

# Guidance for **EVALUATING** Integrated Global Development Programs



#### ACKNOWLEDGMENTS

This guidance document is the product of many thoughtful and knowledgeable contributors. We are incredibly grateful for every contribution, large and small. First, we thank the primary authors, who spent many hours thinking creatively on this topic:

Tessa Ahner-McHaffie Aurélie Brunie, PhD, MS Mario Chen, PhD, MS Gina Etheredge, PhD, MA Greg Guest, PhD, MA Theresa Hoke, PhD, MPH Rick Homan, PhD Christine Kim, MSPH Tricia Petruney, MA Merywen Wigley, MSc

Next, we thank the editors: Kerry Aradhya and Natasha Mack.

We also express our extreme gratitude to the reviewers for their thoughtful and constructive comments: Sohail Agha, Laura Ahearn, Joy Noel Baumgartner, Aurelie Brunie, Cindy Clapp-Wincek, Elizabeth Larson, Kate MacQueen, Mike Merrigan, Emily Namey, Hallie Preskill, Diana Rutherford, Pamela Velez-Vega, and Lucy Wilson.

The participants of a workshop on evaluating integration merit our thanks as well: Sohail Agha, Laura Ahearn, Joy Noel Baumgartner, Mario Chen, Cindy Clapp-Wincek, Bob Engelman, Samuel Field, Nathanael Goldberg, Paul Hutchinson, Daniel Kidder, Elizabeth Larson, Katherine Lew, Kristin Lindell, Kate MacQueen, Shira Mitchell, Veronica Olazabal, Hallie Preskill, Kyle Rearick, Laura Reichenbach, and Joseph Sany. Their keen insight and discussion laid the foundation for this document.

Finally, we thank the Bill & Melinda Gates Foundation for providing the guidance and funding to develop this document.

SEPTEMBER 2016

#### ACKNOWLEDGMENTS ABBREVIATIONS

#### **EXECUTIVE SUMMARY**

INTRODUCTION... Purpose of This Guidance Structure of This Guidan What Is Integrated Devel Types of Integration The Call for Increased an What is Unique about Eva

////

### BEFORE IMPLEMENTATION

# DURING

IMPLEMENTATION

#### 6 | EV Experi AFTER

IMPLEMENTATION

CONCLUSION

APPENDIX A: RESOUR **GUIDANCE ON EVALUA** APPENDIX B: PLANNIN HEALTH PROGRAM AN GLOSSARY

# **CONTENTS**

MENTS S	11 4
MMARY	5
4	15
Guidance Document	16
s Guidance Document	16
ed Development?	17
tion	21
eased and Improved Evaluation of Integrated Programs	21
about Evaluating Integrated Development?	23
1 FORMATIVE RESEARCH	25
Mapping the Stakeholder Environment	27
Mapping the System	30
Informing Program Design and Implementation	31
2   DEVELOPING PERFORMANCE INDICATORS	
Sector-Specific Indicators	40
Harmonizing Indicators	40
Value-Added Indicators	41
3 PROGRAM MONITORING	45
Methodology	47
Monitoring Processes and Teams	48
Performance Monitoring Plans and Data Collection Tools	49
Analysis and Reporting	50
4   PROCESS EVALUATION	51
Methodology	53
Stakeholders	54
Team Composition	55
5   COST ANALYSES	57
Design Phase	60
Preparation Phase	61
Operational Phase	61
Using Cost Data for Evaluation	62
Cost-Effectiveness versus Cost-Benefit Analysis	64
6   EVALUATING IMPACT	67
Experimental	70
Quasi-Experimental	73
Case Study	76
Systems Approaches	78
Mixed-Methods Approaches	79
7   EVALUATING SCALE-UP	
Focus of Evaluations for Scale-Up	88
	01
	91
ESOURCES, LITERATURE, AND	06
	30
AM AND DESIGNING ITS EVALUATION	99
	101

# **EVALUATION IN A PROGRAM LIFECYCLE**





DURING IMPLEMENTATION



The modules in this guidance document reflect various types of evaluative activities. For the most part, they reflect a logical progression of evaluative activities and processes within a program's lifecycle using a before, during, and after implementation structure. Nevertheless, many of the activities discussed can happen simultaneously and be interdependent. Although these activities have a logical progression (and phases wherein they are most important), time and thought are ideally devoted to the planning and results of each activity throughout implementation.

3



#### ABBREVIATIONS

ASPIRES	Accelerating Strategies for Practical	OFSP	orange-fleshed sweet potatoes
	Innovation and Research in Economic Strengthening	PHE	population-health-environment
ANC	antenatal clinic	PMP	performance monitoring plan
CAS	complex adaptive system	RAIN	Realigning Agriculture to Improve Nutrition
CHW community hea	community health worker	SC	significant change
FHI 360	the organization formerly known as FHI (Family Health International)	SCALE+	System-Wide Collaborative Action for Livelihoods and Environment
HIV	human immunodeficiency virus	SDGs	Sustainable Development Goals
IMAGINE	Improve the Education of Girls in Niger	SRH	sexual and reproductive health
M&E	monitoring and evaluation	TOC	theory of change
MSC most significant change	USAID	U.S. Agency for International Development	
OECD-DAC	Organization for Economic	WASH	water, sanitation, and hygiene
	Development Assistance Committee	WSR	whole-system-in-the-room

### LIST OF TABLES

Table 1. Cross-Sector Stakeholder Matrix (with Illustrative Example of a Project Integrating Activities in the Nutrition and Agriculture Sectors)
Table 2. Experimental and Quasi-Experimental Designs in an Integrated Development Context
Table 3. Case Study and Systems Approaches in an Integrated Development Evaluation Context8C
Table 4. A Comparison of Quantitative and Qualitative Approaches

### LIST OF FIGURES

Figure 1. Conceptual framework of anticipated, desired, or potential benefits of integrated development programs (as compared to vertical programs)	: 19
Figure 2. Example of conceptual framework	20
Figure 3. Common types of integration	22
Figure 4. Causal loop diagram of a health, agriculture, and economic subsystem	31
Figure 5. Impact pathway for Mama SASHA program	34
Figure 6. Measuring value-added	42
Figure 7: The cost and effects of improving WASH in schools	64
Figure 8. Integrated four-arm full factorial experimental design	71

# LIST OF PRACTICAL EXAMPLES

Practical Example 1. A Whole-System-in-the-Room Approach
Practical Example 2. Harmonizing Indicators through Stakeholder Meetings
Practical Example 3. Amplified Effects
Practical Example 4. Network Analysis
Practical Example 5. Process Evaluation of an Integrated Family Planning and Microfinance Program
Practical Example 6. HIV and Sexual and Reproductive Health Integration Cost Analysis
Practical Example 7. ASPIRES Full Factorial Impact Evaluation Design72
Practical Example 8. Quasi-Experimental Fractional Factorial Design
Practical Example 9. Impact Evaluation to Determine Potential for Scale-Up
Practical Example 10. PHE Scale-Up

# Executive Summary

### INTRODUCTION

#### Integrated development is an

intentional approach that links the design, delivery, and evaluation of programs across sectors to produce an amplified, lasting impact on people's lives. Whether explicitly or implicitly, integrated approaches are based on the premise that the interaction between interventions from two or more sectors will generate benefits beyond a vertical intervention, such as improved outcomes or operational benefits. Evaluation is a valuable tool in making evidence-based judgements about the comparative value of integrated versus vertical programming and therefore about whether and how to implement integrated development programs. Importantly, the desired goal of integration varies greatly depending on the perspective, priorities, and ultimate aim of a given decision maker. Funders may emphasize cost efficiencies or enhanced sustainability, whereas program implementers may prioritize time savings, improving user satisfaction, or reducing inequality. Therefore, the evidence they require for informed decision making will vary in nature.

Yet relative to vertical development programs, integrated approaches to development are more complex

in design; more complicated in implementation (with a greater number and more diverse range of actors involved); more diverse in terms of inputs, outputs, and outcomes; and innovative, and therefore possibly requiring more adaptive or emergent thinking. They also bring together different cultures of research and evaluation among sectors. These characteristics have implications for how integrated models are assessed and evaluated with respect to the questions being asked and the methods and designs used to answer those questions.

The purpose of this document is to provide evaluators, funders, and development practitioners with guidance on evaluating integrated development programs. Although general "good evaluation practices" are woven throughout its contents, the document focuses on the unique characteristics of evaluating the complexities associated with integrated, multi-sector program implementation. These unique aspects — centering on the concept of interaction across sectors and activities — provide the common thread that ties this guidance document together.

# **BEFORE** IMPLEMENTATION

# **1** FORMATIVE RESEARCH

**Formative research** informs program content, design, and operation. The primary task is to determine the key problem or nexus of problems, and opportunities and ideas for addressing them prior to implementation.

# **KEY QUESTIONS**

- $\rightarrow$  How can the interaction among sectors be used to enhance efficiencies, outcomes, or sustainability?
- $\rightarrow$  How do stakeholders from different sectors perceive the goals of the integrated intervention and the theory of change that underlies it? What are their expectations and how do they differ?
- $\rightarrow$  What challenges might program staff and beneficiaries face with integration?
- $\rightarrow$  How can logistical efficiencies be enhanced? Conversely, what are potential challenges and negative effects that might arise due to the interactive nature of the design?
- $\rightarrow$  What are potential effects of the integrated program on the community and the larger socioeconomic system?

Early development of an integrated theory of change and/or a logic model, or alignment of different models, can help stakeholders and evaluators identify emergent outcomes, relationships between activities, and best practices.

Inclusive participatory techniques are effective ways to build or revise a theory of change/logic model; enhance crosssector cooperation and communication; and identify potential areas of convergence, incompatibility, or unintended consequences associated with integration.

Through various methods, stakeholders can provide insights into how integration may affect: services within their sector of operation, the social structure of the larger community, local and regional governance, access to and cost of services, and dynamics within the household (such as gender or age).



Mapping the stakeholder environment and the broader system can help develop specific evaluation guestions and measures for a complex project, and allows focus on areas that may have the greatest impact.

### **2 | DEVELOPING PERFORMANCE INDICATORS**

**Performance indicators** are the precise measures used to assess success by a program or activity. Indicators for integrated programs can include sectorspecific indicators and value-added indicators. Collection of sector-specific indicators is either required or recommended for distinct programs, and can sometimes be standardized by sector or funder. Value-added indicators measure amplified effects or synergy beyond what would have occurred in a vertical program.

# **KEY QUESTIONS**

- $\rightarrow$  What indicators can be used to measure the expected integrated outcomes (amplified effects or synergies) of the program?
- $\rightarrow$  Do program designers or evaluators need to harmonize indicators across sectors?
- $\rightarrow$  What indicators can be used to measure activities in two or more sectors?

# **KEY CONSIDERATIONS**

Careful planning is needed when developing indicators for integrated programs, as the potential for greater number and complexity of indicators is high. Harmonizing indicators between activities and/or stakeholders is necessary and can be achieved by using proxy indicators, adapting data collection to streamline as much as possible, and convening stakeholders to prioritize indicators.

Meetings between experts in particular sectors, or meetings with stakeholders on specific projects, can help identify and harmonize key indicators, and can reinforce the common understanding of the intervention's outcomes and impact and how they will be achieved.

In order to choose or develop value-added indicators, program designers need to identify what "integration" means within the context of the program, and the pathways through which the program is intending to have an effect on its goals. An integrated program can use traditionally sector-specific indicators to track integration if the program anticipates that value-added will be measured through those indicators.

The most thoughtfully chosen indicators need to be paired with a monitoring and evaluation system that is designed to show relationships between the outcomes produced (the instances in which a sector's outcome indicator is the result of multiple aspects of integrated programming).

# DURING IMPLEMENTATION

# **3 | PROGRAM MONITORING**

**Program monitoring** is the routine, systematic observation and recording of program implementation and problems using the performance indicators developed and other monitoring processes, including analysis and feedback about the progress of the program to the donors, implementers, and beneficiaries (for example, through site visits and periodic stakeholder meetings).

### **KEY QUESTIONS**

- $\rightarrow$  What data collection processes or forms can be adapted to serve more than one sector's purpose?
- $\rightarrow$  Who should have access to the data to best facilitate integrated monitoring, and who should meet to discuss the data?
- $\rightarrow$  Can progress, or lack thereof, in one sector inadvertently have consequences in another sector (and how can that be monitored)?
- $\rightarrow$  How will the monitoring system track beneficiary or household access to or usage of multiple services within different sectors?

Reporting on an integrated program may be challenging because of a larger number of, and greater variation in, activities being implemented and a wider variety of stakeholders.

#### **KEY CONSIDERATIONS**

Responsibility for the collection, housing, analysis, and reporting of data needs to be clearly documented, and may cross sectors or organizations depending on how the program is designed. Different levels of integration will require different solutions to the problems of integrating monitoring systems.

Integrated programs require integrated program monitoring processes and teams. These teams would ideally comprise both monitoring and evaluation staff with specific expertise in individual sectors and monitoring and evaluation staff experienced in program integration.

It is important to cross-reference changes and identify synergistic interrelationships in order to capture cross-sector changes.

# **4 | PROCESS EVALUATION**

**Process evaluation** is a method of assessing and understanding how a program is being implemented, focusing on the program's operations and service delivery.



- $\rightarrow$  What level of integration is occurring across sectors?
- $\rightarrow$  What is the quality of the program components?
- $\rightarrow$  Can data on the implementation experience explain how any observed amplified or synergistic effects were achieved? If none were achieved, can the data explain why they were not?
- $\rightarrow$  What strategies are working for or inhibiting the cross-sector coordination or collaboration required by the program?
- → What implementation experiences may be unique to crosssector programs?
- $\rightarrow$  Are the target beneficiary population(s) being reached, and with which activities?
- $\rightarrow$  Are households or individuals accessing more than one part of the intervention (and if so, how many)?

# KEY CONSIDERATIONS

Evaluators should maintain a bird's-eye view when seeking and generating evidence, and be aware of what has been achieved by monitoring and what evidence still needs to be explored. Having a good understanding of how integration was supposed to be achieved, and the potential constraints and value added of integration is key for the process evaluator.

Feedback for the process evaluation should be sought from a multi-sectoral group of stakeholders; pains should be taken to make sure no sector dominates the evaluation. Familiarity with systems methods can be an additional asset here to identify parts of the system that need to be evaluated and explored.

# **5 | COST ANALYSES**

Cost analysis is a technique for documenting the extent to which any operational benefits occur in integrated programs and the size of these gains.

# **KEY QUESTIONS**

 $\rightarrow$  What are the costs of integrated versus vertical programming (including short- and long-term costs, cost efficiencies, financial costs or savings associated with negative outcomes or missed opportunities, etc.)?

# **KEY CONSIDERATIONS**

Costs of an integrated program will often be frontloaded as compared to standard vertical programs. The requisite expertise and time needed increases as the complexity of program design increases, although this will vary based on the type and extent of integration.

When thinking about costs, it is often useful to distinguish between different phases of a program, including design, preparation, and operational phases. Any operational benefits may not be realized until the operational phase, suggesting higher start-up costs for integrated programs.

be used.

 $\rightarrow$  What are the cost implications of additional inputs and processes necessary for the successful management, coordination, and delivery of an integrated program that are unique in comparison to the implementation of vertical programs?

Many cost analysis methods require allocating and separating all costs and measures of an outcome to particular sectors or activities. As the degree of integration increases, it is increasingly difficult and subjective to allocate funds to distinct sectors, especially for resources like labor. For this reason, cost-effectiveness analysis of integrated programs is particularly challenging unless a single effectiveness metric can

In most cases, a comparator will be required so that the costs of an integrated program can be assessed with respect to a non-integrated approach.

# AFTER IMPLEMENTATION

# 6 | EVALUATING IMPACT

An impact evaluation assesses the outcomes and impact that can be attributed to a particular intervention by comparing outcomes between intervention group(s) and a counterfactual (control group). However, evaluating impact for complex integrated programs is sometimes best served by employing a methodological perspective that extends beyond the traditional experimental and quasi-experimental design. This document adopts a definition of impact that includes a broader range of methods than included in traditional impact evaluations.



- $\rightarrow$  Were the planned amplified effects or operational benefits from integration realized?
- $\rightarrow$  To what extent were operational benefits and/or amplified or synergistic outcomes due to integration (if they were observed)? Can we demonstrate that integration led to these effects?
- $\rightarrow$  How and why did integration produce effects beyond those observed in a vertical program? If no change was measured, why did integration not produce effects?
- $\rightarrow$  Did integration result in unanticipated effects?



# **KEY CONSIDERATIONS**

Impact evaluations for integrated models should be undertaken thoughtfully, and only if there is a reasonable consensus on what models need to be tested, and what can be tested, rigorously. Experimental or quasi-experimental designs are not always appropriate for dynamic or complex program models.

