach. Although experimental designs may include data on how change happens (e.g., by incorporating client diaries) this is not usually a program’s primary purpose.

The complex and long ToCs in systemic programs often make it impractical to measure effects at the development outcome level through evaluations that take place at the end of a typical 4–5-year program. These measurements could also be misleading. On the other hand, the aim of systemic programs is to promote large-scale, sustainable change at the development outcome level, through development of the financial system and greater financial inclusion. The implications for impact evaluation of systemic programs are threefold:

- The most suitable design is likely to be theory-based evaluation (TBE), which is generally compared against experimental impact evaluations.¹⁶ However, it is possible and often desirable to combine TBE with quasi-experimental approaches, or even randomized control trials (RCTs), to test specific links at lower levels in the results chain.
- TBE may focus more on impact at the inclusive financial system level, than on development outcomes.
- Researching the impact of financial inclusion on poverty and well-being is best approached strategically through a combination of initiatives, including multiple experimental or quasi-experimental evaluations and large-scale longitudinal approaches. Combined, these initiatives would be beyond the scope of many programs because of the significant resources involved. This points to the need for funders and other actors to collaborate strategically to advance knowledge in this area.

Using experimental designs for specific links in the results chain

Experimental methods are not the most appropriate evaluation tool to measure change at the market-system level. Experimental and quasi-experimental methods require a credible counterfactual. When the unit of evaluation is the entire market system, it is impossible to create a counterfactual because there is only one market system.

Nevertheless, evaluators can use these methods to test specific links in results chains. They can tell us whether individual, relatively discrete, interventions contributed in the way we expected them to. Experimental methods are particularly helpful during pilot stages to confirm a hypothesis before scaling up.

For example, if a project objective is to stimulate demand for a specific product (e.g., microinsurance), then evaluators may choose to use experimental methods to compare the effectiveness of different forms of outreach to potential customers. This evaluation would not reveal whether the program had a systemic impact on financial inclusion, but it could clarify whether specific interventions were leading to their intended intermediate outcomes. Because RCTs can be expensive, their utility should be carefully assessed.



KEY TAKEAWAYS

- External evaluation is expensive and needs to be justified by robust use cases. At the project level, particularly for formative evaluation, there may be a case for using external evaluators in collaboration with internal staff.
- TBE is the most appropriate methodology for systemic programs and projects in financial inclusion. Rigor in TBE depends on several approaches, including deeply probing the ToC and capturing data about other interventions and contextual factors to explore plausible alternative contributions to the identified changes. At the heart of TBE is the understanding that evaluators are looking for contribution.
- Theory-based impact evaluation at the project and even program level should focus on the financial system level and financial inclusion. Measuring the impact of financial inclusion on development outcomes, such as increased economic well-being, needs to be done strategically through a combination of methodologies and preferably through collaboration between funders and other stakeholders.
- Experimental and quasi-experimental evaluation designs are unlikely to be relevant to most areas of systemic financial inclusion programming. However, they can contribute to strategic research for testing the link between financial inclusion and development outcomes.

Module 9. References and Resources

- Better Evaluation. n.d. "Contribution Analysis." Retrieved 17 October 2015. http://betterevaluation.org/plan/approach/contribution_analysis
- Carter, Becky. 2012. "Theory-Based Evaluation Approach. Governance and Social Development Resource Centre." London: DfID. <http://r4d.dfid.gov.uk/Output/192576/>
- Collier, David. 2011. "Understanding Process Tracing." *Political Science & Politics*, 44, No. 4: 823–30. <http://www.ukcds.org.uk/sites/default/files/uploads/Understanding-Process-Tracing.pdf>
- Dozois, Elizabeth, et al. 2010. "A Practitioner's Guide to Developmental Evaluation." Montreal: J.W. McConnell Family Foundation and International Institute for Child Rights and Development. http://vibrantcanada.ca/files/development_evaluation_201_en.pdf
- Harrison, Tom. 2015. "RCT and Market Systems: An Opportunity to Gain Insights." Blog post, 22 January. <http://beamexchange.org/en/community/blogs/blog-entry/2015/1/22/tomharrison/>
- Mayne, J. 2008. "Contribution Analysis: An Approach to Exploring Cause and Effect." ILAC Brief 16. Rome: Institutional Learning and Change Initiative. http://www.betterevaluation.org/en/plan/approach/contribution_analysis and https://beamexchange.org/uploads/filer_public/f5/55/f5553494-f9cf-4f2f-90ee-105877d44d71/contribution-analysis.pdf
- O'Sullivan, Fionn. 2016. "Impact Evaluations for Market Systems Programmes." London: The BEAM Exchange. <https://beamexchange.org/guidance/evaluation-guidance/>
- Puri, Jyotsna. 2014. "Quasi-Experimental Methods for Impact Evaluations." New Delhi: Asian Development Bank and International Initiative for Impact Evaluation. http://www.betterevaluation.org/en/resources/guide/quasi-experimental_methods_for_impact_evaluations
- Ruffer, Tim, and Elise Wach. 2013. "Review of M4P Evaluation Methods and Approaches." London: DfID. <http://www.itad.com/reports/review-of-m4p-evaluation-methods-and-approaches/>
- Scriven, Michael. 1991. "Prose and Cons about Goal-Free Evaluation." *American Journal of Evaluation*, 12: 55. <https://thefutureofbusinesseducation.wikispaces.com/file/view/Goal+Free+Evaluation+Scriven.pdf>
- Stern, Elliot, et al. 2012. "Broadening the Range of Designs and Methods for Impact Evaluations." London: DfID. <https://www.gov.uk/dfid-research-outputs/dfid-working-paper-38-broadening-the-range-of-designs-and-methods-for-impact-evaluations>

Note

16. See, e.g., O'Sullivan (2016).

APPENDIX M9-A. Evaluation designs and methodologies

The lexicon of evaluation methodology is dense and can be confusing. At the risk of over-simplification, it is possible to see a convergence in recent years on two principal types of evaluation, particularly where the focus is impact: experimental and quasi-experimental evaluation on the one hand, and nonexperimental TBE on the other. In development, generally, these two evaluation types have the same goal. In many areas of market systems development, TBE is considered to be the most appropriate approach to external evaluation (Ruffer and Wach 2011).

Experimental and quasi-experimental designs

Experimental and quasi-experimental evaluation seek to identify the effects of the intervention by isolating the interventions from other factors. It uses focused baseline and end-line definitions of the status of what is being evaluated, and it uses data in the same way data are used in a counterfactual or control group. In some designs—particularly RCTs—the control group at baseline needs to be a close (ideally identical) analogue of the “treatment” group (i.e., the intended beneficiaries of the intervention). In RCTs, the control group is drawn from the same population as the group that receives the treatment. In other designs, differences are permitted between the treatment and control groups, but these differences need to be clearly identified.

These conditions are difficult to achieve in systemic programs and projects because these programs and projects have long ToCs and results chains (respectively), their environments are complex and dynamic, and their boundaries are blurred. Depending on the state of development of the financial system, it may take years before the effects of the interventions on customers and their enterprises can be detected. Systemic programs do not provide the controlled environment that experimental, or even quasi-experimental, designs require. It is often difficult to find a true counterfactual in the conventional sense of the word. Furthermore, the necessarily adaptive, opportunistic nature of systemic financial inclusion interventions often diminishes the role of original baselines because the focus of the intervention shifts in response to obstacles or new opportunities.

Experimental and quasi-experimental designs in isolation tell us only whether an intervention has had an effect, not how. They are often described as “black box” studies—leaving the box containing the effect mechanism unopened. In complex environments, it is important to know how change happened or why it did not. Without this knowledge, replication and scale-up are unlikely, because context and process are important. Even if these evaluation designs met the controlled environment criterion, they would need to be accompanied by qualitative assessment of process and a consideration of other influencing factors, which will always be unique to the intervention’s immediate environment.

As such, experimental evaluation is unlikely to be relevant to most areas of systemic financial inclusion programming. It can, however, contribute to strategic research for testing the link between financial inclusion and development outcomes.

Theory-based evaluation

In sharp contrast, TBE is well-suited for systemic programs. At its simplest, TBE follows the path of the ToC or results chain from the interventions through the various levels of expected outcomes, gathering evidence as to whether the changes at each level have, as hypothesized, been influenced by the intervention and whether the supporting assumptions have held up. TBEs will be able to test the ToC or results chain. Some forms of TBEs are able to estimate the strength of the intervention’s influence and that of other contributing players and factors.

TBE depends on a ToC or results chain. In adaptive programs like those typically found in financial market development, evaluators should ensure that the ToC or results chain is up to date. TBEs can be used not only to assess whether expected results have occurred, they also look at how they occurred or why they did not.

Appropriate rigor needs to be applied in TBEs through several actions:

- Having a robust ToC or results chain that has been tested and adapted, where necessary, during the intervention.
- Defining measurement questions at the beginning of the program or project to enable monitoring to maximize its contribution to the evaluation.

- Identifying baselines for key anticipated changes.
- Capturing and assessing data about anticipated and unanticipated outcomes.
- Capturing data from a sufficient number of different sources, both quantitatively and qualitatively, to enable “triangulation” and use of credible strategic informants.
- Establishing plausible connections between the interventions and the identified changes.
- Capturing data about other interventions and contextual factors (e.g., economic and political) to explore other possible contributions to the identified changes.

Contribution analysis, process tracing, and realistic evaluation are three prominent TBE methodologies.

Contribution analysis, like other forms of TBE, does not set out to prove causality but rather to provide “evidence and a line of reasoning from which a plausible conclusion can be drawn that, within some level of confidence, the program has made an important contribution to the documented results” (Better Evaluation n.d.).