The most effective way to measure the degree to which amplification occurs - if at all - is through a full factorial experimental design. Depending on the priorities of the study, however, utilizing a fractional factorial design or testing an integrated program against a counterfactual can be enough to demonstrate success for some aims.

Challenges for impact evaluation include the larger numbers of data points necessary for the counterfactual and the compounding problem of uneven exposure when multiple interventions are involved.

Case studies are appropriate to: identify the how and why of the added value of integration, document nonlinear pathways, illustrate context, and describe the process behind observed changes. Systems approaches can also be used to augment the evaluation of integrated programs, and better understand documented amplified and synergistic effects.

Combining qualitative and quantitative methods is important particularly for integrated programs, as combining methods has the potential to reveal not only what occurred, but why.

# 7 | EVALUATING SCALE-UP

**Scale-up** is the process of expanding the reach of a successfully tested practice in order to benefit more people and to develop sustainable and institutionalized programs and policies long-term.



- $\rightarrow$  Does the pooling of resources by multiple sectors lead to more robust development solutions that can be sustained over time when implemented on a broad scale, under real-world conditions?
- $\rightarrow$  Do integrated models that show promise as proof-of-concept

pilots retain their added value and feasibility when scaled up?

 $\rightarrow$  Are there priorities or perspectives that supported integration at the pilot level that may not be present everywhere (and what can be done to mitigate these differences)?

**KEY CONSIDERATIONS** 

Evaluators must take into account that the coordination of different actors may differ from pilot to scale-up, as the feasibility of decision makers from multiple sectors conducting joint budget exercises or sharing supervisory responsibility will be influenced by different factors in different areas. Examining the success or failure of scale-up should include examination of these integrated management and coordination activities.

Organizations or partners supporting an integrated pilot intervention may not represent the motivations of their broader communities; preparing for scale-

up requires examination of support for cross-sector collaboration beyond the pilot area, to ensure that it is not merely an idiosyncrasy of a small group of actors.

Expectations should be managed for any project. Ideally, for the scale-up of an integrated project, managers should be able to assign interventions to sites that promise the most success, and should communicate that similar comparison sites may be difficult to find and that variation in implementation is expected. This variation is not necessarily a burden, but is rather an opportunity to better understand the factors favoring and impeding effective implementation of an integrated intervention.

#### **Purpose of This Guidance Document**

16

The purpose of this document is to provide evaluators, funders, and development practitioners with guidance on evaluating integrated development programs. Although general "good evaluation practices" are woven throughout its contents, the document focuses on the unique characteristics of evaluating the complexities associated with integrated, multi-sector program implementation. These unique aspects — centering on the concept of interaction across sectors and activities — provide the common thread that ties this guidance document together.

Importantly, the guidance provided here builds upon — rather than duplicates — existing models, frameworks, and recommendations for the evaluation of complex, integrated approaches. For select examples of these, see Appendix A.

#### Structure of This Guidance Document

The modules in this guidance document reflect various types of evaluative activities. For the most part, they reflect a logical progression of evaluative activities and processes within a program's lifecycle using a before, during, and after implementation structure. Nevertheless, many of the activities can happen simultaneously and be interdependent. This document is also predicated on the assumption that not every program will require all of the evaluative activities presented. For example, an evaluator may not have enough time or funds to do an impact evaluation; conversely, an impact evaluation may be conducted without doing formative research first, if much is already known about the program context and target populations. For this reason, each section of the document is written in modular form and is intended to stand on its own. Readers can choose which module(s) is/are relevant for a particular evaluation context.

Readers can refer to Appendix B for a real-world example of the design and implementation of a multisector evaluation. The journal article "Planning an integrated agriculture and health program and designing its evaluation: Experience from Western Kenya" provides extensive detail on components of the evaluation process used to assess an integrated approach, including formative research, theory of change, program design, indicator development, mixed-method quasiexperimental impact evaluation, and costing.<sup>1</sup> As such, the journal article provides a concrete example that illustrates the respective modules in this guidance document.

Key terms appear throughout the text in **bold**, and a glossary is provided at the end of the document.

# What Is Integrated Development?

Rapid changes in the global economy, migration, climate, demography, and technology are forcing the international community to reconsider the multifaceted challenges of global development and the growing need for solutions that recognize the complex relationships among them. For example, the Sustainable Development Goals (SDGs) — an ambitious framework of 17 goals to end extreme poverty, fight inequality and injustice, and reverse climate change over the next 15 years — emphasize the integration of previously distinct development aims. The SDG agenda states that the "goals and targets we have decided on are integrated and indivisible and balance the three crucial dimensions of sustainable development: the economic, social

and environmental." Indeed, a network analysis of the 169 SDG targets reveals a web of closely related objectives and intertwined relationships among the 17 goals.<sup>2</sup>

Acknowledging the interconnected nature of our world, however, is merely a first step toward realizing the "winwin cooperation" among the social, economic, and environmental sectors. A full realization of the integrated SDG agenda requires critical changes in the way we think, make decisions, and act, based on a fundamental understanding of the linkages among these sectors. FHI 360 believes that understanding the deep connections among addressing global challenges, customizing our responses in collaboration with communities, and simultaneously addressing multiple aspects of people's lives are powerful ways to operationalize the new crosssector global agenda. We refer to this as integrated development.

Emphasis is placed on integration across the core or primary sectors of global development: agriculture and food security; economic development and livelihoods; education; environment; governance; health; nutrition; and water, sanitation, and hygiene (WASH). Examples include the Millennium Villages Project<sup>3</sup> and the holistic model used by Nuru

2 Le Blanc D. Towards integration at last? The sustainable development goals as a network of targets. NY: United Nations Department of Economic & Social Affairs; 2015. DESA Working Paper No. 141 ST/ESA/2015/DWP/141.
3 Millenniumvillages.org [Internet]. [cited 2016 August 1].

Millenniumvillages.org [Internet]. [cited 2016 August 1]. Available from: http://millenniumvillages.org/



#### INTEGRATED DEVELOPMENT

An intentional approach that links the design, delivery, and evaluation of programs across sectors to produce an amplified, lasting impact on people's lives. 17

<sup>1</sup> Cole DC, Levin C, Loechl, C, Thiele G, Grant F, Girard AW, et al. Planning an integrated agriculture and health program and designing its evaluation: experience from Western Kenya. Evaluation and Program Planning. 2016 Jun;56:11-22 [cited 2016 Aug 1]. Available from: http://www.sciencedirect.com/ science/article/pii/S0149718916300581

International.<sup>4</sup> Multi-sector integration can be valuably informed by lessons from experiences in intra-sector integration (for example, linking two health-related fields such as family planning and HIV). Contemporary integration can also be informed by lessons from past efforts, such as the integrated rural development programs undertaken in the 1970s and 1980s. Though many assessments of integrated rural development programs suggested failures based on top-down strategies and poor implementation, most agree that the original rationale for taking multi-sector approaches was and remains valid.<sup>5</sup>

Generally speaking, this document contrasts "integrated approaches" with "vertical approaches" rather than using the term "single-sector approaches." This is so as not to exclude relevant approaches that integrate programs within the same sector. The guidance in this document can be applied to both multi-sector and intra-sector integration.

Importantly, we consider integrated development a possible means to an end, and therefore neither a goal in itself nor necessarily the most appropriate approach in all cases. In other words, it will not always be the best path, but at a minimum it should be on the table as an option for consideration among all possible approaches. We want decision makers to routinely consider integrated solutions, rigorously explore them for synergies, and systematically support them when they add value and produce a greater impact.

The rationale for advancing integrated development will vary based on the perspectives, priorities, and ultimate aims of the decision makers.

Funders may emphasize amplified outcomes, cost efficiencies, or enhanced sustainability, whereas implementers may prioritize saving time, improving user satisfaction, or achieving greater equality. In practice, vertical models and integrated approaches will each have particular advantages and disadvantages in a given setting. For example, a vertical approach may be more affordable, whereas an integrated effort may reach more people or save time. Depending on their priorities, decision makers can explore whether an integrated approach may offer sufficient advantages over a vertical model with respect to one or more of the anticipated, desired, or potential benefits listed in Figure 1. Figure 2 shows an example of the potential benefits of a project integrating activities in WASH and education.

FIGURE 1

Conceptual framework of anticipated, desired, or potential benefits of integrated development programs (as compared to vertical programs)

# SECTOR A PROGRAM ACTIVITIES

# Operational Benefits

TIME SAVINGS Prevention or reduction of duplicative activities or services VALUE FOR MONEY Benefits outweigh the costs or economies of scale

# Improved Outcomes

SATISFACTION Providers or users prefer a more integrated model REACHEEnhancedFaccess tomservices ormhigher overallpnumber of(ii)people servedo

EQUITY Reached poorer, more vulnerable, or more underserved people/populations (including women or youth)

The activities illustrated in Figure 2 are drawn from those planned for the Improve the Education of Girls in Niger (IMAGINE) program; however, the operational benefits and improved outcomes are purely illustrative examples from a relatively straightforward two-sector project.<sup>6</sup>

# SECTOR B PROGRAM ACTIVITIES

ENHANCED HUMAN RESOURCES Improved motivation, skills, or retention of staff STRENGTHENED CAPACITY OF LOCAL COMMUNITY OR GOVERNMENT Better management of crosssector services

**SUSTAINABILITY** Longer lasting effects or changes AMPLIFIED IMPACT Synergistic effects on one or more dimensions of the program SYSTEMS CHANGE Broader-scale impact or more institutionalization of approaches 19

<sup>4</sup> Nuru International [Internet]. Irvine (CA); c2008-2014 [cited 2016 August 1]. What we do. Available from: http://www. nuruinternational.org/what-we-do/nuru-model/

<sup>5</sup> Brinkerhoff DW (Agency for International Development Office of Rural Development and Development Administration). The effectiveness of integrated rural development: a synthesis of research and experience. 1981 Aug [cited 2016 Aug 1]. Available from: http://pdf.usaid.gov/pdf\_docs/PNAAS642.pdf

<sup>6</sup> Bagby E, Dumitrescu A, Orfield C, Sloan M. Long-term evaluation of the IMAGINE project in Niger. Washington, DC: Mathematica Policy Research. 2016 Jun [cited 2016 Aug 1]. Available from: https://mathematica-mpr.com/ourpublications-and-findings/publications/imagine-report-englishjune-2016

### FIGURE 2

Example of conceptual framework<sup>7</sup>



# **Improved Outcomes**

SATISFACTION Children and parents more likely to report positive feelings about local schools

REACH EQUITY More students More girls enrolled in school enrolled in and More students attending have access to a school school that is not prohibitively far away from their home

SUSTAINABILITY Improved longterm WASH outcomes in communities with WASH integrated into schooling

AMPLIFIED IMPACT More students enrolled in school Decreased absenteeism

of students

sending children to school

# More parents

# CHANGE reporting interest in

SYSTEMS

empowered to provide input on decision making

educational system and

# **Types of Integration**

Integration can occur at various levels, including funding, policy, and programming levels. The guidance in this document is specifically aimed at linking sectors in the field through program integration (such as linking agricultural extension workers with health educators to deliver a bi-sector project). Different types of program-level integration are appropriate depending on the context for which they are considered. The most appropriate types strongly depend on the goals of a program and the capacity for collaboration among relevant stakeholders. As will be demonstrated in the subsequent modules, the degree of integration will in turn influence what is monitored and evaluated and how. For reference, a general integration typology is presented in Figure 3.8

Note that different types of integration are not mutually exclusive; they can be deployed alone or simultaneously in combination with others. For example, a single national program aimed at linking WASH with environmental education may both cross-train midlevel managers and planners from each sector and ensure that the key services are co-located.

8 Adapted from Agrilinks [Internet], USAID [cited 2016 August 1]. Nutrition-sensitive agricultural programming online training course. Module 4, part 4. Available from: https://agrilinks.org/ training/nutrition-sensitive-agriculture

7 Ibid.

#### The Call for Increased and **Improved Evaluation of Integrated Programs**

As the development community implements an SDG agenda with increased cross-sector collaborations, more evidence will be needed to understand the optimal circumstances, at all levels of the system, for integrated solutions. In this regard, integrated approaches should be considered for potential effectiveness alongside other options to address development challenges. Yet, generalized statements about the effectiveness of integration may be impossible because integrated development is an umbrella term that describes many combinations of interventions (from health and microfinance to nutrition and education to conservation and livelihoods). Given this extreme heterogeneity, the interpretation of evidence for integrated approaches will need to be tailored to the sectors being combined and to their specific contexts. It is encouraging that to date, a majority of randomized evaluations have found that in many diverse contexts and via various models, integrated interventions have produced positive outcomes.<sup>9</sup> Where there is good evidence on the effectiveness of specific approaches (such as offering conditional cash transfers that boost short-term and

9 Guest G, Petruney T, Ahner-McHaffie T, Dooley B. Evaluating integrated development: are we asking the right questions? A systematic review. Forthcoming 2016.

21

#### FIGURE 3 Common types of integration

# COMPLETE INTEGRATION

A program's staff members receive substantial high-quality training and skills-building in topics outside their sector of expertise. People are then served with multiple services by the same provider.

• Example: A program committed to advancing youth-friendly development builds the capacity of its staff to offer integrated, holistic support for young people by covering their needs in health, life skills, education, and workforce development.

# CROSS TRAINING

A program's staff receive basic orientation to and training in the additional sector/s so they can reinforce complementary messages and offer multi-sector information when they conduct their regular sectoral work in a community.

• Example: A marine-conservation program recognizes that high rates of unintended pregnancies and large families can increase the overfishing of vulnerable stocks. They train their current environmental workers to provide information, referrals, and some basic family planning methods to their remote communities who are underserved by health services.

# **CO-LOCATION**

This is the simplest way to bring sectors together. By intentionally targeting the same community with programming from multiple sectors, we are more likely to see benefits from the complementarities between them. The two programs, however, are not necessarily coordinated and there may not be overlap between all of the program participants. • **Example:** A nutrition program learns of the importance of sanitation for the absorption of micronutrients by children, so it uses spatial mapping to deliver its nutrition interventions to the same districts that benefit from a new WASH campaign being delivered by the government.

# COORDINATION

This approach involves the joint planning of different programs to harmonize interventions across sectors, but keeps the implementation separate. This strategy is often used by an organization that plans its interventions centrally, but employs experts from separate sectors to implement the interventions independently.

• Example: Local health and education specialists recognize the link between good nutrition and cognitive performance. They discuss together what each can do to improve the situation. The education program adds nutritional status to the list of possible reasons for a student's poor performance and identifies undernourished students. The students' families are then referred to local health units for support and services.

# COLLABORATION

In addition to joint planning, this approach includes instances when the implementation of activities is carried out together (but the services are still delivered by sectorspecific staff).

÷¥

• **Example:** A large agricultural company has committed to improving the health (and subsequent productivity) of its workers. The company already arranges for monthly on-site visits to farmers' homes by agricultural trainers, so it collaborates with the district health office to establish a program in which local health care workers also join the monthly home visits, to offer health screenings, basic services, and referrals. long-term economic, health, nutrition, and education outcomes), decision makers should actively use the data to inform the design of funding, policies, and programs. For integrated models with less evidence available, funders, researchers, and policymakers should collectively prioritize the most pressing needs and allocate resources to fill the knowledge gaps through evaluation and research.

# What is Unique about Evaluating Integrated Development?

Relative to vertical development programs, integrated approaches to development are generally:

- $\rightarrow$  More complex in design
- More complicated in implementation (with a greater number and more diverse range of actors involved)
- → More diverse in terms of inputs, outputs, and outcomes
- → More innovative, and therefore possibly requiring more adaptive or emergent thinking
- → Confronting different cultures of research and evaluation between sectors
- → Based on the premise that the interaction between interventions from two or more sectors will generate benefits beyond a vertical intervention, such as improved outcomes or operational benefits (Figure 1)

These characteristics have implications for how integrated models are assessed and evaluated with respect to the questions being asked and the methods and designs used to answer those questions. For example, the desired goal of integration varies greatly depending on the perspective, priorities, and ultimate aim of a given decision maker. Funders may emphasize producing cost efficiencies or enhanced sustainability, whereas program implementers may prioritize saving time, improving user satisfaction, or reducing inequality. Depending on the specific aim of a given effort, integration can be considered effective if it offers advantages over a vertical programming model with respect to one or more of the operational benefits or improved outcomes listed in the conceptual framework (Figure 1).

The evaluation of these effects should be designed to detect not only the positive effects that integration may have in that area, but also whether integration creates unanticipated negative consequences (for example, if adding a new service degrades the overall quality of program delivery). This dimension has proven particularly important to stakeholders in the "base" program upon which additional activities are being integrated. The issue of what type and level of evidence are deemed "acceptable" by different groups of stakeholders and decision makers is also important, and they vary with regard to what is

 $\checkmark$ 

#### VERTICAL PROGRAM/ ACTIVITY

An activity that is designed, supervised, and executed, either wholly or to a great extent, by a specialized, often sectorspecific group with a focused purpose.

considered sufficient when deciding to replicate or scaleup. This is a general issue for development research, but the issue becomes amplified when decision makers come from multiple sectors, each with its own culture of evidence. See Appendix B to learn more about how this issue was handled in a real-world setting.