Appropriate evaluation questions may include the following:

- Was the intervention and its pre-identified supporting factors and assumptions sufficient to produce the intended result?
- What role did the intervention have in this? How did its contribution compare to those of other factors, both pre- and post-identified?
- How and why did the intervention make—or not make—a difference?

Implicit in this approach is that the evaluator should look beyond the intervention for contributions to the changes discovered. Before concluding that the intervention played a central role, the evaluator needs to look at other plausible explanations for the changes.

Also implicit is that evaluators are looking for plausible attribution, or in some cases sole contribution, rather than strict causality. Most financial inclusion interventions attempt to influence a system that has many interdependent parts because numerous nonprogram factors contribute to change. Change often happens episodically, after periods of inaction. This makes it difficult to isolate the change the program has caused. Evaluation expectations need to be set so that plausible attribution or contribution can be discovered.

Strict causality, beyond the most direct of intermediate outcomes, is difficult if not impossible to prove.

Process tracing is similar to contribution analysis in that it analyses whether results are consistent with the program theory and whether alternative explanations can be ruled out. It does this by applying a highly structured methodology that involves the systematic examination of “diagnostic evidence selected and analysed in the light of research questions and hypotheses posed by the investigator” (Collier 2011).

Realistic evaluation also takes a similar approach to contribution analysis, but it places even more emphasis on context. The following evaluation questions would be asked in combination: What works? For whom? In what circumstances? In what respects? How? A ToC in realistic evaluation is highly context-specific and is usually seen to have weak external validity. Replication would require the ToC to be heavily recontextualized and adjusted.

Inward-outward perspectives in TBE

Conventional effectiveness-focused evaluation largely follows the ToC from the intervention upwards to the highest level of evaluation interest (e.g., advances in financial inclusion). This outward approach is also the perspective of conventional approaches to monitoring.

However, explicit in TBE, and particularly in contribution analysis, is that an intervention-centric perspective needs to be set aside at some point to examine the possibility of other explanations for the changes discovered. In this inward approach, measurement is not led by pre-set outcomes, it looks for changes that have taken place in the program environment and works backwards to identify the plausible contributions to the changes, including the intervention in question.

Evaluation that includes inward perspectives allows for greater triangulation and reduces the intervention centricity that can distort the findings and conclusions of effectiveness-oriented evaluation. These approaches capture information on outcomes in the project or program terrain, but without reference at first to the intervention itself. The inward perspective can be applied in evaluation and in enhanced monitoring.

There are advantages and disadvantages in both inward and outward approaches to measurement. A combination of these approaches is most likely to address the inherent draw-

backs of applying either approach on its own, although this implies requiring more resources. Figure M4-1 and Table M4-1 in Module 4 illustrate measurement questions that could be applied in this way.

Developmental methodologies

In this handbook, the term “developmental methodologies” refers to a range of approaches that challenge the conventional separation of monitoring from evaluation. Developmental methodologies can work well with the collaborative model described earlier in this module. The key principle underlying these developmental methodologies is that they are intended to have an immediate impact on programming. The evaluator’s primary role is to bring evaluative thinking and the use of data into program and project management on a regular, or even continuous, basis. The most common examples of these approaches are real-time and developmental evaluation.

Real-time evaluation originated in the humanitarian sector; the Office of the United Nations High Commissioner for Refugee institutionalized it in 2000. Emergency relief is arguably the most complex, dynamic, programming environment. Real-time evaluation in these situations emphasizes the use of constant feedback loops to improve processes. Real-time evaluation has been applied more recently in nonhumanitarian contexts.

Like its real-time equivalent, developmental evaluation is designed for complex, dynamic environments, which makes it suitable for much of financial inclusion programming, particularly for measurement at the lower levels of project results chains. It fits neatly into a collaborative model because it complements enhanced monitoring. It requires measurement itself to be flexible so that the focus and mechanisms for measurement can change as intended outcomes evolve or shift.

“Developmental evaluation facilitates assessments of where things are and reveals how things are unfolding; helps to discern

which directions hold promise and which ought to be abandoned; and suggests what new experiments should be tried” (Dozois et al. 2010).

Although there is a range of methodological approaches to real-time and developmental evaluation, there are some commonalities:

- They take place during implementation.
- They take place iteratively rather than as a one-off, and they are integrated into the program cycle.
- The emphasis is on immediate lesson-learning more than on impact assessment or accountability.
- They aim to produce context-specific understandings that inform ongoing innovation rather than generalizable findings across time and space.

With these methodologies, evaluation questions and intended results should be allowed to evolve as evaluation progresses. They remain theory-based.

Goal-free methodologies

The term “goal-free evaluation” applies to a range of methodologies in which the evaluator conducts the evaluation without particular knowledge of, or reference to, predetermined results. The emphasis is on “finding out what the program is actually doing without being cued as to what it is trying to do. If the program is achieving its stated goals and objectives, then these achievements should show up; if not, it is argued, they are irrelevant” (Scriven 1991).

“Outcome harvesting” is a recent variation of the goal-free approach. Evaluators or project stakeholders collect evidence of what has changed, and work backwards to determine whether and how the project or intervention contributed to the change. Goal-free approaches in financial inclusion should complement outward perspectives in both monitoring and evaluation; it should not be a stand-alone evaluation methodology.



PORTFOLIO-BASED MEASUREMENT



THIS MODULE . . .

- Explains how portfolio-based measurement, using standardized indicators and/or scoring systems, can produce insights for funders to improve their interventions and make decisions about their overall strategy.
- Presents two portfolio measurement tools: portfolio dashboards, which usually are applied at least once a year, and portfolio reviews, which take place less frequently.

For funders and other organizations that implement multiple programs, taking a portfolio approach to measurement provides insights and helps report results to constituencies across the whole portfolio or across specific subsets of programs.¹⁷ Comparing performance across systemic programs is challenging. Context is important, and progress is often measured in terms of changed practices, processes, and structures—making considerable use of qualitative indicators that do not lend themselves to direct comparison or aggregation. Nevertheless, at higher levels in the ToC, there is scope for direct comparability through standardized indicators. In addition, well-designed and consistently applied methods of indirect comparison can be used.

THE IMPORTANCE OF A PORTFOLIO APPROACH TO MEASUREMENT

Unlike individual program evaluations from which lessons may be difficult to extrapolate beyond the program, portfolio-based monitoring and review compares results across all or substantial parts of a portfolio. Portfolio reviews are used to identify reasons for variations in results. These evaluative reviews can produce insights that are broadly valid and that help funders make evidence-based decisions. Portfolio reviews also fulfill an overarching accountability function. They can be used to assess whether funders are achieving their objectives, whether they are achieving results efficiently, and to the extent possible, whether these results are relevant in their sectoral and market contexts.

TOOLS FOR PORTFOLIO-BASED EVALUATIONS AND MONITORING

Portfolio reviews and dashboards are two complementary tools funders and program units can use to aggregate results at the portfolio level.

Portfolio reviews of financial inclusion are thorough assessments of completed or well-advanced programs that are done by independent evaluators and are repeated every three to five years. They can be used to assess whether a funder is delivering on its financial inclusion strategy.

Portfolio reviews typically assess the composition and the performance of the portfolio. They include five phases: preparation, portfolio snapshot, performance scoring, stakeholder consultation (including field visits as much as possible), and overall analysis and reporting (Figure M10-1).¹⁸ A portfolio review can take up to a year from initiation to the submission of the final report to complete.

Portfolio dashboards are designed to regularly inform boards and constituencies about overall portfolio progress, and to help senior management identify programs that need attention and to react appropriately. They can indicate whether current programs are “on track.” Deeper scrutiny can be triggered if the dashboard indicates that a program appears to be diverging from expectations. Some dashboards incorporate a facility to automatically “drill down” to greater program—or even project-specific—data.

Because dashboards are meant to be read and understood quickly and easily, they are usually graphic presentations of data. Dashboards that illustrate program aspects that are subject to change within a short timeframe are often revised more than once a year.

Dashboards typically include the following:

- Program categories (e.g., by region, type of program, type of partner)
- Disbursements
- Values of standardized indicators
- Performance against projections (see Module 6)
- Timeline tracking
- Overall performance ratings (based, for example, on a red/amber/green “traffic light” rating scale)
- Risk ratings
- Short summaries of progress and of issues that arise

Dashboards are usually created by program or project staff. They should not require much, if any, additional research beyond existing monitoring. Dashboards should incorporate as many visual elements as possible, for example, color-coding to highlight high or low performers and important changes made since the previous reporting period.

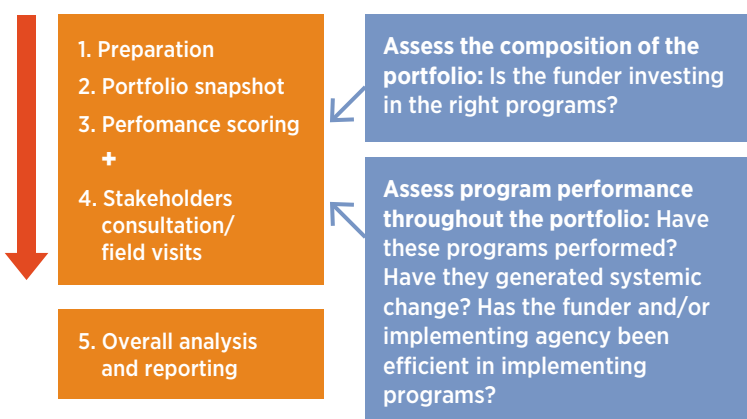
Because dashboards rely on quantification, scoring through categorization, and monitoring against predicted timelines, it may be difficult to illustrate standardized indicators at the intermediate outcome level because comparisons of program progress need to be put in context. Change often happens episodically, so judging progress against predicted timelines can be misleading. Dashboards should be viewed as an entry point for deeper analysis.