Greater complexity and diversity, coupled with a focus on interaction, has implications for how integrated programs are monitored and evaluated. Evaluating these programs goes beyond following good evaluation practices (which are, of course, essential). The unique nature of integrated development affects all components of evaluation – from developing logic models and costing approaches to choosing indicators and an evaluation design. Additional research and thoughtful evaluation objectives and design that are specific to integration also need to be considered.

Viewing integrated models through a **systems** lens can help stakeholders to better conceptualize the complexities involved in integrated programs. **Systems thinking** is a perspective in which a range of approaches and methods that aim to describe and develop an understanding of the underlying structure of a system are used to make inferences about the system, to develop programs that work best in the system, and to most effectively affect change. It can help:

- → Define what the program is and is not
- → Identify the components of the program and their interdependent relationships
- Understand multiple perspectives on what is happening within the program and outside the program over time

The next seven modules in this guidance document discuss the particular evaluative approaches that will lead to the strongest evaluation of integrated programming. Certain potential advantages (and potential positive or negative cost implications of integrated programming) can be determined better by some of the evaluative approaches than by others. Some approaches, such as formative research, could benefit any program. It should be noted than any evaluative approach could be designed to be useful for assessing many of the advantages or costs. As mentioned, these can be used independently or in combination depending on the needs of the evaluation.

#### SYSTEMS THINKING

 $(\mathbf{V})$ 

A perspective in which a range of approaches and methods that aim to describe and develop an understanding of the underlying structure of a system are used to make inferences about the system, to develop programs that work best in the system, and to most effectively affect change.

# V

#### SYSTEM

A set of connected components or elements that form a complex whole.

#### COMPONENTS

The pieces that make up a system, including stakeholders, activities, and settings.







# **1 | FORMATIVE RESEARCH**

**Formative research** informs program content, design, and operation. The primary task is to determine the key problem or nexus of problems to be addressed, and opportunities and ideas for addressing them prior to implementation.



# $\rightarrow$ How can the interaction

- among sectors be used to enhance efficiencies, outcomes, or sustainability?
- → How do stakeholders from different sectors perceive the goals of the integrated intervention and the theory of change that underlies it? What are their expectations and how do they differ?
- → What challenges might program staff and beneficiaries face with integration?
- → How can logistical efficiencies be enhanced? Conversely, what are potential challenges and negative effects that might arise due to the interactive nature of the design?
- $\rightarrow$  What are potential effects of the integrated program on the community and the larger socioeconomic system?



# **KEY CONSIDERATIONS**

Mapping the stakeholder environment and the broader system can help develop specific evaluation questions and measures for a complex project, and allows focus on areas that may have the greatest impact.

Early development of an integrated theory of change and/or a logic model, or alignment of different models, can help stakeholders and evaluators identify emergent outcomes, relationships between activities, and best practices.

Inclusive participatory techniques are effective ways to build or revise a theory of change/logic model; enhance cross-sector cooperation and communication; and identify potential areas of convergence, incompatibility, or unintended consequences associated with integration.

Through various methods, stakeholders can provide insights into how integration may affect: services within their sector of operation, the social structure of the larger community, local and regional governance, access to and cost of services, and dynamics within the household (such as gender or age).

**Formative research** is often conducted before a program begins to inform its content, targeting strategies, and program processes and procedures. In general, the purpose of formative research is twofold: 1) to maximize the impact and efficiency of an intervention and 2) to identify and subsequently minimize potential logistical, ethical, and methodological issues that might arise. For integrated programs, there is an additional focus on the interactive effects of delivering interventions from two or more sectors.

More specifically, program development data from formative research is used to:

- → Clarify the combination of problems to be addressed
- $\rightarrow$  Better understand the local community and stakeholder environment
- → Identify improvements to program content and design
- → Improve recruiting and other program processes
- $\rightarrow$  Inform selection and refinement of indicators
- $\rightarrow$  Identify potential problems with, and unintended consequences of, a program

Doing formative research for integrated programs involves the same basic methods and principles as for any program — primarily focus groups, in-depth interviews, and participant observation. However, the

### **Mapping the Stakeholder** Environment

One of the first steps in the formative process is to identify and characterize key stakeholders. For integrated programs, the stakeholder groups are often larger in number and more diverse. Cross-sector interactions both positive and negative — are also likely. The matrix below provides a structure to help identify stakeholder groups and relationships or conflicting goals among groups (Table 1). The example in the matrix demonstrates the challenges of and some potential opportunities in integrating nutrition activities with agriculture and food security activities. Either focus groups or qualitative individual interviews can be used to complete this matrix.

complexity of integrated programs imposes additional focuses related to interaction between sectors.

# $(\mathbf{V})$

#### FORMATIVE RESEARCH

Research that informs program content, design, and operations prior to implementation The primary task is to determine the key problem or nexus of problems to be addressed, and opportunities and ideas for addressing them.

27

# TABLE 1

Cross-Sector Stakeholder Matrix (with Illustrative Example of a Project Integrating Activities in the Nutrition and Agriculture Sectors).

STAKEHOLDER	SECTOR A:	SECTOR B:	RELATIONSHIP	
DUP	NUTRITION	AGRICULTURE & FOOD SECURITY	BETWEEN SECTORS	
	[Specific group(s) and their goals and expectations for the program]	[Specific group(s) and their goals and expectations for the program]	[Note degree of communication and potential differences in goals and expectations]	
OGRAM ₹TICIPANTS	Families (especially women) with young children They want to provide for their families and keep them healthy.	Farmers (often men) who work small plots of land in target communities Farmers want to provide for their families and maximize profit.	Farmers (often men) who work small plots of land in arget communities Farmers want o provide for heir families and maximize profit. The successful implementation of this project would mean that the groups targeted would not be separate but would be integrated, so that the activities that touch different parts of the household would reach the family in a united fashion. That being said, these groups communicate closely (as they are often in the same household), but there are different power dynamics. Men in these communities usually control the income from farming, the products being produced (food versus cash crops), how much land is used for the production of household items to use versus to sell, etc. Women typically prepare the food products that are consumed by the family and household, and they tend to know which ones would be most nutritious and beneficial. However, the crops available for consumption are largely dependent on what the men have already decided to grow. Therefore, discussions about crop selection and use need to be carried out with both men and women together.	
JGRAM STAFF	Community nutrition workers They expect to perform their duties according to how they have been trained.	Agricultural extension workers They expect to perform their duties according to how they have been trained.	The agricultural extension workers have been offered cursory training on working with nutrition workers in the past. However, not all of them have undergone the training, and it was conducted more than two years ago. The community nutrition workers have not been trained to work with agricultural extension workers. There has been limited communication between the groups in the past. This project may include interacting not only with new staff members, but also with new community groups for each of these staff groups. The different groups of staff also have different organizational cultures and expectations.	
L LEADERS DING POLITICAL, MIC, RELIGIOUS)	Local health providers, elected officials, tribal/ village leaders, priests, women's organization leaders	Established farming community leaders, local leaders of commodity organizations	In some districts, communication between these parties is already well established (for the local leaders as compared to the policymakers). Community leaders have their own priorities but are also familiar with the communities' broader problems. Health officials are resistant to integration	
	They want to improve the health and nutrition of the local community.	They want to increase frey want to mprove the health and maximize profit for local producers. bcal community.		

**COMPLEX ADAPTIVE** SYSTEM A set of interrelated, interacting, and interdependent components and subsystems that make up a whole and can adapt, learn, and change over time.



#### SUBSYSTEM

A self-contained group of interrelated, interacting, and interdependent components within a larger system (such as the primary health care system or tertiary education system).

# OUTCOME

The measurable short-and medium-term result(s) or effect(s) of program activities.



#### THEORY OF CHANGE

The description of a sequence of events that is expected to lead to a particular desired outcome. It typically includes the context for the initiative, the longterm change the initiative supports and from which it will ultimately benefit, the process/sequence of change anticipated to lead to the desired long-term outcome, assumptions about how these changes might happen, and a diagram and narrative summary that captures the outcomes

#### **Mapping the System**

On a broader level, designers of integrated programs may want to understand and describe the whole system during formative research, including the larger socio-economic and environmental context. A complex adaptive system (CAS) is an unpredictable and dynamic system whose components are constantly interacting and adapting to changes.<sup>10</sup> For instance, an integrated project with activities in agriculture, health, and nutrition may involve stakeholders, institutions, and participants across the health **subsystem**, agriculture and economic subsystem, and community subsystem, all of which are interrelated and connected. A change in one component of a subsystem, such as better trained health workers in nutrition and counseling on vitamin A, may have direct and indirect effects (positive or negative) on all of the other subsystems (see Figure 4 for a causal loop diagram illustrating how systems tools can be used for this type of project). In order to know how and why desired **outcomes** may or may not be produced, it is important to understand the context of the systems, as well as their interactions and how they are connected and influence each other. There are several helpful tools for observing systems changes that can be used to capture and describe these dynamic elements.

Systems mapping is a participatory and iterative process conducted with stakeholders. It helps to first understand the system in order to identify systems changes, such as policies, power structures, relationships, and values, in addition to identifying program targets. The process involves developing a graphical representation of the system that includes its components and connections. Building the systems map can be done through interviews and workshops with stakeholders and should align with other tools, such as the program's theory of change (TOC). Meetings undertaken to develop or align other planning tools can also be used to develop systems maps. Approaches to mapping the system include mapping the actors within a system — the key individuals and groups — and how they are connected (social network analysis); mapping the political, social, or economic issues that affect the program area; or mapping the actors with a focus on feedback loops leading to changes in behaviors and functions (causal loop diagram, see Figure 4 for an illustrative example).<sup>11</sup> Systems mapping captures the complexity, dynamism, and nonlinearity of integrated programs within the environmental context in which they are implemented. Systems maps can also serve as a guide for developing

#### FIGURE 4

#### Causal loop diagram of a health, agriculture, and economic subsystem<sup>12</sup>



evaluation guestions and measures that focus on areas that would have the greatest impact, while keeping in mind the possible negative direct or indirect effects on other subsystems.

#### Informing Program Design and Implementation

Good practice for evaluation includes a well-informed TOC and/or a **logic model**. This is especially true for integrated interventions because they are usually more complex in design than vertical interventions. More

important, however, is ensuring that the multiple and diverse stakeholders involved with an integrated intervention are informed of, and agree with, the theory. Formative assessments are relevant to TOCs, logic models, and program design, as they provide a mechanism to use stakeholder input and guidance to inform the development process. In ideal scenarios, this stakeholder engagement with regard to the integrated aspects of a TOC or logic model (or other similar models such as impact pathways, Figure 5) takes place very early in the process. In many settings, however, a program will

and care of neonate



#### LOGIC MODEL

A planning tool for evaluating the effectiveness of a program that lays out what the program hopes to accomplish and its intended impact. It includes the following basic components: goal, resources/inputs activities, outputs, outcomes, and impact.

31

<sup>10</sup> Swanson RC, Cattaneo A, Bradley E, Chunharas S, Atun R, Abbas KM, et al. Rethinking health systems strengthening: key systems thinking tools and strategies for transformational change, Health Policy Plan, 2012 Oct:27(Suppl 4):54-61. doi:10.1093/heapol/czs090. Available from: http://www.ncbi nlm.nih.gov/pmc/articles/PMC3529625/

<sup>11</sup> de Pinho H. Participant guidelines. Systems tools for complex health systems: a guide to creating causal loop diagrams. NY: Columbia University. 2015 Feb [cited 2016 Aug 1]. Available from: http://www.who.int/alliance-hpsr/resources/ publications/CLD\_Course\_Participant\_Manual.pdf

#### **IMPACT PATHWAY**

A planning tool for evaluating a program that includes many of the same components of a logic model, but also shows the progression or chain of what the project will do, through outputs that achieve outcomes and the desired impact.

#### APPRECIATIVE INQUIRY

"[A] group process that inquires into, identifies and further develops the best of 'what is' in organizations in order to create a better future. Often used in the organization development field as an approach to large-scale change, it is a means for addressing issues, challenges, changes and concerns of an organization in ways that build on the successful. effective and energizing experiences of its members."

already have one or more established TOCs or logic models. In these cases, formative approaches can be used to solicit feedback, answer questions from stakeholders, and revise the model, as necessary. The complexity of integrated programs may be difficult to fully capture in a linear or predictive logic model, which can be limited in presenting potential systems level changes. Aligning TOCs and logic models with other systems-related tools, such as **appreciative inquiry** and ripple-effect mapping exercises, can help stakeholders and evaluators identify emergent outcomes, relationships between activities, and best practices.

One particularly useful technique at this stage of integrated program design is a whole-system-in-theroom (WSR) workshop. Rather than convening experts in one thematic area, this method convenes a diverse set of stakeholders from a variety of sectors who work on or are affected by the development problems that the program plans to address. A key result of a WSR workshop should be a clear statement that articulates a balance between the possible advantages and potential constraints of integrated programming (see Practical Example 1). The workshop

conclusions should also recommend how to proceed with integration, including a recommendation on which type of integrated programming should be implemented (see "Types of Integration" in the Introduction). If the recommendation is for less rather than full integration, individual logic models might be elaborated on by respective design teams and subsequently reviewed together. These logic models should include associated performance indicators. When the two design teams come back together, they would discuss the logic and the choices of indicators. World Café is another inclusive participatory approach that uses small group work to achieve the same ends.<sup>13</sup>

In the case of formative research for the development of integrated programs, inclusive participatory group techniques are effective ways to:

- $\rightarrow$  *Build* or revise a TOC/logic model, including revealing possible non-linear and multi-directional program effects
- → Enhance cross-sector cooperation and communication
- $\rightarrow$  *Identify* potential areas of incompatibility and unintended consequences associated with the interaction between intervention activities within two or more sectors. For example, any development program can

#### **PRACTICAL EXAMPLE 1** A Whole-System-in-the-Room Approach

The System-Wide Collaborative countries, SCALE+ uses a Action for Livelihoods and Environment (SCALE+) methodology utilizes WSR as an essential component in facilitating cross-sector planning for integrated initiatives.<sup>14</sup> SCALE+ is a systems methodology used to accelerate broad stakeholder engagement in sustained collaborative action to address a complex development issue. Designed by USAID and Academy for Educational Development (now FHI 360), SCALE+ originated in 2004 as a tool for integrating work on livelihoods and the environment, but it has since been effective in other cross-sector initiatives for health, education, nutrition, and governance. Successfully applied in more than 15

have unintended consequences. Because integrated programs are generally more complex and diverse than vertical programs, the potential for unanticipated consequences, whether positive or negative, is greater.

locally driven approach that ensures the meaningful participation of groups that are often excluded, such as youth, women, and the poor. SCALE+ can identify policy actions that complement rather than duplicate existing efforts; boost cross-sector buy-in and accountability measures; ensure that integrated approaches resonate with and correspond to local realities; and increase the likelihood that support for cross-sector programs will be institutionalized and sustained.

14 Scaleplus.fhi360.org [Internet]. FHI 360 [cited 2016 Aug 1]. Available from: http://scaleplus. fhi360.org/Futuresearch.net [Internet]. Future Search Network; c2001-03 [cited 2016 Aug 1]. Future Search Methodology. Available from: http://www.futuresearch.net/method/ methodology/index.cfm

Although formative research cannot identify all of the unforeseen corollary effects, it can certainly identify and mitigate some. Participant observation in areas where the intervention might take place or within the target population can provide insights in this regard. Another approach to



#### **RIPPLE-EFFECT** MAPPING

Uses elements of appreciative inquiry mind mapping, and qualitative data analysis to engage individuals to map the intended and unintended changes a program targets.



#### WHOLE-SYSTEM-**IN-THE-ROOM** WORKSHOP

A method that convenes a diverse set of stakeholders from a variety of sectors that work on or are affected by the development problems that the program plans to address.

<sup>13</sup> W.K. Kellogg Foundation. Using logic models to bring together planning, evaluation, and action: logic model developme guide. Battle Creek (MI): W.K. Kellogg Foundation; 2004 [cited 2016 Aug 1]. Available from: http://www.smartgivers. org/uploads/logicmodelguidepdf.pdf



# Impact pathway for Mama SASHA program<sup>15</sup>

15 Cole DC, Levin C, Loechl, C, Thiele G, Grant F, Girard AW, et al. Planning an integrated agriculture and health program and designing its evaluation: experience from Western Kenya. Evaluation and Program Planning. 2016 Jun;56:11-22 [cited 2016 Aug 1]. Available from: http://www. sciencedirect.com/science/article/pii/

uncovering these corollary effects could be to work with stakeholders who represent different sectors and who operate at various levels to create a collective map of the social structure and larger systems (such as the systems mapping tool described in the section "Mapping Systems" earlier in this module). Using the map (or other illustrative methods including process diagrams or fictional vignettes) as a reference, questions such as 'What would happen in your sector if we integrated services from another sector?' can be posed. Stakeholders can be asked how the integrated intervention might affect:

- Services within their sector or level of operation
- → The social structure of the larger community
- → Local and regional governance

- $\rightarrow$  Access to and cost of services
- Dynamics within the household (particularly with respect to gender and age)

In particular, if the program is integrated such that staff deliver at least some activities from more than one sector, designers should consider beforehand whether these staff have the skills and the time to do both. Process mapping could be extremely useful here. Program designers could diagram the anticipated work flow of program activities. While creating the diagram, they might ask questions about efficiencies and potential conflicts and problems. It will be important to explore the possible effects of integrated services on staff efficiency, quality of service, and morale.

# Developing Performance Indicators

# BEFORE IMPLEMENTATION

\_\_\_\_\_



# **2 | DEVELOPING PERFORMANCE INDICATORS**

**Performance indicators** are the precise measures used to assess success by a program or activity. Indicators for integrated programs can include sector-specific indicators and value-added indicators. Collection of sector-specific indicators is either required or recommended for distinct programs, and can sometimes be standardized by sector or funder. Value-added indicators measure amplified effects or synergy beyond what would have occurred in a vertical program.



- $\rightarrow$  What indicators can be used to measure the expected integrated outcomes (amplified effects or synergies) of the program?
- → Do program designers or evaluators need to harmonize indicators across sectors?
- $\rightarrow$  What indicators can be used to measure activities in two or more sectors?

# **KEY CONSIDERATIONS**

Careful planning is needed when developing indicators for integrated programs, as the potential for greater number and complexity of indicators is high. Harmonizing indicators between activities and/or stakeholders is necessary and can be achieved by using proxy indicators, adapting data collection to streamline as much as possible, and convening stakeholders to prioritize indicators.

Meetings between experts in particular sectors, or meetings with stakeholders on specific projects, can help identify and harmonize key indicators, and can reinforce the common understanding of the intervention's outcomes and impact and how they will be achieved.

In order to choose or develop value-added indicators, program designers need to identify what "integration" means within the context of the program, and the pathways through which the program is intending to have an effect on its goals. An integrated program can use traditionally sector-specific indicators to track integration if the program anticipates that value-added will be measured through those indicators.

The most thoughtfully chosen indicators need to be paired with a monitoring and evaluation system that is designed to show relationships between the outcomes produced (the instances in which a sector's outcome indicator is the result of multiple aspects of integrated programming).

**Performance indicators** are part of the program monitoring system that will be fully addressed in Module 3. Performance indicators are explored separately here because they play such an essential role in making programs measurable and achievable. **Output** and outcome statements take on greater realism, and in general greater specificity, when the discussion turns to the indicators used to measure them. Because of the inherent complexity of integrated programs, in-depth discussion of measurability and achievability are essential steps to designing a manageable integrated program. Developing performance indicators and their **targets** early in the program design process strengthens the program design, and serves as the basis for quality monitoring and evaluation (M&E).

In an integrated development program, performance indicators must capture both the planned changes from the program being implemented and the additional value derived from the choice to integrate. The indicators should be valid, reliable, precise, measurable, timely, and programmatically relevant. This will enable evaluators to tell whether any progress is being made toward achieving program objectives, and where this progress is occurring. The indicators will also enable evaluation of whether the integration of development sectors has produced outputs and outcomes that are more effective, more efficient, or of a wider variety than those measured in vertical programs.

Indicators for integrated programs will include sector-specific indicators and

value-added indicators. In addition, the M&E system must be designed to be able to show the relationships between indicators (for example, the instances in which a sector-specific outcome indicator is the result of multiple aspects of integrated programming). Integrated programs must be particularly careful when selecting and prioritizing indicators or else they may succumb to the pitfall of **indicator creep** — that is, including so many indicators that the M&E staff, the frontline staff collecting the data, and even the beneficiaries become overwhelmed by data collection needs. The potential for a greater number and complexity of indicators, as well as increased complexity in tracking and analysis, should be considered when planning for and resourcing M&E for integrated programs.

# $\checkmark$

#### PERFORMANCE INDICATORS

The precise measures used to assess success by a program or activity.



#### OUTPUT

The products, goods, and services that results from an intervention.

# 

#### TARGETS

The amounts of expected change within a given timeframe.



#### **INDICATOR CREEP**

The inflation of obligations as new measurements are required to account for expansion of activities.

39

# Developing Performance Indicators 2

#### **Sector-Specific Indicators**

Generally, development sectors have their own standardized indicators that are either required or recommended for programs to collect and that may vary by funder. These are the **sector**specific indicators. Between sectors, some indicators may be shared (for example, if a nutrition project and a health project are tracking breastfeeding rates). However, even if the objective is the same, the indicators often have distinct nuances that can hinder harmonized data analysis. A majority of the indicators in the performance monitoring plan (PMP) of an integrated program will be sector-specific.

#### Harmonizing Indicators

Integrated programming compounds the usual risks of unmanageably long lists of indicators and overly complex analysis needs. Careful planning can mitigate this. First, program designers and M&E specialists identify those indicators that will measure the outputs and outcomes in early draft design. They should then review the definitions of any existing standard indicators linked to international, national, or funder norms (for example, the standard Foreign Assistance indicators, Demographic and Health Surveys, indicators from private funders, or USAID-required definitions) to identify any differences that may pose problems later during data

collection or analysis. If the detailed indicator definitions already being used by separate sectors are identical, the indicator can simply be selected for the integrated program without issue. If the detailed indicator definitions are completely or even slightly different (such as if unmet need for family planning is tracked by both sectors but the programs define the indicator as applying to different age ranges), integrated program staff will need to decide how to move forward.

Some nuanced differences in indicator definitions have workable solutions. For example, one sector may decide that it could use another sector's indicator as a **proxy indicator** for what it would usually measure in a stand-alone program. Or, in the case of different disaggregation, it might be possible to adapt data collection forms to enable collection and reporting of individual data elements, rather than to aggregate data (expanded in Module 3). Where there are no workable solutions, it may be advisable to convene stakeholders and funders to agree on whether both indicator variations are necessary, or to compromise. It is always good practice to condense the list by requesting that stakeholders from each sector designate their respective indicators as either required or recommended, and as either primary or secondary data. The M&E staff can then carefully delineate responsibilities for data collection for specific indicators that will be collected by a specific team,

# PRACTICAL EXAMPLE 2

### Harmonizing Indicators through Stakeholder Meetings

The Clean, Fed & Nurtured community of practice works to bring together experts in WASH, nutrition, and early childhood development. In 2015, this community convened a meeting that provided an overview of research in the three sectors, looked at field examples, and discussed impact and outcome indicators for programs integrating the three sectors.<sup>16</sup> The meeting identified existing frameworks and sets of indicators that could be used, a list of both sector-specific and integrated indicators, and strategies for projects in different contexts to choose indicators based on their needs.

particularly at the output level. This is most workable when the program is not completely integrated or when there are specific teams for specific activities. These stakeholder conversations can be folded into conversations to align program content and integration opportunities detailed under the section "Informing Program Design and Implementation" in Module 1 (see Practical Example 2).

#### **Value-added Indicators**

Value-added indicators measure effects beyond what would have occurred in a vertical program. These indicators can be quantified both in terms of **amplified effects** and

IDANCE FOR EVALUATING INTEGRATED GLOBAL DEVELOPMENT PROGRAN

#### SECTOR-SPECIFIC INDICATORS

 $|\psi|$ 

Indicators, sometimes standardized, that are either required or recommended for programs to collect and that may vary by funder.

# PROXY INDICATOR

A stand-in or approximate indicator for an indicator

of interest. It may use an alternative definition and/or be derived from a different data source.

# V

**PRIMARY DATA** Information collected directly by a program.



#### SECONDARY DATA

Information derived from other sources rather than collected directly by a program. These kinds of meetings for technical groups and for specific projects can help identify key indicators and look for places where indicators can be unified. They can also reinforce a common understanding of the integrated intervention's outcomes and impacts and how they will be achieved.

16 Indicators for Clean, Fed, and Nurtured: Impact, Outcome, and Output Indicators. [2015 Oct 5; cited 2016 Aug 1]. Available from: http://www.coregroup.org/storage/ documents/Indicators\_CFN\_05Oct2015v2.pdf

Clean, fed & nurtured. Joining forces to promote child growth and development. A report on a consultative meeting. Alive & Thrive, The Global Public-Private Partnership for Handwashing with Soap (PPPHW), Save the Children, The WASHplus Project; 2013 May [cited 2016 Aug 1]. Sponsored by the Bill & Melinda Gates Foundation, FHI 360, and USAID. Available from: http://aliveandthrive. org/wpdev/wp-content/uploads/2014/11/2013-Clean-Fed-Nurtured-Meeting-Report.pdf

in terms of **synergy**. For example, amplified effects could be reaching more people, while synergy could be reaching new population groups (see Figure 6). In an integrated program, it is important to identify what "integration" means within the context of a thoughtful logic model and an understanding of the pathways through which the program is addressing its goals in a TOC prior to developing appropriate and relevant indicators (see "Types of Integration" in the Introduction). Guided by stakeholders (possibly through a WSR workshop detailed in Module 1), individuals developing indicators should particularly keep in mind the potential advantages of integration as described in the Introduction. as well

41

V

#### VALUE-ADDED INDICATORS

Indicators that measure a synergy or amplified effects above and beyond what would have occurred in a vertical program.



#### AMPLIFIED EFFECTS

A potential benefit of integration where the act of integrating produces a result that is larger than what would have occurred in vertical programs.



#### SYNERGY

A potential benefit of integration where the act of integrating produces a result that would not have been seen in vertical programs.



as the potential costs and unintended consequences. Reviewing these considerations in the specific context of the program can help capture the synergies and amplified effects of integrated programming.

Which indicators are considered valueadded indicators is interwoven with the program's conceptual framework and logic model (see Module 1, "Formative Research") and is the responsibility of the designers of the integrated program. These will change from program to program, even if they cover the same domains or sectors, based on the program's context and set of activities. An integrated program can use traditionally sectorspecific indicators to track integration if the logic model or impact pathway anticipates that value-added will be measured through those indicators. Indicators for the amplified effects aspect of an integrated program may

include measurement to document groups of people who are reached by the intervention beyond the original mandate (improved access to target populations) or greater ease or use of the program activities (improved beneficiary satisfaction). For example, more women gaining access to agricultural extension services or livelihood activities when a project integrates these services in health can add value through integration (see Practical Example 3). Indicators for the synergistic aspects of an integrated program may include measurement to document social norms change. In the same example, different measures of women's empowerment (agency, assets ownership, decision making on health care choices, etc.) can demonstrate change that may not be seen in a vertical program. Listing all imaginable value-added indicators is not advisable; the indicators should be considered carefully in the context of

#### **PRACTICAL EXAMPLE 3 Amplified Effects**

A quasi-experimental study conducted in the Philippines examined a project integrating coastal resources management the value-added indicator "use with reproductive health services to address overfishing and high population growth through "food security from the to the single-intervention sea."<sup>17</sup> The study included three arms: one arm integrating resource management with reproductive health, one arm providing resource management activities alone, and one arm providing reproductive health activities only. The study demonstrated that outcomes in resource management, health, and food security were amplified when integrated, as compared to the single-intervention arms. In the integrated area, all nine food security and health indicators reached the desired level; five of the nine indicators performed better than the

### Systems-Level Measures

the individual program by a group of people familiar with the possible and likely interactions and synergies. Not all integrated development programs will have value-added indicators. Depending on what benefits the integrated program is targeting (see Figure 1, "Conceptual Framework"), value-added indicators may not be necessary to document some intended outcomes of integration.

single-intervention arms and the other four performed at the same level. For example, of contraceptives at first sexual encounter" was higher in the integrated arm as compared arms. Environmental conditions, measured with 18 indicators, also improved the most in the integrated area as compared to singleintervention arm areas. The thoughtfully chosen impact indicators, when compared through the study design, were able to demonstrate the amplified effect of integrating for health and coastal resource management in these communities.

17 D'Agnes L, D'Agnes H, Schwartz JB, Amarillo ML, Castro J. Integrated management of coastal resources and human health yields added value: a comparative study in Palawan (Philippines). Environmental Conservation. 2010 Dec;37(4):398-409.

Social network analysis can be used to generate quantitative systems-level measures. The method mathematically examines systems structures and the interaction of multiple components within a system through the use of network and graph theories. Social network analysis measures the strength and/or nature of relationships and interactions between individuals. institutions, or other social entities. Common examples include friends, colleagues, and economic, kinship, and disease relationships. Network

43

# Developing Performance Indicators 2

#### **PRACTICAL EXAMPLE 4 Network Analysis**

Organizational network analysis was conducted in Malawi for an integrated nutrition project to assess the links between clinical sites and community services for economic strengthening, livelihoods, and food security. Identification of key organizational actors and their linkages to one another helped to better identify synergies for improved health status outcomes and to sustain those outcomes due to the centrality and focus on connecting local organizations. In Malawi, a

baseline network analysis was conducted to understand the accessibility of providers and their referral systems. The analysis found that there was a disconnected network, which the project aimed to address.<sup>18</sup> A follow-up network analysis can show to what extent the relationships and referral system have been established and are functional.

data can provide various systems-level measures for integrated programs such as network density or the centrality of key actors/points across multiple sectors. Network analysis tools can measure a wide range of relationship characteristics, such as quantity of interactions, quality of relationships (such as trust), as well as the geographic boundaries of the networks. Network data can also be visualized through sociograms, which depict items in the network (with dots as nodes) and the nature of and strength between each node (with lines as ties), which can also be a powerful advocacy tool (see Practical Example 4).

To use systems measures for monitoring systems changes brought about via integrated approaches, evaluators can collect social network measures pre-intervention, at intervals during the intervention, and at intervention completion. For

evaluation, trends in results would be analyzed and triangulated with other data on systems change.

Measuring changes to policy or the influence of policy change, which do not lend themselves to simple performance indicators, can also be measured through simulation models of the system. This is done by quantifying the causal loop diagram developed with stakeholders during the formative stage (see resource C in Appendix A for a guide to developing causal loop diagrams and see Figure 4 for an illustrative example). Similarly, feedback loops, dynamic relationships, and delays among variables and determinants of the target outcome can also be developed as performance measures to assess, and monitor over time, which interventions and combinations of interventions would most efficiently produce the desired outcomes.



# DURING IMPLEMENTATION

# Program Monitoring



<sup>18</sup> Reynolds HW, Curran J, Thomas JC. Organizational network analysis: MEASURE Evaluation's experience. 2010-2014. Chapel Hill (NC); 2014 Aug [cited 2016 Aug 1]. Available from: http://www.cpc.unc.edu/ measure/resources/publications/sr-14-103

# **3 | PROGRAM MONITORING**

**Program monitoring** is the routine, systematic observation and recording of program implementation and problems using the performance indicators developed and other monitoring processes, including analysis and feedback about the progress of the program to the donors, implementers, and beneficiaries (for example, through site visits and periodic stakeholder meetings).



- $\rightarrow$  What data collection processes or forms can be adapted to serve more than one sector's purpose?
- $\rightarrow$  Who should have access to the data to best facilitate integrated monitoring, and who should meet to discuss the data?
- $\rightarrow$  Can progress, or lack thereof, in one sector inadvertently have consequences in another sector (and how can that be monitored)?
- $\rightarrow$  How will the monitoring system track beneficiary or household access to or usage of multiple services within different sectors?



Responsibility for the collection, housing, analysis, and reporting of data needs to be clearly documented, and may cross sectors or organizations depending on how the program is designed. Different levels of integration will require different solutions to the problems of integrating monitoring systems.

Integrated programs require integrated program monitoring processes and teams. These teams would ideally comprise both M&E staff with specific expertise in individual sectors and M&E staff experienced in program integration.

It is important to cross-reference changes and identify synergistic interrelationships in order to capture cross-sector changes.

Reporting on an integrated program may be challenging because of a larger number of, and greater variation in, activities being implemented and a wider variety of stakeholders.

**Program monitoring** achieves several purposes. One purpose is to help a program continuously assess whether it is on track to achieve the stated desired outcomes. This information can then be reported to key stakeholders, such as funders and host governments. Monitoring data is also the fundamental basis for evaluating a program in terms of outputs, efficiencies, and outcomes over the course of the program. Here we focus on developing a performance monitoring system specific to an integrated development program and how that system can be used to continually assess program achievements.