STANDARDIZATION

Standardization is a controversial concept in market systems development because the systemic approach requires flexibility and context-specificity throughout the program. However, the ability to standardize and aggregate performance information is necessary for both dashboards and reviews. Standardization and aggregation are required for a holistic perspective of the portfolio, and to enable comparisons. They do not need to undermine flexibility in implementation.

The focus should be on the middle levels of the ToC (as defined in Module 3)—later intermediate outcomes, systemic change, and the inclusive financial system—to enable standardization in portfolio-based measurement. Development outcomes require different evaluation instruments, as discussed in Module 9.

FIGURE M10-1. Portfolio review components



Experience with portfolio reviews

Portfolio reviews to help revisit strategic priorities—Agence Française de Développement (AFD)

A six-month portfolio review of AFD's microfinance operations was conducted by independent consultants in 2010–2011. The review was intended to provide, for the first time, an external and critical viewpoint on the performance of AFD's microfinance projects. The review involved the use of a predetermined methodology to score a representative sample of active projects and was based on documents and field visits in a selected number of countries. The operational diagnostic made it possible to identify several strategic priorities, based both on market needs and AFD's strength and added value. The findings contributed to developing the definition of AFD's financial inclusion strategy.

Portfolio reviews to evaluate strategy implementation—African Development Bank (AfDB)

In 2013, AfDB leaders decided to evaluate its microfinance policy, strategy, and operations from 2000 to 2012. There was no automated system in place that allowed for the extraction and consolidation of information about all financial inclusion projects. The evaluation included a thorough portfolio review. In this review, microfinance projects were identified, performance data were collected, and the projects were assessed through standardized indicators. The review identified several discrepancies between the strategy and the operations. For example, whereas the strategy specified private-sector department leadership for financial inclusion operations, most of the projects were still channeled through governments. AfDB used the findings to update its strategy.

At these middle levels, the tension between standardization and flexibility can be reconciled through “smart standardization” that accounts for context and does not require projections to be set in stone from the onset of the program. Smart standardization can be achieved through a mix of aggregation of standardized quantitative indicators (A in Figure M10-2) and subjective scoring, which reflect context (B).

Although there is no standard methodology for results aggregation, funders can use the following guidance and generic set of scoring factors to develop a methodology that meets their specific needs and context and that accounts for data, time, and budget constraints. The following is a four-phase approach to aggregate performance information while accounting for context:

1. *Design*—objectives are set and tools are developed.
2. *Snapshot*—portfolio composition and performance data are collected, and where possible, aggregated using standardized indicators.
3. *Scoring*—program performance is rated using standardized scoring criteria, taking into account other monitoring data and context, and then aggregated.
4. *Analysis*—performance is compared, patterns are identified, and recommendations

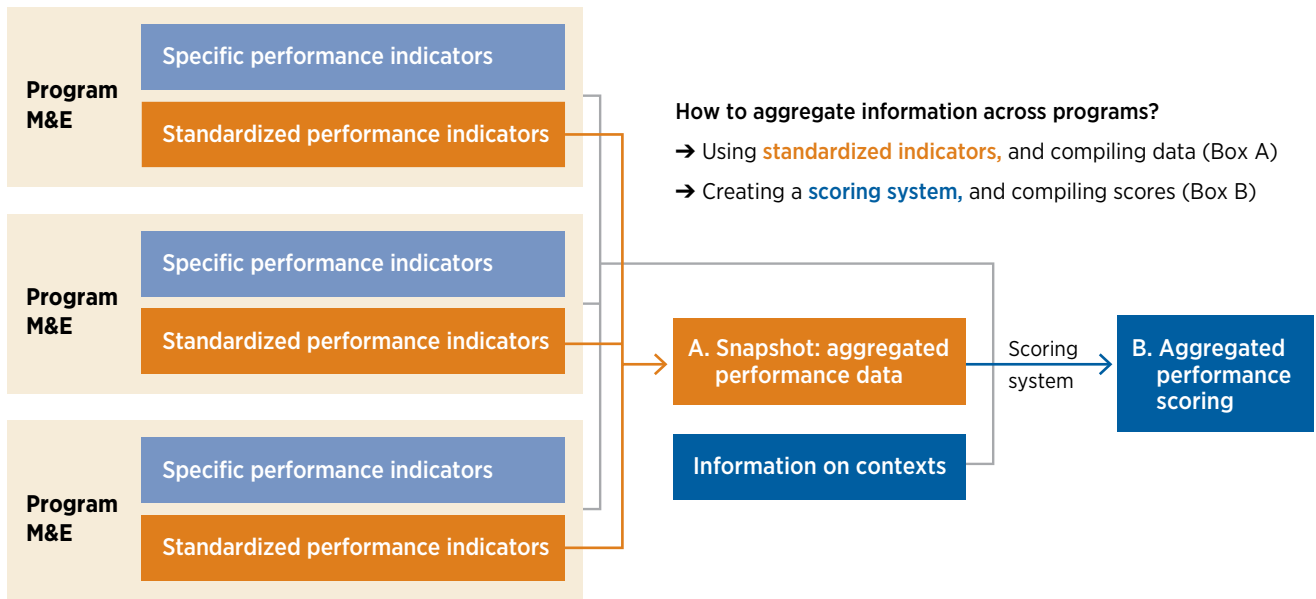
are made for changes in program management and strategy.

Whereas the snapshot phase focuses on objective and quantifiable information, the scoring and analysis phases include contextual and other qualitative information and subjective assessment.

Design. In the design phase, the evaluator (or the program managers in the case of portfolio dashboards) and the funder (senior management) work together to adapt the evaluation framework to the funder's information needs and determine the scope of the review. A central element of this phase is agreement on the evaluation questions to be answered. The overarching question a portfolio review is designed to answer is whether the funder delivers on its financial inclusion strategy (i.e., is the funder investing in the right programs and projects and is it achieving its objectives?). Funders may have other questions that portfolio-based measurement can be used to answer (e.g., about specific links in a portfolio-wide ToC). Like all measurement activities, the design and use of portfolio reviews and dashboards should be guided by measurement questions from the funder's measurement strategy (see Module 4).

Snapshot. The evaluators collect information on programs and consolidate this information. The objective is to get a good understanding of the composition of the funder's portfolio

FIGURE M10-2. Smart standardization for performance aggregation



Note: M&E = monitoring and evaluation

and its performance. What are the characteristics of the programs? What were their initial objectives in terms of intermediate and market system outcomes? Who were the funder's partners? How have funds been used? Have program objectives been achieved?

Table M10-1 summarizes the types of data needed for the snapshot. Consolidating these data enables a thorough analysis of the composition of the portfolio and helps gather key information for program scoring. Data included in the snapshot should be transparent and objective. Funders should make sure their own and their partners' data management systems can automatically generate what is needed (program documents should be structured in a way that guarantees the necessary information is collected).

Scoring. In this phase, evaluators (i.e., program managers in the case of portfolio dashboards) score the performance of programs and consolidate scores across the portfolio. Using scoring criteria makes it possible to include results that are not measurable through quantitative indicators. Using a well-defined scoring methodology reduces potential bias in the assessment process.

Evaluators score all programs, based on the data collected during the snapshot phase and on information from key informant interviews and field visits. The scoring grid example in Appendix M10-A can be used as a starting point for funders to adapt and develop their own grids. This example is based on an adaptation of OECD's Development Assistance

Committee evaluation criteria:

- *Relevance and coherence.* Are programs that are aligned with a portfolio-wide ToC designed to promote financial system development? Are they relevant to country needs and coherent with country long-term strategies?
- *Effectiveness.* Have objectives been achieved? or How likely are they to be achieved?
- *Efficiency.* Are program interventions commensurate with allocated resources?
- *Impact.* What is the nature and scale of change at the relevant levels of the ToC to which the programs have contributed?
- *Sustainability.* Are programs' positive outcomes being sustained or are likely to be sustained?

Criteria that assess the *extent to which* a program met a certain objective, rather than granting scores on a binary (i.e., yes or no) basis, should be used in the scoring grid. Evaluators should use their judgment in scoring and incorporate contextual factors.

The scoring process should follow clear guidelines to allow for flexibility and to ensure the comparison is relevant.¹⁹ Guidelines may address the following:

- A scoring scale that explains the meaning of each score should be designed (see Figure M10-3).
- Qualified evaluators should be selected and trained. Evaluators are necessarily a key ele-

TABLE M10-1. Data for the portfolio snapshot

DESCRIPTION			Performance against results frameworks
Program	Program unit	Program partners	
Information about the program. The measurement system should collect this information from programs in a standardized way. Examples: region or country, key dates, funding amount, funding instruments, type of program.	Information on who is implementing the program. Examples: funder staff, consultancy firm, local implementer, government unit.	Information on the partners (e.g., regulators, FSPs, sector coordination facilities, membership organizations, universities, statistics bureau). Examples: type of institution, years of operation, range of services. For private-sector entities, there should also be information on outreach (size and type), size (number of employees, etc.), and financial situation.	For private-sector entities, there should also be information on outreach (size and type), size (number of employees, etc.), and financial situation. Information on the levels of achievement of the program's expected outcomes, normally at the intermediate, systemic change, and inclusive financial systems levels. ^a Standardized indicators, if available, are used so that performance can be aggregated. Data normally come from partners' own monitoring systems.

a. Intermediate outcomes focus on changes in attitudes, behaviors, institutional processes, etc., that are difficult to capture with standardized indicators and, therefore, are unlikely to be included in the snapshot phase. Data on outcomes at this level should be introduced selectively during the scoring phase.