#### **Methodology**

Discussions about monitoring methodology begin when a program is being conceived. Designers decide which indicators should be collected (discussed in Module 2), how they will be collected, who will collect them, and with what periodicity they will be collected. In addition, program designers have to discuss with whom the indicators will be shared, and when, where, and by whom they will be analyzed. The designers must anticipate constraints to an ideal system and make appropriate modifications. The performance monitoring plan (PMP) lays out the framework but does not generally specify where or how the data are archived, how they are analyzed, or who has access to the data. This is generally the point at which the

discussion of targets (the amount of expected change) is begun. Project monitoring discussions clarify these and other issues.

An integrated program offers more options to all of these decisions on collection, analysis, and access, and complicates the "chain of custody" of the data. In some integrated programs, one sector or program will have a monitoring system established that can be built out to include the extra indicators and data sources that integration requires. In other cases, new systems will need to be developed or links will need to be developed between two relatively equal systems. Different levels of integration (see Figure 3) will require different solutions to the problems of integrating monitoring systems. Key monitoring stakeholders can

#### PROGRAM MONITORING

The routine, systematic observation and recording of program implementation and problems using the performance indicators developed and other monitoring processes, including analysis and feedback about the progress of the program to the stakeholders.



#### PERFORMANCE **MONITORING PLAN**

A document designed to support the measurement of progress in achieving the results specified in the program's logic model and specifying performance indicators. definitions, data sources, reporting frequencies and responsibilities, and targets.

47

develop a clear picture of what already exists in the sectors and areas being integrated, and then advise on the best way to move forward. Established integrated logic models or impact pathways can serve as a start to monitoring conversations. Programs should attempt to document a clear understanding of the responsibilities around the collection, housing, analysis, and reporting of the data, while including open access to data for different groups within the program.

Ideally, data are routinely analyzed to monitor progress toward achieving outcomes, outputs, and other benchmarks (perhaps mitigations of negative side effects), and to highlight domains or other areas of success or challenges. How often this analysis is performed varies with the need for updated data. The decision on periodicity rests mainly on what questions the data can answer, how quickly a negative situation can be rectified or a positive situation replicated, and when programmatic decisions are scheduled to take place. Program monitoring data should be readily available to managerial and technical stakeholders and program personnel from all program/ sector components so they can easily track progress and do not always have to rely on M&E teams to provide answers to their questions. Making all programming data (not just sector-specific) available to technical and programmatic staff and organizing reporting of program-wide achievements can tease out some of

the added value of integration. Wide availability of data from all areas of the program can help technically specialized staff members working on integrated activities see where programs produce cross-sector or cross-activity benefits, and can reveal any unintended results. On a set schedule, program monitoring data should be presented and discussed among program staff and stakeholders. Discussions of the data with monitoring and program staff across integrated sectors or activities can further illuminate what has been found and what needs to be further measured. From these discussions and presentations, decisions can be made about the need for more qualitative lines of inquiry to discern sources of the problems or where best practices are found.

#### **Monitoring Processes** and Teams

Because integrated programs will inevitably have sector-specific and value-added indicators, program monitoring will likely be conducted within and across sectors. The exact nature of this division ("within" versus "across" monitoring) will depend on how services are integrated and what efficiencies can be gained. However, regardless of how the performance indicators are distributed, integrated programs require integrated program monitoring processes and teams. A team would ideally comprise M&E

staff with specific expertise in the individual development sectors, as well as M&E staff experienced in program integration.

The M&E team may decide to conduct ongoing routine monitoring reviews for each development sector to assess achievement within the sectors. In addition, the M&E team will need to 1) assess progress in integration and value-added indicators and 2) assess where progress, or lack thereof, in one sector may have inadvertently had consequences in another sector. Until there is greater experience with integrated programs, good practice would suggest that the M&E team, and possibly the program team, conduct frequent site visits and stakeholder consultations to identify unintended consequences and whether the identified value-added indicators are indeed capturing the potential advantages and costs of integrated programming.

#### **Performance Monitoring Plans** and Data Collection Tools

In PMP development, it is necessary not only to develop detailed indicator definitions, but also to define the data sources, the timing and process by which data will be collected and reported up through the program, and the targets of expected change. PMPs should also have detailed analysis plans that capture the cross-sector contributions and synergies of the integrated programming. Depending

→ If the *same implementer* is responsible for providing services across development sectors, having a single streamlined data collection form with all of the necessary data elements may be easier on the provider than having him/her complete multiple and separate forms, which may contain repeated data elements (for example, name and age).

on how integration is being defined and implemented within an integrated program (see Figure 3), data collection tools may be affected. In some cases, it may make sense to combine primary data collection tools, and in other cases it may be more effective to keep the data collection tools distinct. This decision will largely depend on the indicators selected in the PMP, how services are being integrated, and the phase of scale-up. At the service delivery level, options include the following:

 $\rightarrow$  If, however, services are being integrated through the introduction of services from one development sector into a site currently offering services from another sector, and those services will be provided by *different providers*, then retaining the distinct data collection forms may be more practical. In this case, integration is occurring more at a management level than at a service provider level. This will require a tool/process for integrating the data, but one that the frontline

# Program Monitoring **3**

data collectors will not be responsible for creating.

Wherever possible in an integrated development program, the aim is to reduce redundancies in data collection forms regardless of the number of sectors involved.

#### **Analysis and Reporting**

Because of the need to capture cross-sector changes, the analysis plans will have distinct requirements above and beyond those of a program implemented vertically. Aggregation and making connections between diverse aspects of a program are always necessary, but in integrated programs, the need to cross-reference where the changes occur and identify synergistic interrelationships requires greater attention to this part of the program monitoring process.

By working out an analysis tool or format early in the process of developing a program monitoring system, the need to modify some indicators, definitions, or frequencies is likely to become apparent. It is highly recommended that, if feasible, a unique identifier system (may or may not be linked with names) be developed, so that each service or activity can be linked to the individual who receives it. Being able to identify beneficiary data at the individual level provides greater analytic clarity if the integrated program is targeting higher-level social structures such as households, organizations, or communities.

Programs are often expected to report on the total number of beneficiaries served. This may prove challenging to an integrated program, particularly when a funder requires sectorspecific totals, which is often the case when that funder is accountable to multiple funding streams. Ideally, the program would count the total number of beneficiaries served with any programming, and disaggregate by the type of services received (for example, for sector A, sector B, both sector A and B). This can be used to capture both direct and indirect beneficiaries.

Reporting on an integrated program may be challenging because of a larger number of, and greater variation in, activities being implemented. There may well be more funders and other stakeholders to report to, each with their own reporting format and requirements.

# **DURING** IMPLEMENTATION

\_\_\_\_\_

# Process Evaluation



## **4 | PROCESS EVALUATION**

Process evaluation is a method of assessing and understanding how a program is being implemented, focusing on the program's operations and service delivery.



→ What level of integration

is occurring across sectors?

- $\rightarrow$  What is the quality of the program components?
- $\rightarrow$  Can data on the implementation experience explain how any observed amplified or synergistic effects were achieved? If none were achieved, can the data explain why they were not?
- $\rightarrow$ What strategies are working for or inhibiting the cross-sector coordination or collaboration required by the program?
- What implementation experiences may be unique to crosssector programs?
- $\rightarrow$  Are the target beneficiary population(s) being reached, and with which activities?
- $\rightarrow$  Are households or individuals accessing more than one part of the intervention (and if so, how many)?



Evaluators should maintain a bird's-eye view when seeking and generating evidence, and be aware of what has been achieved by monitoring and what evidence still needs to be explored. Having a good understanding of how integration was supposed to be achieved, and the potential constraints and value added of integration is key for the process evaluator.

Feedback for the process evaluation should be sought from a multi-sectoral group of stakeholders; pains should be taken to make sure no sector dominates the evaluation. Familiarity with systems methods can be an additional asset here to identify parts of the system that need to be evaluated and explored.

While monitoring is usually thought of as a continuous process, **process evaluation** is considered a more discrete activity. As a method of assessing and understanding how a program is being implemented, process evaluation focuses on the program's operations and service delivery and how they affect the likelihood of achieving outcomes and impacts. It is important to include the experiences of beneficiaries, non-beneficiaries, staff, and other stakeholders to assess how the program is perceived and to identify possible barriers to success. The evaluation looks at not only what the program has completed or delivered to date, but also why certain elements are working well and others not. There may be some overlap with monitoring in some programs, but that overlap can be utilized by evaluators interested in process. Planning for or conducting a process evaluation for an integrated program is similar to doing so for any program, but with some unique aspects (see Practical Example 5).

#### Methodology

As for most programs, the methodology of an integrated program's process evaluation should be mixed. This may include document review, secondary data analysis, quantitative surveys, and qualitative inquiry. Systems approaches to evaluation can be particularly useful in process evaluations to understand interrelationships, perspectives, and

and P).

boundaries that help and hinder change. Integration may be used in resiliency programs and complex or conflict-affected areas where there are many interconnected needs and the situations are particularly fluid. These issues should be considered when choosing methods. Further resources on qualitative and systems approaches can be found in Appendix A (particularly resources B, C, H, M, N, 53

#### **PROCESS EVALUATION**

A method of assessing and understanding how a program is being implemented, focusing on the program's operations and service delivery.

Process Evaluation 4

#### **PRACTICAL EXAMPLE 5**

#### **Process Evaluation of an Integrated Family Planning** and Microfinance Program

A 2013 study tested the feasibility of training members of a microfinance self-help group, called village health guides, to deliver family planning information and provide linkages to family planning services.<sup>19</sup> Although the primary outcomes measured were related to family planning use, a secondary objective assessed the feasibility of training the village health guides to deliver family planning information and provide linkages to family planning services as a regular part of their interaction with clients. Results showed that adding family planning information to the existing health information activities of a microfinance organization increased uptake of modern family planning methods and dramatically reduced unmet need among the cohort observed. With regard to implementation, findings showed that semi-literate *health workers successfully* 

delivered the family planning information after one week of training and with ongoing support. Building on an existing cadre of health workers within the organization who were already seen as trusted sources of information was a key part of the intervention. The semiliterate village health guides were able to deliver simple family planning messages and make referrals to existing service providers. Through the referrals, this intervention complemented investments made in service delivery projects. The process-related results from this study have shown that adding a family planning information campaign to a microfinance project is feasible and can be successful.

19 Delivering family planning information and services through a microfinance program: lessons from Uttar Pradesh, India. FHI 360/PROGRESS, Network of Entrepreneurship and Economic Development (NEED), Institute for Reproductive Health (IRH): 2013 Apr [cited 2016 Aug 1]. Available from: https://www.fhi360.org/ sites/default/files/media/documents/india-needmicrofinance-family-planning.pdf

#### **Stakeholders**

The stakeholders targeted in an integrated program's process evaluation should be multi-sectoral, with input and feedback sought from the various sectors addressed in the program. The process evaluation should include stakeholders from the various levels of program

implementation — beneficiaries, frontline data collectors (such as service providers), site management staff, systems-level representatives, and program staff — as well as local leaders and other key actors in the program's context. This group of stakeholders can be informed by work done in the formative research phase (see Module 1, "Mapping the

Stakeholder Environment"). This will ensure a broad perspective on the effects of multi-sector integration. Stakeholder consultation would likely be more in-depth in a process evaluation than in the performance monitoring processes discussed in Module 3.

#### **Team Composition**

Ideally, the process evaluation team would be multi-sectoral and would include an evaluator experienced in integration. Familiarity with systems change evaluation is an additional asset. The process evaluation team should be formulated particularly for the evaluation and, besides the lead evaluator, be composed of some of the larger M&E team, sector experts, sector-specific M&E team members, and possibly an individual from outside the program. Because integration is a key component of the program, the evaluators must maintain a bird's-eye view when seeking and generating evidence. This is particularly necessary when addressing the evaluation objectives, asking questions about how the program has achieved integration between sectors, and understanding the perceived constraints or value added to the integration.

An integrated program offers more options to all of these decisions on collection, analysis, and access, and complicates the "chain of custody" of the data. In some integrated programs, one sector or program will have a monitoring system established that can be built out to include the extra indicators and data sources that integration requires. In other cases, new systems will need to be developed or links will need to be developed between two relatively equal systems. Different levels of integration (see Figure 3) will require different solutions to the problems of integrating monitoring systems.

55

# Process Evaluation 4

# DURING

# Cost Analyses



# **5 | COST ANALYSES**

**Cost analysis** is a technique for documenting the extent to which any operational benefits occur in integrated programs and the size of these gains.



- $\rightarrow$  What are the costs of integrated versus vertical programming (including short- and long-term costs, cost efficiencies, financial costs or savings associated with negative outcomes or missed opportunities, etc.)?
- $\rightarrow$  What are the cost implications of additional inputs and processes necessary for the successful management, coordination, and delivery of an integrated program that are unique in comparison to the implementation of vertical programs?



Costs of an integrated program will often be frontloaded as compared to standard vertical programs. The requisite expertise and time needed increases as the complexity of program design increases, although this will vary based on the type and extent of integration.

When thinking about costs, it is often useful to distinguish between different phases of a program, including design, preparation, and operational phases. Any operational benefits may not be realized until the operational phase, suggesting higher start-up costs for integrated programs.

Many cost analysis methods require allocating and separating all costs and measures of an outcome to particular sectors or activities. As the degree of integration increases, it is increasingly difficult and subjective to allocate funds to distinct sectors, especially for resources like labor. For this reason, cost-effectiveness analysis of integrated programs is particularly challenging unless a single effectiveness metric can be used.

In most cases, a comparator will be required so that the costs of an integrated program can be assessed with respect to a non-integrated approach.

Cost savings are a key potential benefit of integration, and **cost analysis** can be used to determine if these kinds of cost efficiencies are achieved. Cost analysis techniques are ideally suited for documenting the extent to which any operational benefits occur in integrated programs and the size of these gains. This is important given that the added value of integration can manifest through operational benefits, as noted in Figure 1. We further expect the degree of integration (Figure 2) to influence the likelihood of these gains.

In general, regardless of whether or not one is talking about an integrated program, costs reflect the value of the resources (for example, labor, supplies, equipment, and infrastructure) used to design, prepare for, and operate a program. The costs are influenced by the following three factors:

- $\rightarrow$  The type of resources used
- → The amount of resources used
- $\rightarrow$  The value assigned to those resources

To do a comprehensive cost analysis, one must consider the phases of program implementation (design, preparation, and operation including M&E), as well as the three factors influencing cost. The key issue when selecting an appropriate unit value for a resource is perspective. In a full economic evaluation (a cost-effectiveness analysis or cost-benefit analysis), an economic

perspective is used and every resource is assigned a value representing its opportunity cost. Capital equipment and infrastructure are valued based on the current cost to replace them and the appropriate **annualization** factor. The full economic evaluation will also consider costs to the clients (such as fees, transport, and time), as well as societal costs associated with lost productivity because of illness or other reasons. For an analysis from the payer's perspective or financial cost analysis, a resource is valued according to the costs that were actually incurred to secure it. When considering multiple sectors, the perspectives needed to assign value can be even more complex than they are in vertical programs.

This module focuses on how integration influences an ingredientsbased approach to activity-based costing. It also discusses some of the unique challenges to overcome

# $\checkmark$

#### **COST ANALYSIS**

Technique for documenting the extent to which any operational benefits occur in integrated programs and the size of these gains.

#### **COST-EFFECTIVENESS** ANALYSIS

A form of economic analysis that compares the costs and outcomes of two or more courses of action.

# $(\downarrow$

#### **OPPORTUNITY COST**

A value assigned to a resource that reflects the value of the resource in its best alternative use. For example, a person's time spent waiting is valued according to what that person could earn through paid employment but cannot because he or she is waiting.

#### ANNUALIZATION FACTOR

Similar to depreciation, the annualization factor is used to spread the acquisition cost of a resource, which lasts multiple years across its expected useful life. Unlike depreciation, it also considers the value of the money tied up in the investment that is no longer available for other uses (opportunity cost of the investment).

60

when trying to evaluate an integrated intervention.

#### **Design Phase**

Depending on the type of integration envisioned (see Figure 3), designing an integrated program could take longer than designing a vertical program. It could also involve more groups or organizations. As a result, the list of resources used is likely to be longer (reflecting inputs from all parties involved) and may require assigning different values to a resource depending on where the resource originates.

For example, a person with a particular job in an international nongovernmental organization may have her time valued differently than a person with the same title in a local government entity, reflecting different compensation packages (for instance, salaries plus benefits plus allowances) across the two organizations. Moreover, integrated programs are often supported by an array of funders across multiple sectors. This can potentially have implications for how resources are valued or assigned a monetary value in the absence of expenditure information.