FIGURE M10-3. Sample scoring scale

- 6 Exceptionally successful.
- 5 Very good. Worth extending or replicating.
- 4 Satisfactory/all right. Good for this specific situation, but not worth replicating/extending.
- 3 Weak but minimally acceptable. The evaluator wouldn't support such a project, but wouldn't veto it or stop it.
- 2 Unacceptable. Unlikely to produce a decent development return on the funds invested. The evaluator would veto such a project.
- 1 Exceptionally poor. Probably does serious damage, or represents extreme bad practice.

FIGURE M10-4. Portfolio review—Evaluations profiles

The operational responsibility for a portfolio review should be with an independent team of evaluators. Independence is important not only to achieve accountability, but also to gain an external view on the relevance and performance of programs.

The evaluation team needs to combine evaluation skills with expertise in financial inclusion, a good knowledge of the market system development approach, and, last but not far from least, a good knowledge of the countries where the funder operates.

ment of the “smart standardization” process. Accounting for context relies on their knowledge of the context and their analytical skills (see Figure M10-4).

- The scoring process should be transparent. Evaluators should list available evidence and justify their scores. In the case of large portfolios, to save time, the evaluators might focus on justifying the lowest and highest scores. To test whether the scoring system is used consistently across different evaluators, evaluators could score a sample of projects and check whether the scores and justifications are consistent. For quality assurance, to the extent possible, all projects should be scored independently by two evaluators before a final score is assigned.

Analysis and action. At this stage, sufficient information should be available for analysis. Evaluators need to extract findings that will help the funder achieve better results with its portfolio. The evaluators use data collected in

previous phases to answer the measurement questions defined for the portfolio review. They combine findings and analyze the compiled data for trends and correlations to extract lessons on what works, what does not, and why. It is important at this stage to triangulate findings with country-level market data and to verify them with key stakeholders. The evaluators should develop recommendations for the funder based on these findings.

The findings and recommendations should be consolidated in the final report. It should be a coherent standalone product that a reader can understand without having to refer to intermediary reports and other documents produced during the review process. It should be concise and draw the reader's attention to the main findings the evaluators consider important and realistic for the funder to act on. The reports are usually shared with key stakeholders. By contrast, portfolio dashboards are usually shared internally.



KEY TAKEAWAYS

- Portfolio-based monitoring and evaluation, based on aggregated information on programs and their performance, are useful for both learning and accountability purposes.
- Portfolio reviews and dashboards are two examples of results measurements at the portfolio level. Dashboards aim to detect whether or not existing programs are going in the right direction. Reviews aim to assess whether the programs are delivering on funder strategies, how, or why not.
- Like all measurement activities, the design and use of portfolio reviews and dashboards should be guided by measurement questions from the funder's measurement strategy.
- Some degree of data standardization of program composition and performance indicators is necessary to enable aggregation at the portfolio level. But there are limits to the practicability and validity of standardized indicators in systemic financial inclusion programs, particularly at the intermediate outcome level. Portfolio reviews need to combine standardized indicators with scoring, which, while systematic and based on clear criteria, allows evaluators to use their judgment based on contextual factors and non-standardized performance data.

Module 10. References and Resources

- Biscaye, Pierre, et al. 2015. "Evaluating Donor-Level Results Measurement Systems." Seattle: Bill & Melinda Gates Foundation. https://evans.uw.edu/sites/default/files/EPAR_UW_300_Donor%20Results%20Measurement%20Systems_08.21.15_0.pdf
- Burjorjee, Deena M., and Barbara Scola. 2015. "A Market Systems Approach to Financial Inclusion: Guidelines for Funders." Washington D.C.: CGAP. https://www.cgap.org/sites/default/files/Consensus-Guidelines-A-Market-Systems-Approach-to%20Financial-Inclusion-Sept-2015_0.pdf
- Development Gateway. 2016. "What Systems and Processes Do Donors Use for Measuring Aid Effectiveness?" Blog post, 25 March. <http://www.developmentgateway.org/blog/what-systems-and-processes-do-donors-use-measuring-aid-effectiveness>
- Nègre, Alice, and Barbara Scola. 2012. "Portfolio Reviews: Resource Guide for Funders." Washington D.C.: CGAP. <http://www.cgap.org/publications/portfolio-reviews-resource-guide-funders>
- OECD (Organization for Economic Co-operation and Development). 1991. "The DAC Principles for the Evaluation of Development Assistance." Paris: OECD. <http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>
- Smart Campaign. "Smart Campaign Client Protection Principles." Washington, D.C.: Smart Campaign. <http://www.smartcampaign.org/about/smart-microfinance-and-the-client-protection-principles>
- UNPRI (United Nations-Supported Principles for Responsible Investment). n.d. "Principles for Responsible Investment." London: UNPRI. <https://www.unpri.org/about>

Notes

17. This module builds on Scola and Nègre (2012).
18. See Scola and Nègre (2012) for detailed guidance for each of these phases.
19. Scola and Nègre (2012) provide additional guidance on how to define a scoring grid and processes.

APPENDIX M10-A. Scoring grid example

RELEVANCE AND COHERENCE	Scoring Criteria
Do programs follow an appropriate ToC?	R1 Extent to which the program, based on a portfolio-wide ToC, meets a set of effectiveness criteria (see Module 3).
Are program objectives consistent with the funder's overall strategy or ToC?	R2 Extent to which the program results framework is aligned with objectives in the funder's strategy and ToC.
	R3 Extent to which the program complements the funder's other programs, especially in the same country.
Are the programs in a given country aligned with national strategies to promote financial inclusion?	R4 Extent to which the objectives of programs in a given country are aligned with or complementary to the objectives in national strategies for financial inclusion where they exist.
	R5 Extent to which the program design intentionally addresses diagnosed market system development needs.
Are the programs in a given country aligned with the efforts of other funders and non-government agencies?	R6 Extent to which the program is based on clear and documented analysis of the funder's value-added, taking into account the funder's comparative advantage and other agencies' contributions.
Are programs following internationally recognized best practices?	R7 Extent to which the program is in line with internationally recognized best practices (see Burjorjee and Scola [2015]); Principles for Investors in Inclusive Finance; Principles for Responsible Investment ²⁰
Are program designs relevant to meet their intended objectives?	R8 Degree of consistency between program objectives and partners' needs, strategy, and capacity to implement the program and extent to which consultation ensured effective ownership by recipients.
	R9 Extent to which programs included a clear crowding-in strategy beyond the program partners.
	R10 Extent to which instruments, funding amounts, and planned timeframe were consistent with the program objectives and provided the necessary flexibility.
EFFECTIVENESS	Scoring Factors
Are measurement systems fit to track the effectiveness of the programs?	ES1 Extent to which outcomes were clearly specified, at all levels of the ToC, with relevant indicators, baseline values, projections, and measurement sources.
	ES2 Extent to which the program has established a measurement system that allows for enhanced monitoring of unintended, positive, and negative outcomes, as well as intended outcomes.
	ES3 Extent to which performance data are sufficient in all relevant categories, quality assured, and available for comparison across programs.
Are planned outcomes up to and including the systemic change level achieved or likely to be achieved?	ES4 Extent to which intermediate outcomes have been or are being achieved.
	ES5 Extent to which systemic change outcomes have been or are being achieved.
	ES6 In case of loans or refinancing, extent to which the recipient has repaid or is repaying the installments on time.
EFFICIENCY	Scoring Factors
Is the funder enabling partners and third-party implementers to deliver on their commitments in a timely and efficient manner?	EY1 Extent to which the program and its projects were designed, negotiated, and finalized within reasonable time and/or agreed timeframes.
	EY2 Extent to which financing was provided in line with activities or other financing needed or extent to which disbursement took place according to plan.
	EY3 Extent to which the funder has an accountability regime that matches criteria for effectiveness in the context of market systems development.
Are contracting systems designed to ensure efficiency?	EY3 Extent to which the funder has used transparent, competitive procurement processes for contracting resources to implement the program (e.g., consultants, services providers, etc.).
	EY4 Extent to which the funder's contracting (e.g., performance-based contracts) and accountability mechanisms incentivize high performance.
Have program interventions been delivered according to plan?	EY5 The extent to which planned interventions have been delivered according to planned factors: timing, scale, specification, quality.
Have program intervention results been commensurate with the resources used?	EY6 Extent to which resources used (e.g., money, staff, and materials) are commensurate with the interventions.

APPENDIX M10-A. Scoring grid example, *continued*

RELEVANCE AND COHERENCE	Scoring Criteria
IMPACT	Scoring Factors
What are the impacts of the programs on inclusive financial system development and (where relevant) on development outcomes?	I1 Extent to which the program has had a positive impact on partners' provision of services, regulation, etc.
	I2 Extent to which the program has had a positive impact on the range and the scale of inclusive financial services.
	I3 Extent to which the program has had a positive impact on the financial capability of individuals, households, and enterprises.
	I4 Extent to which the program has had a positive impact on market infrastructure and supporting functions.
	I5 Extent to which the program has had a positive impact on rules and norms.
	I6 Extent to which the program has had a positive impact on the uptake and use of financial services by poor and low-income people.
	I7 Extent to which the program has contributed to positive development outcomes.
	I8 Extent to which the program has contributed to negative displacement effects and other negative outcomes, socially, economically, or environmentally.
SUSTAINABILITY	Scoring Factors
Are there clear exit strategies for the funders' support?	S1 Extent to which exit options were assessed and defined in program design and subsequent work plans, and carried out.
Are partners able to sustain their operations without continued funding?	S2 Extent to which partners have increased their financial sustainability over the funding period, and have the financial capacity to address recurrent costs.
	S3 Extent to which partners have enhanced capacity to innovate and withstand shocks beyond the boundaries of program interventions.
Has market sustainability and resilience been strengthened by the programs?	S4 Extent to which the program has influenced systemic change and sustainable change in the inclusive financial system.

Note

20. <https://www.unpri.org/about> & <http://www.smartcampaign.org/about/smart-microfinance-and-the-client-protection-principles>

GLOSSARY

Adaptive management	Purposeful experimentation to answer questions that are critical to the achievement of results; includes pivots in strategy and tactics based on feedback loops coming from the results of detailed, planned experiments or pilots.
Attribution	The degree to which an outcome can be shown to be caused by one or more interventions .
Catalyzing	Increasing the rate of a reaction through the participation of an additional substance: a catalyst. With a catalyst, reactions occur faster and uses less energy. The catalyst itself does not become part of the change.
Contribution	The demonstration of a plausible link between an outcome and one or more interventions , but without fully isolating the effect of the intervention(s) from other factors.
Contribution hypothesis	An explanation of the logic behind the progression from one level of a theory of change or results chain to the next.
Development domain	A sector, subsector, or broad focus area, such as financial inclusion, for development interventions.
Facilitation	Working with system actors to catalyze the desired systemic change process.
Financial inclusion	A state in which all individuals, households, and businesses have the choice to access, and the ability to use, a range of appropriate financial services, responsibly and sustainably provided by institutions enabled to offer such services.
Financial system	An umbrella term that includes a market system for different financial services (e.g., payments, savings, credit, insurance, leasing, Sharia-compliant financial products, etc.).
Funder	Public or private organizations that support financial inclusion to achieve a developmental mission or mandate, including bilateral and multilateral development agencies, private foundations, and development finance institutions. Funders typically fund the programs; they may also act as an implementer. Funders are not considered market actors because they only temporarily intervene in market systems.
Impact	Positive and negative, primary and secondary, medium- to long-term, intended or unintended effects produced by interventions.
Innovation	New practice or approach adopted by a market player. May include new goods, services, supporting functions, rules, and norms produced as a result of the new practice or approach.
Intervention	One or more activities, carried out by an implementing organization, that involve interactions with and including products, services, or funding delivered to, targeted groups or institutions.
Market system	The interaction of multiple market actors performing multiple market functions, including a core function (such as the demand and supply of financial services), supporting functions, and rules and norms. Broadly describes the complex and dynamic interactions between all market actors, including those from the private and public sectors.

Measurement plan	Complements the strategy as it provides more detailed information at the project level: defining what is to be measured, how, when, by whom, and at what cost.
Measurement strategy	Provides broad and long-term guidance for measurement at the level of an organization or large program. Principally covers matters relating to measurement throughout the organization or program, such as principles and standards, standardized indicators, and program-level evaluation.
Measurement system	Consists of frameworks (e.g., a ToC or results chain, a results framework, and a risk register), processes (e.g., for data collection and management, review and adaptation, and evaluation), and their related tools (e.g., data collection instruments, reporting templates, data processing, and storage applications).
Outcome	A change in an organization or persons brought about, at least partly, by an intervention .
Portfolio	A set of programs in a given development domain or country or region.
Program	A coherent set of projects with a common, but often broad, focus. Usually has a longer life than any single project within it.
Program unit (or implementer)	Units implementing financial inclusion programs; may be operated by funders themselves, but more typically they are found in Financial Sector Deepening Trusts, intergovernmental organizations, international or national nongovernmental organizations, educational or research organizations, consultancy companies, or government agencies. Commonly referred to as implementer or implementing partner.
Progression narrative	The description (explicit or implicit) of the transition from one level of a theory of change or results chain to the next.
Project	A tightly focused set of interventions with a common work plan, managed by a single implementing agency.
Results chain	A sequence of changes, usually in the form of a diagram, that are expected to be influenced by specific interventions in a project . The expected changes— outcomes —along with the interventions and the assumptions behind the changes, are more context-specific and narrowly focused than in a program theory of change. (See Module 3.)
Results framework	A planning and results measurement instrument, usually in matrix form. Includes a summary of a program or project's intended interventions and outcomes, together with measurement-related information, such as indicators, means of obtaining data for these indicators, baselines, and targets for the indicators. May also contain assumptions and risks. (See Module 5.)
Rules and norms	Shape incentives for market actors and determine who can participate in a financial system and under what conditions. Formal rules include laws and regulations issued by the legislator and public authorities (e.g., banking regulation, licensing criteria for financial institutions, or know-your-customer procedures). Also includes rules issued by industry bodies (e.g., industry standards or codes of conduct). Informal rules are the product of local culture and generally accepted practices.
Supporting functions	A range of functions that fall outside of the core of a market system, but that significantly affect the strength or weakness of that market. Support, shape, inform, and enable transactions between demand and supply actors.
Systemic approach (or market development approach)	An approach to development that aims to catalyze systemic change . It aims to promote inclusivity, scale, sustainability, and resilience in the financial system and to have positive development impacts. Key features include the role of facilitation, adaptive management, and monitoring of a long-term view. Also referred to as a market development approach, a market systems approach, a market facilitation approach, or making markets work for the poor.
Systemic change	A change in the underlying dynamics of a financial market that leads to significant change in terms of scale, sustainability, and resilience of the financial system. Systemic change occurs if market players beyond those interacting with program interventions adopt new practices that improve poor people's participation in the financial services market.
Theory of change	A set of pathways from interventions to different levels of expected outcomes, usually presented in the form of a diagram. Progress along the pathways is based on contribution hypotheses and is governed by assumptions, the most important of which are made explicit. In this Handbook, the term is used in connection with programs rather than projects (see results chain). (See Module 3.)

SELF-ASSESSMENT TOOL FOR FINANCIAL INCLUSION MEASUREMENT SYSTEMS

The Self-Assessment Tool (SAT) is designed to help funders and program managers **improve the measurement strategies, systems, and practices used in financial inclusion portfolios, programs, and projects**, particularly those using systemic approaches.

The composition of the tool

SAT is based on a set of building blocks. For each building block, SAT proposes a series of assessment questions (first column of the tool's grid). The aspect of measurement covered by a question is then scored according to a standard methodology that is set out in the second column. Depending on the score, brief action steps or other comments are entered in the third and final column, to be followed up later with more detailed action planning.

SAT is closely aligned to CGAP's *Measuring Market Development Handbook*. Although some terms are defined in footnotes in SAT, more comprehensive explanations of concepts and terms are given in the Handbook.

A few questions in the tool are in bold to indicate that they are critical for effective measurement. However, it is important to view the blocks as interdependent and every question as potentially relevant.

Questions that are closely related are separated by broken lines.

Audience

A majority of questions in SAT are oriented around implementation and will be more relevant to **program units**, whether they are funders themselves or external implementers (e.g., facilitators, consulting firms, grantees of

funders, or government bodies). However, funders could also benefit from collaborating with program units in these assessments. Furthermore, some of the questions will be relevant to funders' own "upstream" measurement systems, including portfolio-based monitoring and evaluation. It is important that the SAT process is sanctioned by, and preferably has the participation of, senior management. Changes triggered by the process may have resourcing and other strategic implications.

As a whole, SAT proposes a comprehensive set of measurement practices. It is unlikely that all of the practices will be highly relevant to every user. Furthermore, some will be relevant but not a high priority for action (e.g., because of the stage in the program cycle). For example, the development of ToCs and results frameworks are most relevant at the beginning of a cycle; detailed evaluation planning will be more relevant once the program is well under way.

SAT enables the collaborative identification of priorities for improvement and the triggering of action. SAT could be conducted with or without external stakeholders, and with or without external expert facilitation.

Scoring

Assessors should consider two aspects of each of the questions.

First they should consider the extent to which the question is **relevant** to their portfolio, program, or project, given the scope of the assessment under way. The SAT scale for relevance is **A** = highly relevant; **B** = moderately relevant; **C** = not relevant.

If a question is considered "not relevant" in the context of the review, there is no need to

go further. If the question is considered to be fully or moderately relevant in the assessment context, assessors should consider the degree of **fitness** of measurement in the portfolio, program, or project in relation to the question.

Fitness has two dimensions. The first is the extent to which the portfolio, program, or project is equipped with any necessary **frameworks, processes, or tools** in relation to a specific assessment question. The second is **human capacity**: Are there sufficient staff with the necessary knowledge, skills, and mind set in relation to the assessment question? The SAT scale for **fitness** is **3** = fully fit (no action needed on frameworks, processes, tools, or capacity); **2** = moderately fit (some action needed); **1** = unfit (considerable work needed).

A score for a question combines the two elements (e.g., a score of A2 would mean highly relevant and moderately fit, and therefore some action is needed). Where action is needed, the next steps and any other comments should be recorded in the last column of the tool. The scale of action and the urgency may depend on whether the relevant score is A or B. Action may mean putting in place or improving frameworks, processes, or tools; it may also mean supporting human capacity development.

SAT building blocks

SAT comprises 13 building blocks that are widely considered to be essential for an effective measurement system.