If the design phase takes longer for an integrated program than for a vertical program, we would expect the integrated program to use more resources (more person hours, more meetings held, and more travel reimbursements). A final challenge unique to integrated programs is that separate estimates may be needed for the different sectors involved. This may be necessary if funding will come from multiple sectors and each sector needs a separate estimate of resource requirements. In this case, one would have to determine which resources in the resource list are relevant to each individual sector and how the remaining resources should be allocated across the different sectors. A naïve approach would be to divide shared resources equally across sectors, but that should be an explicit decision rather than the default.

The central challenges during this phase will come from ensuring that the necessary information from all parties involved is captured and that the data are kept separate by source. This is so that if the data will be used in secondary analyses, such as estimating the cost of scale-up, adjustments can be made to both the quantities of resources required and the valuations used to reflect changes in the context. A final issue to consider is whether the experiences gained during implementation of the integrated program suggest modification to the approach used to implement and operate the program.

#### **Preparation Phase**

Once decisions have been made about how an integrated program will be structured and the roles of the various organizations involved, organizations typically prepare for the introduction of the integrated program. The duration of the preparation phase may be longer than that of a vertical program, but by varying degrees depending on the type and extent of integration. The tasks to be completed in this phase can include:

- $\rightarrow$  Preparation of training materials
- $\rightarrow$  Development of guidelines and informational materials
- $\rightarrow$  Development of branding material for the new or newly integrated program
- $\rightarrow$  Procurement of new space or modification/renovation of existing space to be used in delivery of the program
- $\rightarrow$  Training of service providers in the provision of the integrated program (including modifications to or introduction of new reporting formats to be used to document activities)
- $\rightarrow$  Planning for integrated sectoral M&E and additional resources for integration, analysis, and potentially multiple reporting
- $\rightarrow$  Orienting the community/ communities to be served by the

As with the design phase, the central challenge will be to document all the types and quantities of resources (including labor, supplies, equipment, and infrastructure) by the source of the resource (which organization or organizations are providing the resource) and the assignment of a unit value to each resource. Allocation of resources across sectors will also be necessary if sector-specific cost estimates are needed.

# **Operational Phase**

In the operational phase, the integrated program is up and running and the beneficiaries are being served. As the program begins operation, the need to document the resources used and to attribute the quantity of resources used by each piece of the program continues. As described above, a unit value needs to be assigned to each resource. If there are distinct services being provided as part of the program (for example, food supplements and antiretroviral therapy), which service(s) are being supported by each resource needs to be documented on the resource list. Costs of the operational phase are typically estimated on an annual basis. Thus, even if data are collected

integrated program about what is changing and why

 $\rightarrow$  Promotion of the program to be launched

#### **FINANCIAL COSTS**

 $(\mathbf{V})$ 

COUNTERFACTUAL

A control or comparison group that matches the

treatment group on as

many dimensions and

variables as possible.

Costs that consider the transfer of funds required to secure the use of a resource. Resources for which no transaction occurs (donations) would be assigned a unit cost equal to zero.

#### ECONOMIC COSTS

Costs that consider the opportunity cost of securing a resource. A resource is assigned a unit value equal to what it would cost to obtain the resource in the local market. A donated input would be valued at what the user would need to pay to obtain that resource if it was not being provided to the program at no cost.

for only a few months, this should be extrapolated to an annual amount.

As the degree of integration increases, the challenge of allocating to different sectors or services is likely to become more subjective, especially for labor and other shared resources (like equipment or infrastructure). The alternative is to conduct a rigorous, direct observation of service provision including provider time (such as in a time-motion study).

#### **Using Cost Data for Evaluation**

How cost data are used to evaluate an integrated development program will depend on the purpose of the evaluation.

#### IF THE OBJECTIVE OF THE **EVALUATION IS TO MEASURE THE** SIZE OF OPERATIONAL EFFICIENCIES FROM INTEGRATED PROGRAMMING. then one will need a counterfactual

for comparison (presumably the provision of services within separate sectors). One may begin by comparing the cost per service in the operational phase between the integrated program and the counterfactual. However, at some point any potential savings in the operational phase should be compared to the costs required in the design and preparation phases, with an eye toward how long it would take to offset these "investment" costs through the savings obtained during

the operational phase. In this type of analysis, the valuation of resources should be based on financial costs rather than economic costs, and the implementer perspective is usually used (see Practical Example 6).

For example, a project might wish to compare the costs of providing sexual health counseling through a peer educator versus through a clinician. The peer educator's time is likely to be assigned a lower value than the clinician's time, but the counseling session may take longer with the peer educator, especially if the client is more willing to discuss his or her issue in depth with a peer. In this case, the operational efficiency would be influenced by the product of time spent with the counselor and the cost per minute of the counselor's time.

#### IF THE GOAL OF THE EVALUATION IS TO ASSESS THE COST-EFFECTIVENESS OF THE INTEGRATED PROGRAM.

a counterfactual is still required, as the term "cost-effectiveness" is a relative term. In this case, we would want to compare the total cost of activities in the integrated program with the total cost of these same activities as provided separately in the different sectors. It will be important to control for potential scale effects across the alternatives being compared; this is often done by using a standard population size (such as per 1,000 clients).

If a common measure of effectiveness is not being used across the sectors being compared, then the costs and the outcomes or impacts of the integrated program will need to be reported separately for the separate sectors, and the same indicator of outcome or impact will be needed for each alternative being evaluated (see Module 2). Again, it will be important to control for potential scale effects and to consider expected outcomes for a standard population size.

The cost-effectiveness analysis is summarized by an incremental costeffectiveness ratio. This is when the change in total cost moving from the less expensive to the more expensive alternative (controlling for scale or volume of output) is divided by the change in the outcome or impact indicator. The result is a dollar amount per unit change in the indicator (for example, the cost per client successfully treated for a sexually transmitted infection). If the change in impact is negative, this indicates that the more expensive alternative is less cost-effective and should be rejected. If positive, it is left to the decision maker to decide if the rate at which additional resources are translated into additional results is acceptable.

Of note, concluding that a program is cost-effective does not guarantee that it is affordable. For example, gold is an excellent conductor of electricity, but we do not wire our households with

**EVALUATION IS TO ASSESS THE** AFFORDABILITY OR COST OF SCALE-UP, then an explicit counterfactual is not required. The valuation of resources would be on a financial cost basis, and the value used should reflect the cost to the most likely source of the resource. The magnitude of resources required can be adjusted to reflect several changes:

gold wires. It is simply unaffordable.

# IF THE GOAL OF THE

a. Differences in the intensity of the design phase. For example, for scale-up, partner buy-in or the development of training materials and other support tools may no longer be needed. The design costs may even be omitted in the case of replication to a new site with the current partners.

**b.** Changes in the preparation phase reflecting any changes in the scale of the program.

c. Changes in the operational phase corresponding to changes in the volume of services provided. If the length of trainings will change, or if parts of the preparation phase are no longer necessary, the resources required for those activities may be omitted.

#### **INCREMENTAL COST-EFFECTIVENESS RATIO**

A ratio that summarizes cost-effectiveness analysis by calculating the change in total cost divided by the change in the outcome or impact indicator, resulting in a dollar amount per unit change in the indicator.

Cost Analyses 5

The cost and effects of improving WASH in schools



An important part of any evaluation of an integrated program may be separate cost estimates by sector so that, for example, the cost to the education sector may be reported separately from the cost to the health sector. If these costs per output, outcome, or impact are lower under the integrated program, this suggests that the integrated program may have more efficient operations and in turn the potential for greater sustainability over time.

For example, an integrated project that improves water and sanitation in schools may have fewer students missing school because of diarrheal diseases. The costs in the health component are probably higher to

cover the health intervention, but the cost per student completing a grade level may actually fall, as the number of students successfully completing the grade level may increase. Therefore, the education sector will become more effective/efficient because of the investments in the health sector (see Figure 7).

#### **Cost-Effectiveness versus Cost-Benefit Analysis**

In a full economic evaluation of a program, the costs, outcomes, and impacts of the program and a comparator need to be considered. A challenge when evaluating an integrated program is that there are often multiple measures of outcomes and impacts using different metrics. For example, a health outcome might be measured by **quality-adjusted life year** and an economic strengthening outcome might be measured by income relative to poverty. One of the rationales for integration is that we expect a positive feedback loop between the two measures. One strategy to assess this would be to try to disentangle the costs related to the health component from the costs related to the economic strengthening component and perform two separate cost-effectiveness evaluations with sector-specific comparators. This allows the evaluator to use the two metrics in their original form.

Alternatively, the evaluator could decide on a cost-benefit analysis, in which case a monetary value would need to be assigned to each unit of outcome or impact and then these monetary values can be aggregated across the two measures. This could be highly subjective but does allow for a single "net cost" assessment of the integrated program versus its comparator. Currently there is no ideal solution; the solution one chooses would depend on the purpose of the evaluation and potentially on the specific questions being asked.

#### QUALITY-ADJUSTED LIFE YEAR

A common measure of effectiveness used in costeffectiveness analyses of health interventions that considers not only life expectancy but also the 'quality" of the remaining life years based on the presence of diseases or other conditions.

#### PRACTICAL EXAMPLE 6 HIV and Sexual and Reproductive Health Integration Cost Analysis

The Integra Initiative is a research project about the benefits and costs of a range delivering integrated HIV and of models for delivering integrated HIV and sexual and reproductive health (SRH) services in settings with medium and high HIV prevalence, to reduce HIV infection (and associated stigma) and unintended pregnancies. The initiative performed a retrospective facility-based costing study in 40 health facilities in Kenya and Swaziland to estimate the unit costs of six integrated HIV and SRH services. The services were family planning, postnatal care, cervical cancer screening, HIV counseling and testing, sexually transmitted infection treatment, and HIV treatment and care. Determinants of costs of integrated HIV and SRH services were evaluated. and the economies of scale and scope associated with integrated delivery were

explored.<sup>20</sup> Finally, the efficiency of health facilities SRH services were explored, taking into account structural and process measures of quality of care and the determinants of efficiency.

For all HIV and SRH services, variability in unit costs and cost components suggest the potential to reduce costs of delivery through better use of both human and capital resources. These findings reveal that contrary to expectation, efficiency gains from the integration of HIV and SRH services, if any, are likely to be modest. In addition, efficiency gains from joint production are dependent on the specific combination of resources used for different services. Efficiency gains are likely to be most achievable in settings with substantial fixed costs. The finding that there were efficiency

gains for only three service combinations highlights the need to carefully consider the setting, specific clinical practices, and the extent to which services can be combined when deciding which services to integrate. Regarding economies of scope, this study suggests that planners in all settings need to carefully consider the detailed processes and clinical practices required by each service, and which of these can be combined when services are integrated, before assuming substantial efficiency gains. The extent to which the findings of this study can be generalized will depend on how similar service delivery and clinical practices are to those observed here.

20 The Integra Initiative. Cost and technical efficiency of integrated HIV and SRH services in Kenya and Swaziland. 2014 [cited 2016 Aug 1]. Available from: http://www.integrainitiative.org/wp/ wp-content/uploads/2015/09/Integra-StepsO4-Costs.pdf





# Evaluating Impact



# 6 | EVALUATING IMPACT

An **impact evaluation** assesses the outcomes and impact that can be attributed to a particular intervention by comparing outcomes between intervention group(s) and a counterfactual (control group). However, evaluating impact for complex integrated programs is sometimes best served by employing a methodological perspective that extends beyond the traditional experimental and quasi-experimental design. This document adopts a definition of impact that includes a broader range of methods than included in traditional impact evaluations.

# **KEY QUESTIONS**

- → Were the planned amplified effects or operational benefits from integration realized?
- $\rightarrow$  To what extent were operational benefits and/or amplified or synergistic outcomes due to integration (if they were observed)? Can we demonstrate that integration led to these effects?
- $\rightarrow$  How and why did integration produce effects beyond those observed in a vertical program? If no change was measured, why did integration not produce effects?
- → Did integration result in unanticipated effects?



Impact evaluations for integrated models should be undertaken thoughtfully, and only if there is a reasonable consensus on what models need to be tested, and what can be tested, rigorously. Experimental or quasi-experimental designs are not always appropriate for dynamic or complex program models.

The most effective way to measure the degree to which amplification occurs - if at all - is through a full factorial experimental design. Depending on the priorities of the study, however, utilizing a fractional factorial design or testing an integrated program against a counterfactual can be enough to demonstrate success for some aims.

Challenges for impact evaluation include the larger numbers of data points necessary for the counterfactual and the compounding problem of uneven exposure when multiple interventions are involved.

Case studies are appropriate to: identify the how and why of the added value of integration, document non-linear pathways, illustrate context, and describe the process behind observed changes. Systems approaches can also be used to augment the evaluation of integrated programs, and better understand documented amplified and synergistic effects.

Combining qualitative and quantitative methods is important particularly for integrated programs, as combining methods has the potential to reveal not only what occurred, but why.

As part of the evaluation process, research studies are sometimes used to demonstrate the causal effects of interventions. Conventional approaches for this type of confirmatory research often favor experimental designs and statistical methods. They focus on establishing causation through the measurement of change that has occurred as a result of the program (outcome evaluation) or of longterm program goals (impact evaluation), usually for a few key variables.

**Impact evaluation** has different meanings for different parts of the evaluation community. For many, impact evaluation became synonymous with randomized controlled trials in the 1990s as the call for greater rigor in evaluating development programs started labeling experimental and guasiexperimental approaches impact evaluations. This guidance document has kept with this more traditional definition, wherein impact evaluations assess the outcomes and impact that can be attributed to a particular intervention by comparing outcomes between intervention group(s) and a counterfactual (control group). This means that for the purposes of this document, impact evaluations include only experimental and quasiexperimental designs. At the same time, we also discuss the concept of evaluating impact in a broader, more

inclusive, sense. Evaluating impact for complex integrated programs is sometimes best served by employing a methodological perspective that extends beyond the traditional experimental and quasi-experimental design.

This document adopts OECD-DAC's definition of impact as the "positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended."21 The use of the term "produced by" infers that evaluating impact attempts to document causal inferences. In the context of integrated programs, this should be achieved through a full examination not only of causation (whether the program makes a difference) but also of explanation

21 OECD-DAC. Glossary of key terms in evaluation and results based management. Paris: OECD; 2002

 $\checkmark$ 

#### IMPACT EVALUATION

An evaluation that assesses the outcomes and impact that can be attributed to a particular intervention by comparing outcomes between intervention group(s) and a counterfactual (control group). For the purposes of this document, this is limited to experimental and guasi-experimental evaluation designs, which rigorously determine causal attribution.



#### IMPACT

The "positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.'

69

**EXPERIMENTAL** DESIGN

 $(\downarrow)$ 

Randomly allocating interventions to a sample of participants either individually or as part of a larger group or cluster (such as a community).

(why and how it makes a difference). Explanation is important in providing decision makers with the information they need to change course in an existing program or to modify designs for future programs. It also enhances internal and external validity, as well as extends inquiry to permit exploration of unintended consequences. Due to their complex nature, the full effects of integrated programs may be realized through direct and indirect pathways at different levels of the system, and may require extended periods of time to be observed. Moreover, multiple interacting components may have unintended consequences. For this reason, this module includes not only experimental and quasi-experimental impact evaluations, but also case studies, systems-level approaches, and mixed-methods considerations.

This module presents a range of evaluation strategies and methods for making causal inferences while highlighting considerations specific to their use in the context of integrated programs. Evaluations of this nature should only be undertaken when there is reasonable consensus on integration models that need to be, and can be, tested rigorously. Specific questions and a well-formulated TOC should already have been articulated as part of program design, and are important to guide the choice of appropriate strategies and methods. For some

22 Consort-statement.org [Internet]. [Cited 2016 Aug 1]. Consort 2010. Available from: http://www.consortstatement.org/consort-2010

integrated programs, processes may be too dynamic, requiring constant adaptation. In such cases, experimental or quasi-experimental designs may not be appropriate.

#### Experimental

Impact evaluation studies with experimental designs involve randomly allocating interventions to individuals or to groups or clusters (such as communities), and evaluating impact by comparing the "exposed" group to the "unexposed" control group. Randomized studies have the advantage of avoiding selection biases and characteristics (such as education), as well as possible targeting biases of program implementers. Ensuring the integrity of the randomization process is essential and should be closely monitored and properly documented.<sup>22</sup>

Experimental designs typically require larger sample sizes than other designs (case studies, for example) and extensive resources (both time and money). Moreover, for practical, political, or ethical reasons, certain research contexts and hypotheses may also preclude the use of a control group or the randomization process required for experimental designs. For example, if a planned evaluation is examining a government-funded conditional cash transfer program's effect on health care decisions, it may not be possible to randomly allocate

#### **FIGURE 8**

Integrated four-arm full factorial experimental design

CONTROL HEALTH **INTERVENTION** ARM

households or individuals into groups to access that kind of social program.