1. Key concepts

There is no clear consensus on some concepts, such as financial system, systemic change, and financial inclusion, which are nevertheless important for funders to clarify to be able to effectively measure progress in market system development for financial inclusion. Funders need to have a plausible model of the system and concepts of systemic change and financial inclusion that are clear and that stakeholders have bought into. To be able to compare progress across interventions, the model and definition of the concepts should be applied consistently in ToCs and results frameworks. SAT lists questions around key concepts to check whether these are properly defined and communicated to relevant program or project stakeholders.

For more information on these key concepts see Module 1 and Module 3.

2. The measurement system

Measurement needs frameworks (e.g., ToCs, sets of measurement questions, and results frameworks) and processes (e.g., for data col-

lection, data management, review, and evaluation). They need to be seen as a system, interdependent, and connected. SAT maps the key components of the system with available resources, measurement status and location in the organization, capacity building for measurement (including at the partner level), and linkages with reporting and the way evidence produced by the system influences decision-making.

For more information, see Module 2.

3. The environment for measurement

The measurement system should be set in a conducive environment that enables and incentivizes staff to participate fully in measurement activities. This is particularly important for systems in financial inclusion programs that adopt systemic approaches—where measurement requires innovation and initiative by a range of program staff. SAT covers support received by management, incentives for staff to participate, collaboration between funders and program units, and a broader notion of accountability.

For more information, see Module 1 and Module 2.

4. Measurement strategies and plans

Organizations, portfolios, and large programs need measurement strategies that address measurement matters that transcend individual projects. These may include measurement policy, if that policy is not set out elsewhere. Measurement strategies should be complemented by detailed and regularly updated project measurement plans. The boundaries between the two instruments will vary. SAT covers how measurement strategy, plans, and questions were developed and their content.

For more information, see Module 2.

5. Theories of change (ToCs) and results chains

ToCs and results chains play foundational roles in measurement and link with program and project planning for financial inclusion. They pave the way for results frameworks and are the platform for enhanced monitoring and evaluation, particularly theory-based. ToCs and results chains can be viewed as similar instruments: ToCs are broader in scope and less context-specific than results chains and are, therefore, found in programs rather than projects. SAT covers the quality and relevance of ToCs and results chains and the way they are used (hypothesis rather than blueprints) and updated.

For more information, see Module 3.

6. Measurement questions

Measurement questions clarify what we want to learn about our programs, projects, and portfolio of programs. They help to define the scope and direction of monitoring and evaluation. Framing measurement questions at the beginning of program and project cycles is important.

For more information, see Module 4.

7. Results frameworks: Structure

Results frameworks are the most practical ways to summarize the chain of expected results and the means of measuring them. The frameworks need to reflect the related ToC or results chains. This means allowing for more than one level of outcome and a reasonable number of results at each level. The most common version of a results framework is the logical framework or logframe. This building block covers the structure and terminology of the results framework.

For more information, see Module 5.

8. Results frameworks: Indicators and targets

This building block looks inside the results framework at indicators and targets. It assesses how indicators and targets are selected and reviewed, how relevant they are, and whether the framework links to adequate data sources.

For more information, see Module 5 and Module 6.

9. Data sourcing

Sourcing data for the measurement of financial inclusion programs presents few conceptual but many practical difficulties. Data are needed in three stages: before the program or project begins, during the program or project, and after it has been completed. SAT assesses data quality, including baseline data, and the processes for collecting them.

For more information, see Module 5 and Module 6.

10. Monitoring, review, and adaptation

Market systems, especially immature financial systems, are dynamic and unpredictable. Because of this project monitoring takes on additional responsibilities compared to traditional approaches. Monitoring based exclusively on results framework indicators is inadequate for market system development and similar program approaches, and the results can be misleading. Enhanced monitoring that goes beyond tracking predefined indicators and straddles the boundaries between monitoring and both formative evaluation and risk management should be considered. Indicators need to be complemented

by the monitoring of a wider range of results data: qualitative aspects, unplanned and unexpected outcomes, assumptions and risks, and other aspects of the external environment that may affect the intervention. SAT covers enhanced monitoring's content and processes, including feedback into intervention design and management.

For more information, see Module 4 and Module 7.

11. Reporting and knowledge sharing

Reporting to stakeholders plays a big part in measurement. It is fundamental to accountability, which is one of the most important functions of measurement, and it can feed stakeholders' learning needs and appetites. Combining these two functions need not be a challenge, but other outreach channels may need to be used to reach broader audiences.

For more information, see Module 8.

12. External evaluation

External evaluation is a core component of the measurement system because it adds value in several areas, including an unbiased standpoint to calibrate results data, assess contribution, validate performance for accountability, and provide expertise and experience to turn findings into recommendations for change. Evaluation is expensive in terms of time and other resources and needs to be led by robust use cases. Evaluation should be closely integrated with monitoring for the lower levels of the ToC because external evaluators bring a detached perspective to bear on the program staff's deep knowledge of the intervention environment. Collaboration in, and reciprocal sharing of knowledge from, evaluation is particularly important in the relatively unmapped territory of market system development programs for financial inclusion. SAT covers the policy governing external evaluation, the type of methodologies used, integration with monitoring, and collaboration with other stakeholders.

For more information, see Module 4, Module 7, and Module 9.

13. Portfolio reviews (only for funders)

Portfolio-based monitoring and evaluation, based on aggregation or systematic scoring of information on programs and their performance, can be useful to funders for both learning and accountability. SAT encourages funders to check whether there is a regular, systematic, and "smart" monitoring of interventions at the portfolio level, and whether findings are applied in decision making.

For more information, see Module 10.

Building block 1. KEY CONCEPTS		
Assessment questions	Scoring	Next steps and/or comments
1.1 Do we have a clearly defined model of a <i>financial market system</i> that we use consistently across our program(s) and, where relevant, share with key stakeholders?		
1.2 To what extent is the model clearly identified in the program ToC? ^a		
1.3 Do we have a definition of systemic change that we use consistently across our program(s) and, where relevant, share with key stakeholders?		
1.4 Is the definition evaluable (e.g., is it possible to represent it in a TOC or results chain and translate it into indicators)?		
1.5 Do we have a definition of <i>financial inclusion</i> that we use consistently across our program(s) and, where relevant, share with stakeholders?		
1.6 Is the definition evaluable (e.g., does it easily translate into indicators)?		

a. "Theory of Change" covers constructs that may also be called results chains, logic models, or impact pathways. Results frameworks—such as logframes—also contain indicators and other measurement information. For more information, refer to modules 3 and 5.

Building block 2. MEASUREMENT SYSTEMS		
Assessment questions	Scoring	Next steps and/or comments
2.1 Is there a clear picture of what the measurement system is composed of and how it is connected?		
2.2 Is this picture shared with stakeholders who need to know?		
2.3 Is the system adequately resourced and managed?^a		
2.4 Is system management appropriately positioned in the organizational structure to give it the necessary credibility, authority, and access?		
2.5 Is the system proportionate in terms of depth and cost in money and human resources? ^b		
2.6 Are the data collected adequately managed within the system (e.g., quality assured, accessible where and when needed, effectively analyzed and presented)?^c		
2.7 Given that sustained changes at the financial system level may not be detectable until after an intervention is over, are there institutionalized processes for continuing to monitor beyond the life of specific interventions? ^d		
2.8 Is evidence produced by both monitoring and evaluation systematically reviewed at a strategic level and within projects? Are findings and conclusions purposefully applied to new decision-making?^e		
2.9 Is there a systematic approach to identifying and monitoring risks in both the delivery and outcomes of interventions? ^f		
2.10 If risks are identified, is appropriate action taken (e.g., adjust the intervention to mitigate or avoid the risk)?		

Building block 2. MEASUREMENT SYSTEMS, <i>continued</i>		
Assessment questions	Scoring	Next steps and/or comments
2.11 Does internal and external reporting flow from review processes, rather than being conducted as an isolated exercise without learning benefits? ^g		
2.12 Is reporting to different stakeholder groups sufficiently streamlined to avoid unnecessary duplication of effort?		
2.13 Is capacity building for measurement sufficiently resourced and institutionalized? ^h		
2.14 Does this include capacity building for stakeholders (e.g., project partners) who are engaged in measurement with us?		
2.15 Is it backed by accessible, user-friendly tools, templates, and written guidance?		
2.16 Are there ready sources of human support, including peer-to-peer (e.g., through a community of practice)?		

- a. All measurement systems need resources and need to be managed. In small projects, this can sometimes be done without specialist expertise and dedicated measurement. But this is unlikely to be the case with larger, more complicated interventions.
- b. The measurement system needs to meet defined needs and no more. Overly elaborate measurement requirements are unsustainable and can damage the whole system.
- c. The quality of data—both in its raw state and in the way it is analyzed and presented—affects its reliability and use. Data management is an important part of any system.
- d. One of the benefits of having a program-wide measurement strategy is that data collection can continue beyond the life of individual projects to support impact evaluation.
- e. The most consistently highlighted weakness in measurement systems is in the application of evidence to new decision-making. Yet this is probably the most important reason for measurement.
- f. Identifying and monitoring significant risks are essential for effective program delivery and outcomes. They are also valuable processes for learning about what is feasible and what promotes progress in interventions, especially in the relatively unmapped territory of financial inclusion.
- g. Reporting is one of the least loved activities in the intervention or institutional cycle. It is often seen solely as a compliance activity. Constructing reports as a consequence of a collaborative review process can inject learning into report-writing and improve the quality of the reports. It is often possible to avoid multiple report formats and timing through negotiation with report stakeholders.
- h. Measurement does not come naturally to many of the staff involved. It requires investment in regular capacity-building and other support mechanisms. This is particularly true of partners who may not have any background in measurement. A community of practice can help both specialist measurement staff and program staff without specialist support from feeling isolated.