#### Integrated Development Considerations with Experiments

A key question in integrated development evaluation concerns whether or not amplified or synergistic effects are observed and, if so, whether they can be attributed to the integrated design. This brings us to an essential question of integration: is 1 plus 1 greater than 2? The most effective way to measure the degree to which amplification occurs — if at all — is through a **full factorial** experimental design. In a full factorial integrated design, participants are randomized to either: 1) a control group (no intervention), 2) a single

intervention arm for each activity included in the study, or 3) a multiintervention arm for each permutation of integration. The simplest full factorial design for integrated evaluation will include four arms: one control, one for the first activity, one for the second activity, and one for the integrated activities (Figure 8). In such a design, if true amplification is achieved, the integrated arm(s) should show a degree of change that is greater than the sum of change among all of the arms that are not integrated (see Practical Examples 7 and 9). Depending on what evidence is needed



and on the hypothesis being tested, using a strategically selected subset of arms (a **fractional factorial design**) may be appropriate. For example, such a design may be used if not all combinations of interventions are



#### FULL FACTORIAL INTEGRATED **DEVELOPMENT DESIGN**

Research design in which participants are randomized to 1) a control group (no intervention), 2) a single intervention arm for each activity included in the study, or 3) a multi-intervention arm for each permutation of integration. The simplest full factorial design for integrated development evaluation will include four arms: one control, one for the first activity, one for the second activity, and one for the integrated activities. In such a design, if true amplification is achieved, the integrated arm(s) should show a degree of change that is greater than the sum of change among all of the vertical arms.

#### **ASPIRES Full Factorial Impact Evaluation Design**

The USAID-funded Accelerating Strategies for Practical Innovation and Research in Economic Strengthening (ASPIRES) project (2013-2018) supports evidencebased, gender-sensitive programming to improve both the economic security and health outcomes of families and children living with or affected by HIV/AIDS, as well as others at high risk of acquiring HIV. One effort within ASPIRES is a four-year full factorial randomized study evaluating an intervention integrating economic strengthening and HIV prevention projects for vulnerable youth in South Africa.<sup>23</sup> The research will assess whether the integrated *intervention produces* synergistic effects on economic and health outcomes. The study's primary objectives are:

 $\rightarrow$  To assess whether the integration of an economic strengthening intervention with an HIV prevention

education intervention improves economic and health outcomes beyond singular interventions

- $\rightarrow$  To estimate the resources required at the project level to support the economic strengthening and HIV-prevention education interventions. and to determine if cost efficiencies result from integration
- $\rightarrow$  To understand the nuances of cross-sector synergies through qualitative data collected from study participants, and to explain patterns in the quantitative findings

Study results are expected in mid-2018.

23 A mixed-methods study evaluating an intervention integrating economic strengthening and HIV prevention programs for vulnerable youth in South Africa. [Cited 2016 Aug 1]. Available from: https:// www.microlinks.org/sites/default/files/resource/ files/ASPIRES\_SA\_Factsheet.pdf

# $(\mathbf{V})$

FRACTIONAL **FACTORIAL DESIGN** 

An experimental design consisting of a specific subset, or fraction, of the possible arms of a full factorial design.

of interest (see Practical Example 8). One could also decide that given the gains in efficiency afforded by integration, it is sufficient to measure the effectiveness of a specific integrated program against a control group without measuring the effects of each intervention when implemented individually. Another option is undertaking a standard twoarm integration and control study, and including an examination of the

effects of an integrated approach as compared to effects documented in well-established, vertical approaches.

Because of the increased number of activities and arms associated with using an experimental design for evaluating integrated programs, certain challenges in maintaining the integrity of the study design are heightened. Preventing diffusion effects between arms and isolating

intervention effects is difficult in any field-based experimental study, but the more sectors and arms there are, the more challenging this becomes. There are accepted measures for dealing with contamination, such as documenting beneficiary exposure to activities and adjusting analysis based on that documentation. Another option is to randomize at the community level, which minimizes the likelihood of an individual's unintentional exposure to another arm. Randomizing at the community level, however, requires including sufficient numbers of treatment units (communities) to control for the differences in the variability within and between communities, which can present its own challenges.

Another challenge with a randomized design is that individuals may not accept their assignment into a specific arm and may decline to participate post-randomization. The likelihood of uneven exposure increases when multiple interventions are involved. At the community level, for example, interventions like savings groups or growth monitoring and promotion sessions involve some amount of self-selection. If both interventions are offered in a community, some households may participate in both, while others may join in only one or neither. One option is to focus the analysis on all those intended to receive the intervention, regardless of actual participation (intent-totreat analysis). Another is to analyze

be addressed.

# **Quasi-Experimental**

the impact on those effectively participating (treatment-on-treated analysis), although this approach is vulnerable to selection biases since the integrity of the random assignment is no longer preserved and the implications of such biases must

When exposure to the intervention is not determined by a random process, the study is considered quasiexperimental. Quasi-experimental designs are used in contexts where randomization of individuals or of clusters (such as households or communities) to study arms is not possible. Common approaches to quasi-experimental designs include matching, regression discontinuity, and interrupted time series design, guidance for which can be found in the general evaluation literature (some included in Appendix A). Although there are advantages and disadvantages to each of these approaches, matching and interrupted time series are likely the better fit for integrated programs (expanded below).

Rather than using random assignment, matching methods are used to construct a control group. The idea is to use data from the treatment group to create a comparable counterfactual that matches the treatment group on all relevant confounding factors. Often, 73

#### QUASI-EXPERIMENTAL DESIGN

A research design in which exposure to the intervention is not determined by a random process: instead, a counterfactual not exposed to the intervention is constructed in such a way that it matches the treatment group on as many dimensions and variables as possible.

#### **PROPENSITY SCORE** MATCHING

 $(\mathbf{V})$ 

A statistical matching technique used in nonrandomized experiments that attempts to mimic randomization by creating a control that is comparable on all observed measures.

the counterfactual is selected from an existing large database. Statistical techniques such as **propensity score** matching are employed to find good matches.

#### Integrated Development Considerations with Quasi-Experiments

When evaluating the impact of an integrated program and the potential for amplified effects in the absence of randomizing individuals to study arms, comparable treatment groups must be constructed for each sector intervention, the control group, and each possible combination of interventions. In other words, a valid matching process needs to occur among all arms of the study. Constructing one good equivalent group is difficult. Constructing three or more equivalent matching groups, as would be needed in a simple full factorial design, requires an exceptionally large sample from which to draw and match cases. Matching is one of the better fit quasi-experimental methods for integrated programs because it allows the evaluators to prioritize matching on key characteristics that may affect the outcomes with specific regard to integration. It also works well for arms that are made up of groups as opposed to individuals (such as health facilities or communities). This is beneficial for complex integrated projects that operate at a community level, and where it is infeasible to randomize or match at a lower level.

Interrupted time series designs involve repeated measures both before and after an intervention is implemented. Causal impact is assessed through statistical modeling on the basis of observable change after the intervention is introduced. One advantage of interrupted time series designs is that they allow the use of results for making adjustments to the program, with the effects of these adjustments then assessed through additional data waves. This may be particularly desirable with integrated programs whose complex nature may require adaptations to be made. Time series designs also allow for a historical control, which decreases the number of groups required. This is beneficial for integrated programs because, as mentioned previously with multiple interventions, the number of arms required increases dramatically. However, such designs require the measurement of as many data points before and after the intervention as possible, based on the desired resolution and anticipated time frame for observing intervention

#### **PRACTICAL EXAMPLE 8 Quasi-Experimental Fractional Factorial Design**

The Realigning Agriculture to Improve Nutrition (RAIN) project in Zambia was designed to improve stunting and malnutrition rates and to rigorously document the process and impact of implementation.<sup>24</sup> The project targeted children from conception through 23 months of age with an integrated agriculture, nutrition, and health project. This included district- and individuallevel agriculture activities, community-level health and nutrition promotional activities, group-level health promotional activities, and training on nutrition topics. The study used a hybrid design that combined a cluster randomized design and a plausibility design. Six wards were paired and then randomized into four intervention and two control wards. Some pre-defined areas within the intervention wards were changed at this point to better reduce the risk of spill-over effects and to support the ease of the operation of staff. Then the 26 remaining

areas were randomized into two different intervention arms. One arm received the agriculture intervention only, another arm received the agriculture, nutrition, and health interventions, and a third arm served as the control. The study was designed in this way because the focus of the evaluation was on the added value of integrating the agriculture with the health and nutrition activities, as compared to the agriculture activities alone. The impact of the health and nutrition activities on the population was not in question, so the extra arm that would have made this study a full factorial design was not necessary.

The study found that agricultural production *improved* across the groups (regardless of integration); the study also documented improvements in food security. There were small changes in some knowledge outcomes, and possible protective effects in wasting, but there was no measured impact on stunting for

either intervention arm. Importantly, the process evaluation demonstrated that the burden of integration was placed on the program beneficiaries, who had to make their own links between activities. The coverage in clusters receiving health and nutrition activities was very low (only 13 percent of households in the agriculture, nutrition, and health arm was visited in 2015 by a community health volunteer, for example). This evaluation demonstrates not only how thoughtful fractional factorial designs can be used to evaluate integrated programs, but also the importance of process evaluations in integrated efforts to document the success of integration and considering from the beginning the effect of integrating activities at different levels.

24 Harris J, Nguyen PH, Maluccio J, Rosenberg A, Mai LT, Quabili W, et al. RAIN project impact evaluation report. Zambia: International Food Policy Research Institute of the Realigning Agriculture to Improve Nutrition (RAIN) project. 2016 May [cited 2016 Aug 1]. Available from: https://www.concern.net/resources/finalreport-rain-project-zambia

Evaluating Impact 6

The holistic investigation

real-life context, focusing

on how causes interact to

allow an understanding

of a program within its

 $( \mathbf{V} )$ 

of effects

effects. This in turn may require extensive resources and also limits the applicability of this approach in the absence of sufficient preintervention data. See Table 2 for a summary of the strengths and limitations of experimental and quasiexperimental designs in an integrated context.

#### **Case Study**

A **case study** involves the holistic investigation of a program within its real-life context. Rather than individual causes or variables, as is most common with experimental

and quasi-experimental designs, case studies focus on understanding how causes interact to lead to effects. Case studies are particularly useful in identifying and explaining the wide variation of the effects that can occur in complex programs. They can also capture "statistically 'unobservable' factors, such as those stemming from social networks, motivation, legitimacy, expectations and power."25

#### In designing a case study, a choice will need to be made on the unit

25 Woolcock M. Using case studies to explore and explain complex interventions. 2016 Jul 20 [cited 2016 Aug 1]. In: World Bank Blog [Internet]. Available from: http://blogs. worldbank.org/impactevaluations/using-case-studiesexplore-and-explain-complex-interventions

#### TABLE 2

**Experimental and Quasi-Experimental Designs** in an Integrated Development Context

# STRENGTHS

They can rigorously and quantitatively answer the following questions in a way that can be attributed to the integrated programs:

- $\rightarrow$  Does 1 + 1 achieve outcomes > 2 (does integrating two interventions yield more than the sum of its parts)?
- $\rightarrow$  Does 1 + 1 offer operational benefits > 2 (does integrating two interventions produce operational benefits or efficiencies beyond the sum of its parts)?
- → Does 1 + 1 cost < 2?

# **X I IMITATIONS**

- $\rightarrow$  Implementation can be costly compared to other evaluation approaches
- $\rightarrow$  Only feasible under certain circumstances
- $\rightarrow$  The potential for diffusion between arms can be problematic
- $\rightarrow$  For quasi-experimental designs, may be difficult to construct equivalent groups
- $\rightarrow$  Does not identify the "how" and "why" of observed results
- → Does not capture context

of analysis of the case. Depending on the evaluation guestion, this could for example be an individual, an intervention, a program, the implementation process, or community change. Case study research can include single- or multiple-case designs. Single-case designs are used when cases are new, rare, or extreme. Multiple-case designs are used when a program is implemented at multiple sites, even where there is some level of adaptation across sites.

To evaluate outcomes and impacts of integrated development programs, case studies aim to expand and generalize theories and involve linking data to previously developed theoretical propositions such as a program's TOC. One common approach is **pattern matching**, in which case study data are compared to several possible patterns to determine which one fits best. Multiple case studies follow a replication logic and call for the individual analysis of each case, in addition to cross-case analysis. Replication can occur in two ways: cases providing similar results (literal replication) or cases providing contrasting results but for predictable reasons (theoretical replication). Either occurrence builds theoretical support across cases, but if cases are contradictory, the propositions must be revised and retested. Case studies can also be used to add depth and richness to the understanding of why the causal links did or did not occur.

### Integrated Development Considerations with Case Studies

Case studies are well-suited to the integrated development context. The holistic view is relevant to situations when effects are wide-ranging, cannot be clearly separated from context, and take time to occur. Case studies are appropriate to explain complex causal links when there is no clear, single set of outcomes, or when there is a deliberate purpose to address context to allow a full understanding of causes. Additionally, case studies do not require control over interventions as is required in experimental designs and thus may be a good strategy when control is not feasible. Whether cases are sites, households, or individuals, the methodological framework is similar to evaluate outcomes and impacts. Cases can be systematically selected from a larger sample, through random procedures, or based on explicit

criteria such as those exhibiting the most or least change (for example, the **most significant change** technique described below). Data from cases are then analyzed and aggregated to make causal inferences. In an integrated development context, case study designs can also be used to:

 $\rightarrow$  Identify the "how" and "why" of efficiencies, amplified or synergistic effects, and other consequences of integration

77

 $(\downarrow)$ 

#### PATTERN MATCHING

Case study approach in which case study data are compared to several possible patterns to determine which one fits best.

QUALITATIVE RESEARCH

 $(\mathbf{V})$ 

An inductive method of inquiry that employs open-ended questioning. Qualitative research generates data that are non-numeric in nature, primarily text and images.

- → Document the role of context on outcomes and experience
- → Document complex, non-linear, potential causal pathways
- → Describe processes behind observed changes within cases and identify tipping points

If a large enough number of cases (>10) are available, analysis can use Boolean logic techniques to identify common patterns across cases.

#### Systems Approaches

Changes in the local context and changes in relationships, power, or influence between actors or organizations are also important to assess in order to explain the overall context in which program-specific changes occurred (or did not occur). **Qualitative** approaches — like in-depth interviews, focus groups, participant observation, or inclusive participatory techniques — can be used to map systems pre - and post-intervention. The maps can then be qualitatively compared and changes can be noted. Some specific considerations for evaluating systems within an integrated context are listed below, along with two example methods.

Integrated Development Considerations with Systems Approaches

This guide proposes clear and deliberate planning of outputs, outcomes, and impacts of integrated programs, which would include the amplified and synergistic effects from integration. That said, the interaction effects of integrating interventions are often not well understood. Systems approaches can be used to augment the evaluation of the planned amplified and synergistic effects, as well as unanticipated changes to larger systems. Two methods that can be used to monitor and evaluate systems, and changes within them, are most significant change and outcome harvesting.

The most significant change (MSC) technique systematically

collects stories of significant change from program participants. A subset of stories is analyzed by a series of stakeholders who select those they consider most significant and who identify the criteria they use for the selection. The technique is flexible and adaptive to improvements and the context for implementation, with a focus on participatory approaches and getting at systemic change without pre-determined indicators.<sup>26</sup> MSC is best suited for complex programs with emergent and nuanced outcomes at the community level, such as integrated development programs. MSC offsets quantitative rigor with a greater understanding of what outcomes and impacts the project participants value and can help explain how and under which context change occurs. The technique is particularly helpful in an integrated context as it helps evaluators to identify unexpected changes, while understanding participant values and which are most important or significant for explaining program success as defined by participants.

**Outcome harvesting** is another technique that can be used to evaluate the synergistic and amplified effects of integrated programs. Outcome harvesting is particularly useful when outcomes are not sufficiently specified or cannot be predicted at the time of planning an intervention. Specifically, in the case of integrated development, it is not always clear what synergistic effects will emerge. Outcome harvesting can help discover these synergies and identify the program's contributing factors. Outcome harvesting is designed to capture outcomes — planned or unplanned, positive or negative — and understand how the program contributed to those outcomes.

27 Wilson-Grau R, Britt H. Outcome harvesting. Cairo: Ford Foundation; 2012 [cited 2016 Aug 1]. Available from: http:// www.managingforimpact.org/sites/default/files/resource/ outome\_harvesting\_brief\_final\_2012-05-2-1.pdf

26 Davies R, Dart J. The "Most Significant Change" (MSC) technique: a guide to its use. 2005 April [cited 2016 Aug 1]. Available from: http://www.mande.co.uk/docs/MSCGuide.pdf 79

Using outcome harvesting, the evaluator collects evidence of what has been achieved, and works backward to determine whether and how the project or intervention contributed to the change.<sup>27</sup> Evidence is collected from reports, personal interviews, and other sources to document how a given program or initiative has contributed to outcomes. The outcomes, or a sample of them, are verified by knowledgeable third parties. The method can also be used to judge the significance of the outcomes. There is an emphasis on effectiveness rather than on efficiency or performance. See Table 3 for a summary of the strengths and limitations of case studies and systems approaches in an integrated context.