Building block 3. ENVIRONMENT FOR MEASUREMENT		
Assessment questions	Scoring	Next steps and/or comments
3.1 Is measurement embedded in program management and not seen as a technical function done by measurement specialists? ^a		
3.2 Does senior management visibly promote measurement? ^b		
3.3 Are adequate incentives applied to staff to collect data systematically and to take time to review it? ^c		
3.4 Are staff encouraged to identify weaknesses in implementation and results?		
3.5 Are staff encouraged to take reasoned risks, even if the chances of failure might be considered high?		
3.6 Is there a collaborative approach to measurement between funders and program units? If not, are steps being taken to improve them? ^d		

continued

Building block 3. ENVIRONMENT FOR MEASUREMENT		
Assessment questions	Scoring	Next steps and/or comments
<p>3.7 Are the following program factors considered in interactions about accountability between program units and funders?^e</p> <ul style="list-style-type: none"> a. Designing well-formulated projects based on thorough diagnostics and with an appropriate level of ambition b. Putting in place and operating a proportionate full-cycle measurement and learning system c. Systematically identifying, monitoring, and managing risks d. Having processes that ensure the quality of activities and outputs and delivery at a reasonable cost e. Being a proportionate contribution to valued outcomes f. Delivering robust cases for results frameworks 		

- a. Measurement does not function properly as a technical add-on, particularly in market systems development programs. It should be part of mainstream intervention management.
- b. As with many other activities in organizations, if measurement is not seen to be valued and used by senior management, it will be given a low priority by other staff. Examples: regularly demanding good-quality, timely data on both implementation and outcomes and being seen to use it in operational and strategic decisions; or supporting the production of these data by adequately resourcing and engaging with it.
- c. Many aspects of measurement, especially enhanced monitoring, benefit from a culture of enquiry and openness. This, and the need for time to be devoted to it outside the normal routines of intervention management, requires a re-setting of incentives for staff throughout the program or organization.
- d. Funders and program units have a mutual interest in ensuring that interventions are designed and implemented effectively and that lessons are learned and carried forward. Measurement plays a big part in this. It makes sense to collaborate in ensuring measurement meets these needs and serves accountability.
- e. Program unit accountability in environments like those of most market systems development programs for financial inclusion, should not be equated solely with delivering predefined, quantitative targets. This narrow concept of accountability can even be damaging and can create inappropriate incentives and risk aversion. Accountability conceived more holistically, is likely to be more relevant and effective in developing and maintaining conducive environments and promoting good outcomes. If this approach to accountability is lacking, it should be possible for program units at least to open a dialogue about it with funders.

Building block 4. MEASUREMENT STRATEGIES AND PLANS		
Assessment questions	Scoring	Next steps and/or comments
<p>4.1 Is measurement operationalized through robust and practical measurement strategies and plans?</p>		
<p>4.2 Do measurement strategies and plans include the following?^a</p> <ul style="list-style-type: none"> a. An introduction that includes a description of key measurement issues and the measurement questions that should be addressed b. Resourcing and management details c. Relevant ToC(s) d. Results framework(s), including baselines and targets e. A risk register and clear processes and responsibilities for risk monitoring and management f. Detailed guidance on the monitoring for each indicator g. Guidance on monitoring beyond indicators (unplanned results, etc.) h. Guidance on baseline data collection i. Guidance on processes for data management j. Guidance on internal review processes, including key measurement questions k. Guidance on internal and external results reporting l. Guidance on external evaluation^b and use of evaluators m. Details of further capacity building and support for measurement in the intervention 		

- a. Some items, such as a, b, and c, are more appropriate for a measurement strategy; others, such as f, as applied to project-specific indicators will be found in plans. There is always a degree of overlap between the coverage measurement strategies and plans.
- b. This is only preliminary guidance for evaluation. Every evaluation will have a detailed evaluation plan.

Building block 5. THEORY OF CHANGE AND RESULTS CHAINS		
Assessment questions	Scoring	Next steps and/or comments
5.1 Do we have a ToC for the overall financial inclusion program and, where relevant, broad themes within financial inclusion, such as savings?		
5.2 If there is a strategy that covers financial inclusion? Is the ToC aligned with it?		
5.3 Is the ToC broad and inclusive enough to encompass the full range of interventions we anticipate delivering or supporting under our strategy?		
5.4 Do individual projects have their own results chains?		
5.5 Do they clearly relate to the program-wide ToC? If not, is there an agreed case for divergence (e.g., to pilot an innovative approach)? ^a		
5.6 Are the ToCs and results chains based on adequate evidence? Are they developed with the participation of key stakeholders? ^b		
5.7 Are the progression narratives of the ToCs and results chains plausible? ^c		
5.8 Are they based on explicit contribution hypotheses? ^d		
5.9 Are the assumptions (including the contributions of other stakeholders) behind the progression narratives explicit in the models or supporting documentation? ^e		
5.10 Have the assumptions been identified through a proper process of analysis involving relevant stakeholders?		
5.11 Are the key assumptions—or their counterparts, risks—carried forward into an actionable risk register?		
5.12 Are the building blocks of the ToCs and results chains evaluable to the extent that they can readily be converted into results frameworks and easily translated into indicators?		
5.13 Are the ToCs and results chains readily understood by program staff and key stakeholders? ^f		
5.14 Where detail is left out to prevent the ToCs and results chains from becoming too complicated, is it included in an accessible accompanying document?		
5.15 Are the ToCs and results chains presented and understood as hypotheses and not as blueprints? ^g		
5.16 Are they reviewed at appropriate intervals or after critical events and, if necessary, adjusted to reflect changes in the program strategy, external players, and program circumstances?		

- a. Projects stem from program strategies so their results chains should normally align to some extent, unless a divergence has been agreed, e.g., to test out a new area of intervention.
- b. The process of developing ToCs and results chains is important partly for their credibility and the buy-in of stakeholders; and also because their robustness depends on the quality of inputs to the process: evidence from research and evaluation where available, and stakeholder perspectives about what is plausible and feasible.
- c. The progression narratives (description of the transition from one level and its different elements to the next; it usually includes a set of “If...then” statements) and their underlying hypotheses—including assumptions about factors beyond the intervention—are what bind the blocks of the ToCs and results chains together. They need to be grounded through a proper process of analysis.
- d. A contribution hypothesis is an explanation of the logic behind progression from each output to its related outcome, and outcomes at one level to the next.
- e. Assumptions and/or risks are important in market system development programs and need to be monitored and managed effectively.
- f. ToCs and results chains are valuable as communication vehicles both internally and externally. They should not be too complicated. At the same time they should attract credibility through not appearing simplistic or over-deterministic. Balancing these two considerations may be challenging but is always worthwhile.
- g. ToCs are designed to be reliable, but in complex and unpredictable territory like market systems development for financial inclusion, they are bound to be hypothetical and will need to be tested and adapted if found wanting.

Building block 6. MEASUREMENT QUESTIONS		
Assessment questions	Scoring	Next steps and/or comments
6.1 Are comprehensive sets of measurement questions that need to be addressed by monitoring and different types of evaluation developed at the beginning of program and project cycles? ^a		
6.2 Whenever measurement questions are developed, do they take the OECD-DAC evaluation criteria into account? ^b		
6.3 Do they relate closely to the ToC and its explicit or implicit contribution hypotheses and assumptions?		
6.4 Are measurement questions developed jointly between funders and implementers and, where relevant, with the participation of other stakeholders? ^c		

a. Measurement questions set the overall agenda for measurement of a project or program. It is important to frame them at the beginning of the program or project cycle to maximize the potential of monitoring and support evaluation by ensuring that critical data are collected and analysis is undertaken at optimal times.

b. The OECD-DAC criteria provide the broad evaluation agenda. The ToC contextualizes it enough for measurement questions to gain traction, but without narrowing their focus as much as the results framework. OECD DAC: <http://www.oecd.org/dac/evaluation/dacriteriaforevaluatingdevelopmentassistance.htm>

c. Developing the questions jointly helps to promote coherence and align expectations on accountability and learning priorities.

Building block 7. RESULTS FRAMEWORKS: STRUCTURE		
Assessment questions	Scoring	Next steps and/or comments
7.1 Do results frameworks exist for interventions at both project and program level?		
7.2 Are the results aligned with the program ToC and project results chains?		
7.3 Are the frameworks proportionate in their size and complexity to those aspects of the project or program?		
7.4 Is the structure (the number of levels, the permitted number of outcomes at each level) of the frameworks practical and easy to work with?		
7.5 Is the terminology of the different levels of the framework (e.g., outputs, intermediate outcomes) used consistently across the program? ^a		
7.6 Is the rationale for the levels and their terminology understood and appreciated by stakeholders?		

a. There is no universal language for results framework terminology and this often generates debate and uncertainty. To reduce the uncertainty in the circles in which it moves, every organization or program needs to base its own terminology on a clear rationale, explain it to stakeholders, and use it consistently.

Building block 8. RESULTS FRAMEWORKS: INDICATORS AND TARGETS		
Assessment questions	Scoring	Next steps and/or comments
8.1 Where there are preidentified measurement questions that point to important aspects of the expected results, are they reflected in the results framework indicators? ^a		
8.2 Have the indicators been developed by staff who are involved in and have a good understanding of the intervention? ^b		
8.3 Are the indicators relevant to the related results? For example, is there at least one indicator for each distinct element of a result?		
8.4 Are qualitative indicators used to complement—or instead of—quantitative indicators where the latter are not able to tell the story adequately? ^c		
8.5 Are steps taken to exclude indicators of marginal relevance? ^d		
8.6 Is there appropriate alignment between project and program indicators? ^e		
8.7 Where relevant, is there appropriate alignment (results and indicators) with frameworks of other stakeholders (e.g., between programs and funder portfolios)?		
8.8 Are the indicators precise enough to be understood in the same way by different stakeholders?		
8.9 Are the indicators linked to accessible, reliable, and affordable data sources and instruments? ^f		
8.10 Where there are targets, have they been established through an evidential process? ^g		
8.11 Are current targets and milestones appropriate and feasible, reflecting what is likely to have been achieved at that point in the intervention? ^h		
8.12 Are the contents of results frameworks reviewed at appropriate intervals and adjusted, where necessary, so that projects and programs do not get locked into inappropriate or out-of-date results frameworks?		

- a. Measurement questions can suggest nuances for indicators because of themes such as gender, which are of particular interest and importance, but which may not be explicit in the ToC outputs or outcomes themselves. They can also point to where indicators need to be particularly intensive because of a high degree of interest in the result.
- b. Indicators are central to most measurement systems. In complex environments like market system development, they need to be based on solid knowledge and understanding of where change is likely to happen. Participation is also important for buy-in.
- c. Results cannot always be adequately captured through quantification. Changes in a single policy, institutional structure or process, or the behavior of a group, for example, are often better described in words.
- d. Indicators require much effort and other resources. It is important that they carefully target areas where they are most needed.
- e. It is important to avoid imposing standardized indicators where they are not sufficiently relevant to the intervention because this can lead to inappropriate incentives and distortion in implementation.
- f. Although measurement plans will provide greater detail on data sources and instruments like surveys, it is important to establish that these sources exist and that they can be accessed through data collection instruments, reliably, and affordably, before the indicator is “signed off”.
- g. For example, have baseline data and information on the trend leading to the baseline been collected? Have there been consultations with stakeholders who have a good understanding of what is feasible, and of how reliable the assumptions behind the targets are? Targets create expectations. They need to be realistic.
- h. In market system development programs, milestones on the road to end-line results, such as increased access to services, may need to be about changes in behavior and institutional structure.

Building block 9. DATA SOURCING		
Assessment questions	Scoring	Next steps and/or comments
9.1 Are baseline data collected for indicators and other areas where change is expected to be measured? ^a		
9.2 Are the baseline data collected early enough in the intervention cycle to be valid?		
9.3 Are steps taken to avoid collecting baseline data where there is not a clear use case? ^b		
9.4 Where data—whether for baselines, ongoing monitoring, or evaluation—are collected from samples of a population, is it ensured that they are sufficiently representative of the population for the purpose at hand? ^c		
9.5 Where secondary sources are relied on, are they sufficiently up-to-date and granular for the purpose at hand? ^d		
9.6 If they are not, are steps taken to influence the timing and granularity of these sources, or to fill in the gaps with primary data collection? ^e		
9.7 Where existing data are not available and sourcing data ourselves would be unaffordable or disproportionate, is collaboration with other stakeholders (e.g., funders, government agencies) considered to generate the data, either one-time or regularly?		

a. Baseline data are essential if we are to measure change, e.g., in the status of stakeholders or institutions that already exist at the start of the intervention.

b. To avoid data overload, it is important to be clear what changes we are likely to want to measure before planning baseline data collection.

c. Data sampling is governed by well-established techniques. Standards need to be observed if the data are to be relied on and credible.

d. At the higher levels in financial inclusion ToCs, secondary sources are often the most practical option for measurement.

e. Secondary sources may not give us the data when and in the form we need. It is important to research this at the beginning of the intervention cycle and consider investing in new data sources or extensions to existing ones. Collaboration may be the most feasible way of doing this.

Building block 10. MONITORING, REVIEWING AND ADAPTING		
Assessment questions	Scoring	Next steps and/or comments
10.1 Is there systematic monitoring of indicators?		
10.2 Is there systematic monitoring of qualitative aspects of planned results (outputs and outcomes) not adequately captured by the indicators in the results frameworks?		
10.3 Is there systematic monitoring of outcomes that are unplanned and unexpected, both positive and negative?		
10.4 Is there systematic monitoring of identified assumptions or risks and other aspects of the external environment that may matter for the interventions? ^a		
10.5 Is there regular and systematic review both of the implementation and results of interventions?		
10.6 Do relevant stakeholders participate in the review processes?		
10.7 Do relevant measurement questions drive the reviews?		
10.8 Is relevant available evidence assembled and presented in these reviews in usable formats?		

continued

Building block 10. MONITORING, REVIEWING AND ADAPTING		
Assessment questions	Scoring	Next steps and/or comments
10.9 Are there clear criteria and a robust process, acceptable to key stakeholders, for adapting interventions based on review findings?^b		

- a. The external environment will affect the intervention and its outcomes both positively and negatively, predictably and unpredictably. These external factors need to be monitored mainly through the lens of identified assumptions and risks. During the review process it is important to keep an open mind about what is driving results that may appear at first sight to have been primarily influenced by the intervention.
- b. Market system development interventions need to keep pace with the dynamics of their environment. Data collected through monitoring needs to be regularly reviewed, and the conclusions of the reviews should lead, through due process, to adaptations to the intervention itself and its results framework.

Building block 11. REPORTING AND KNOWLEDGE SHARING		
Assessment questions	Scoring	Next steps and/or comments
11.1 Is reporting designed to serve both accountability and stakeholders' learning needs and interests?		
11.2 Have measurement questions been taken into account in determining the scope and content of reports? Alternatively, have key stakeholders been consulted recently on their information and knowledge needs and interests that could be served by the reports?		
11.3 Do progress reports focus on the results framework as well as provide ample space for reporting beyond planned results, including unplanned outcomes and relevant developments in the external environment?		
11.4 Do reports present a plausible narrative of how the program has contributed to change?		
11.5 Do reports describe what lessons have been learned and how they are to be applied going forward?		
11.6 Is the frequency and scope of reporting required from projects to programs and from programs to funders proportionate and in line with how that information will be used?		
11.7 Is reporting integrated into cycles of review, so that reporting flows in a timely way from evidence-informed reviews, and is not prepared in isolation?		
11.8 Is knowledge stemming from review and evaluation shared with other stakeholders and broader audiences through appropriate dissemination channels?		

Building block 12. EXTERNAL EVALUATION		
Assessment questions	Scoring	Next steps and/or comments
12.1 Is there an up-to-date and relevant policy or strategy governing external evaluation?^a		
12.2 Are there criteria and a robust process for ensuring that there is a clear use case for every evaluation?		
12.3 Are those criteria and that process followed consistently?		
12.4 Are evaluations based on relevant ToCs and measurement questions?^b		
12.5 Where use is made of experimental methodologies, is it appropriate?^c		

continued

Building block 12. EXTERNAL EVALUATION		
Assessment questions	Scoring	Next steps and/or comments
12.6 Do theory-based evaluations: ^d <ul style="list-style-type: none"> a. Capture and assess data about both anticipated and unanticipated outcomes? b. Capture data from a sufficient number of sources, both quantitative and qualitative, to enable “triangulation”? c. Use credible strategic informants? d. Establish plausible connections between the interventions and the changes identified? e. Capture data about other interventions and contextual factors, to explore other possible contributions to the changes identified? 		
12.7 Is evaluation closely integrated with monitoring?		
12.8 Have external evaluation staff been used, when relevant, to facilitate internal reviews and/or conduct developmental or real-time evaluation? ^e		
12.9 Is there collaboration with other stakeholders in evaluating progress in market system development?		
12.10 Are findings and conclusions from our evaluations being shared with stakeholders, including key national actors?		

- a. This is best included in the type of measurement strategy described Building Block II.
- b. If ToCs and/or measurement questions do not exist, they need to be created in the evaluation inception period. If they exist, they should be revisited and refined, if necessary.
- c. Experimental methodologies have limited roles in systemic programs. They are more likely to be used in testing links between financial inclusion and poverty reduction.
- d. TBE is the most appropriate methodology for market system development programs for financial inclusion. Rigor depends on several approaches, including combining top-down with bottom-up perspectives.
- e. The terms “developmental” and “real-time” evaluation are used to describe evaluation work that takes place during project and program implementation, is conducted iteratively, and where the emphasis is on immediate lesson-learning and its application, rather than on impact assessment or accountability.

Building block 13. PORTFOLIO REVIEWS (funders only)		
Assessment questions	Scoring	Next steps and/or comments
13.1 Is there regular and systematic monitoring of the portfolio of financial inclusion interventions at the portfolio level, using an instrument such as a portfolio dashboard, to see if the interventions are going in the right direction?		
13.2 Is there periodic (e.g., once every 2–4 years) evaluation of the portfolio against a strategy for financial inclusion?		
13.3 Is there a system that generates sufficient relevant data to enable comparisons of interventions across the portfolio? ^a		
13.4 Are these comparisons applied “smartly,” bringing context and complexity into play? ^b		
13.5 Are insights from the reviews applied in new strategic decision making?		

- a. Portfolio reviews are more effective if their needs are anticipated at the beginning of intervention cycles through categorization in intervention plans and in the use of standardized expression of results and indicators, where this is appropriate. (See Building Block 8.)
- b. Comparisons across market systems development interventions can be difficult and even invidious. It is important to treat standardized results as the entry points for further analysis and interpretation.



Consultative Group to Assist the Poor
1818 H Street, N.W., MSN IS-700
Washington, DC 20433 USA
Phone: +1 202 473 9594
cgap@worldbank.org
www.cgap.org
www.microfinancegateway.org