#### Mixed-Methods Approaches

All methods bring strengths and weaknesses to the evaluation of integrated development programs. For example, **quantitative** methods accommodate large sample sizes and offer explanatory power, but tend to break programs into sub-parts. They also generate evidence that offers a relatively narrow understanding of the nuances of causality. Qualitative methods are rooted in the experience of typically smaller participant pools, but they provide in-depth,

# $\checkmark$

#### MOST SIGNIFICANT CHANGE TECHNIQUE

A qualitative participatory method used to evaluate complex interventions by systematically collecting stories of significant change from program participants, and then selecting stories that represent the most significant changes to analyze.



#### OUTCOME HARVESTING

A method that identifies, formulates, analyzes, and interprets the outcomes to answer useable questions. It is useful for detecting intended and unintended results and in complex contexts.



#### MIXED-METHODS

The combination of at least one quantitative and one qualitative method in a single research or evaluation project.



#### QUANTITATIVE RESEARCH

A method of inquiry that investigates observable phenomena through statistical, mathematical, or computational techniques.

contextualized insights to evaluate synergies, identify unanticipated consequences, and tease out contextual factors that might influence results (Table 4). Quantitative methods provide the most rigorous methods for measuring the amount of change while qualitative methods address the understanding of why.

Particularly for evaluating programs as complex as integrated development, combining methods has the potential to yield a more comprehensive inquiry and support stronger causal inferences.

#### TABLE 3

Case Study and Systems Approaches in an Integrated Development Evaluation Context

# **G**STRENGTHS

- $\rightarrow$  Can identify the "how" and "why" of  $\rightarrow$  Do not independently measure observed results
- $\rightarrow$  Take into account "context" and multiple potential causal pathways
- $\rightarrow$  Can identify unplanned results, including synergies from integration
- $\rightarrow$  Can be less costly than experimental designs

# × LIMITATIONS

- effects, efficiencies, or variation; likely rely on program monitoring data
- → Cases selected may not represent larger sample from which they are drawn, such that conclusions cannot be generalized
- $\rightarrow$  Bias of some funders, researchers, and evaluators toward *quantitative measures*

#### TABLE 4

A Comparison of Quantitative and Qualitative Approaches

#### **EVALUATING IMPACT WITH QUANTITATIVE APPROACHES**

- Structured surveys
- Biological specimens
- Network analysis

Quantitatively compare study arms and/or preand post-test measures of indicators.

Statistically assess differences between groups and relationships between variables.

Quantitatively measure systems changes (network analysis).

#### INTEGRATE WITH QUALITATIVE DATA

Can be used to inform sampling procedures (for example, select cases) and instrument design for follow-up qualitative data collection methods.

Can provide measure of variation and generalizability of qualitative results.

If both qualitative and quantitative data are collected closely in time and with the same intent, evidence can be triangulated. Evaluators can compare patterns in both sets of evidence and assess for convergence, divergence, or contradiction.

#### **EVALUATING IMPACT WITH QUALITATIVE APPROACHES**

#### **PRIMARY DATA COLLECTION METHODS**

- One-on-one qualitative interviews
- Focus groups
- Participant observation
- Participatory group methods (for example, World Café, MSC)
- Systems and process mapping

#### **PRIMARY ANALYTIC AIMS**

- Understand the "how," "why," and contextual factors associated with change.
- Identify patterns to infer causal mechanism(s), including tipping points.
- Qualitatively assess changes in systems (intended and unintended).

#### INTEGRATE WITH QUANTITATIVE DATA

- Can be used to inform sampling procedures and instrument design for subsequent quantitative data collection.
- Can be used to explain trends in, and provide context for, quantitative findings, including turning points in quantitative time-series.

81

# Evaluating Impact 6

82

# AFTER IMPLEMENTATION

# Evaluating Scale-Up



# **7 | EVALUATING SCALE-UP**

**Scale-up** is the process of expanding the reach of a successfully tested practice in order to benefit more people and to develop sustainable and institutionalized programs and policies long-term.



### **KEY QUESTIONS**

- $\rightarrow$  Does the pooling of resources by multiple sectors lead to more robust development solutions that can be sustained over time when implemented on a broad scale, under real-world conditions?
- $\rightarrow$  Do integrated models that show promise as proof-of-concept

pilots retain their added value and feasibility when scaled up?

 $\rightarrow$  Are there priorities or perspectives that supported integration at the pilot level that may not be present everywhere (and what can be done to mitigate these differences)?

# **KEY CONSIDERATIONS**

Evaluators must take into account that the coordination of different actors may differ from pilot to scale-up, as the feasibility of decision makers from multiple sectors conducting joint budget exercises or sharing supervisory responsibility will be influenced by different factors in different areas. Examining the success or failure of scale-up should include examination of these integrated management and coordination activities.

Organizations or partners supporting an integrated pilot intervention may not represent the motivations of their broader communities; preparing for scale-

up requires examination of support for cross-sector collaboration beyond the pilot area, to ensure that it is not merely an idiosyncrasy of a small group of actors.

Expectations should be managed for any project. Ideally, for the scale-up of an integrated project, managers should be able to assign interventions to sites that promise the most success, and should communicate that similar comparison sites may be difficult to find and that variation in implementation is expected. This variation is not necessarily a burden, but is rather an opportunity to better understand the factors favoring and impeding effective implementation of an integrated intervention.

A fundamental challenge in development is identifying those innovations that not only produce impressive results when implemented on a trial basis but also achieve comparable success when implemented broadly. As with any intervention, integrated projects proven to be effective in a single context (as evidenced, for example, from an impact evaluation) must be further evaluated when scaled up to determine their ultimate utility.

When development interventions are assessed with a view toward scale**up**, several dimensions are examined. These can be classified as dimensions of potential for scale-up (which informs decisions about replication or expansion) and success with scale-up (which assesses efforts at scaling up both in terms of process and outcomes).

Potential for scale-up refers to inherent characteristics of the development intervention that predict whether the intervention can be implemented on a broad scale with success. Regardless of whether the intervention is integrated, generally accepted relevant factors in this assessment include the following:

 $\rightarrow$  *Replicability:* Are the intervention components sufficiently clear, such that the elements required for complete and correct implementation are obvious? Can those components be feasibly duplicated by others?

→ Adaptability: Can the intervention be adjusted to correspond to the needs and resources in different contexts while maintaining elements essential to effectiveness? Can the intervention be managed by new implementing partners working with varying infrastructures and resources? Will the intervention appeal to the priorities and preferences of diverse beneficiaries?

 $\rightarrow$  Affordability: Are the gains achieved by the intervention of sufficient value to warrant the required resource investment? Is sustained and expanded investment realistic as the intervention is extended to more beneficiaries? Can any economies of scale be achieved, such that per-unit cost of output decreases as the intervention is implemented more broadly?

# $(\mathbf{V})$

#### SCALE-UP

The process of expanding the reach of a successfully tested practice in order to benefit more people and to develop sustainable and institutionalized programs and policies long-term.

# $(\mathbf{V})$

#### POTENTIAL FOR SCALE-UP

Inherent characteristics of the development intervention that predict whether the intervention can be implemented on a broad scale with success broad scale with success.



#### SUCCESS WITH SCALE-UP

A dimension assessed under scale-up evaluation that examines the efforts at scaling up in terms of both process and outcomes.

# Evaluating Scale-up **7**

Consideration of these factors illustrates how evaluation supporting the delivery of an integrated program at scale should begin well before that intervention is implemented on a broad basis. These factors are typically used to inform decisions being made about whether an intervention should be scaled up or not.

Other dimensions, such as the following, can be assessed only during the actual implementation of scale-up (see Practical Example 9):

- $\rightarrow$  *Fidelity*: To what extent and how consistently does the intervention conform to specifications that were in place when it achieved success on a small scale?
- → *Coverage:* What proportion of intended beneficiaries actually receive the intervention as intended?
- $\rightarrow$  Effectiveness at scale: When the intervention is replicated in diverse new contexts, to what extent does implementation achieve the same benefits achieved by the piloted intervention?

With regard to evaluating the scale-up of integrated, multisector interventions, the following additional considerations are important:

#### **INTER-SECTORAL COORDINATION:**

Integrated development programs are inherently complex, requiring effective coordination of practices to amplify effects through intended combinations of services. In evaluating the scale-up of integrated programs, that coordination function must be isolated and examined in its own right. Evaluators must assess, for example, the feasibility of decision makers from two different sectors conducting joint budgeting exercises, the acceptability of managers from two historically distinct programs sharing supervisory responsibilities, the challenges of integrated M&E in addition to component M&E, or the effectiveness of referral mechanisms between services. When integrated interventions do not achieve desired outcomes, it is important to examine whether there were shortcomings in the component interventions or whether the coordination mechanisms did not operate as intended.

#### **BROAD-SCALE MULTI-SECTOR**

**SUPPORT:** Multi-disciplinary partners in a single location might agree to collaborate on piloting an integrated development project on a limited scale. The decision to implement integrated solutions could be influenced by longstanding professional relationships at a local level. It is not safe to assume.

however, that individual partners supporting a pilot intervention represent the broader perspectives and priorities of their respective organizations or fields. To prepare for scale-up, the perspectives of stakeholders within each organization or sector should be examined to

Example 10).

#### **PRACTICAL EXAMPLE 9** Impact Evaluation to Determine Potential for Scale-Up

A team of investigators conducted six randomized controlled trials in six sites to test the effectiveness of a package of complementary interventions aimed at reducing poverty among the very poor.<sup>28</sup> The intervention package provided an income-generating asset such as livestock, a cash stipend, training in running a business, health education, and improved access to health services. The study's primary objective was to assess the effect of the intervention package on the level of consumption and the psychosocial well-being of the targeted households. The study found statistically significant improvements in all 10 key outcomes; positive

results were maintained for eight of those outcomes one year following completion of intervention support.

Additionally, the trial had two important features that permitted the investigators to draw optimistic conclusions about the potential for implementing the intervention package at scale. By testing the intervention in six countries (Ethiopia, Ghana, Honduras, India, Pakistan and Peru) with six different implementation partners, the research team was able to examine and confirm the adaptability of the integrated solution to diverse contexts. The investigation further included cost-benefit analysis of the intervention variations,

ensure that support for cross-sector collaboration is in fact pervasive, and not merely an idiosyncrasy of a small group of actors. Such a stakeholder analysis can reveal the partners that need support and the approach to advocacy that is required (see Practical

> showing that the estimated benefits were higher than the costs in five out of six sites. Based on the impressive results of the multi-site trial and the other factors indicating an adaptable, cost-effective intervention, the investigators concluded that similar positive results might be achieved *if the intervention were* implemented at broader scale. Although the intervention was complex and there were many outcomes to monitor in diverse settings, the evaluation team was able to demonstrate scalability.

28 Banerjee A, Duflo E, Goldberg N, Karlan D, Osei R Parienté W et al A multifaceted program causes lasting progress for the very poor: evidence from six countries. Science, 2015 May 14;348(6236): 1260799–1260799. doi: 10.1126. 87

Evaluating Scale-up **7** 

#### MANAGING EXPECTATIONS ABOUT **ASSESSING THE IMPACT OF AN** INTEGRATED DEVELOPMENT

**PROGRAM:** The term "scale-up" typically implies that services are implemented not only broadly, but also under real-world conditions. A consequence is that the conditions required to scientifically measure an intervention's impact are rarely in place. To manage expectations, the following considerations are important:

- $\rightarrow$  Integrated solutions are not meant to be implemented randomly. Ideally, managers assign interventions to sites intentionally, following thoughtful consideration of the sites promising the most success.
- → Highly similar comparison sites that can serve as valid counterfactuals are difficult to find.
- → Variation in implementation can be expected. Program designers and managers must strike a balance between maintaining the integrity of a piloted intervention and adapting it to the needs and resources of the local context.

#### **Focus of Evaluations** for Scale-up

Despite these constraints, examination of scale-up makes essential contributions to the integrated development field and of course the programs themselves. Evaluations must focus on the following:

- $\rightarrow$  Documenting feasible, acceptable integration models, adapted to varying contexts
- $\rightarrow$  Monitoring conformity of implementation to the intended model for a given context
- $\rightarrow$  Identifying the factors favoring and impeding effective implementation of the integrated intervention
- $\rightarrow$  Monitoring variations in outcomes over time for a given context or across contexts at a set time
- $\rightarrow$  Analyzing the factors associated with successful outcomes

Collectively, these lines of inquiry will produce evidence to support continuous refinement of the integration model, as well as decision making about continued expansion of the program.

#### **PRACTICAL EXAMPLE 10 PHE Scale-Up**

The field of populationhealth-environment (PHE) provides a rich body of experience on integrated development, with important lessons about evaluating for scale-up. PHE refers to projects that combine complementary interventions that help couples achieve their desired family size (population); promote, for example, maternal and child health or HIV prevention (health); and improve conservation of local resources (environment). Regions with rich biodiversity at risk of degradation are prioritized for PHE initiatives, which are designed to protect the environment while improving community health and well-being.

USAID commissioned a 2007 case study to examine exemplary expansion of PHE initiatives in Madagascar over nearly two decades.<sup>29</sup> The methods consisted of review of published, gray, and

web-accessible literature; field visits; and collaborative work with stakeholders over a two-year period. The investigation documented how collective support from a host of actors made possible the scale-up of PHE initiatives throughout this island nation. Influential stakeholders included the following:

- $\rightarrow$  Madagascar's (former) president, an outspoken advocate for both conserving Madagascar's unique biodiversity and combatting poverty by reducing family size
- $\rightarrow$  A multi-donor group focused on environment, rural development, and food security engaged in sustained dialogue at high levels with the Government of Madagascar
- → Nongovernmental organizations with their hands-on experience

that afforded them knowledge of both the practical challenges encountered in implementing integrated solutions and the benefits realized by communities

 $\rightarrow$  Communities engaged in the national "Champion Commune" movement supporting local ownership and implementation of integrated community development initiatives

This case study revealed how the success of PHE scale-up in Madagascar can be attributed to a common commitment across these levels, supported by national and local policies, resource allocations, governing mechanisms, and operational tools.

29 Gaffikin L. Scaling up Population and Environment Approaches in Madagascar: A case study. Washington, DC: World Wildlife Fund and Evaluation and Research Technologies for Health (EARTH) Inc.; 2008.

89

# Evaluating Scale-up **7**

90

# Conclusion

91

#### Conclusion

Integrated programming offers unique opportunities for time savings, value for money, enhanced human resources, strengthened capacity of local community or government, provider or user satisfaction, reach, equity, sustainability, amplified effects, and systems change. The integrated nature of the programming also has potential challenges and costs. Indeed, costs for M&E are likely to be greater than for vertical programs.

Evaluation is a valuable tool in making evidence-based judgements about the comparative value of integrated versus vertical programming and therefore whether and how to implement integrated development programs. Importantly, the desired goal of integration varies greatly depending on the perspective, priorities, and ultimate aim of a given decision maker. Funders may emphasize cost efficiencies or enhanced sustainability, whereas program implementers may prioritize time savings, improving user satisfaction, or reducing inequality. Therefore, the evidence they require for informed decision making will vary in nature.

Formative research can help reveal the scenarios and problem sets that are best suited for integrated approaches, and where the need is greatest. Process and impact evaluations can offer proof-of-concept findings to test the feasibility and results of innovative integrated interventions. Implementation science can identify best practices for the replication or scale-up of proven multi-sector models. High-quality monitoring and evaluation within non-research settings can help assess progress and guide subsequent adaptations and improvements. Cost analyses can help to identify the components of a multifaceted program that

offer the best return on investment, and the most efficient means to a desired outcome. Generally speaking, greater complexity and diversity coupled with a focus on interaction or amplification effects — have implications for how integrated development programs are monitored and evaluated. This affects all components of an evaluation - from developing logic models and costing approaches to choosing indicators and an evaluation design. Additional research and evaluation objectives and questions, specific to integration, also need to be considered.

This guidance document raises issues for strategic consideration and provides suggestions derived from experience to help support evaluators, funders, and development practitioners in evaluating the unique aspects of integrated development approaches. Our hope is that this document will be used to improve the evaluation of integrated programs so that we take the opportunity to learn from what is being done now, and so that the evidence base around integration continues to grow.

# Conclusion

94

# Appendices

# APPENDIX A

#### **Resources, Literature, and Guidance on Evaluating Integrated Approaches**

A	Adamou B, Curran J, Wilson L, Apenem Dagadu N, Jennings V, Lundgren R, et al. Guide for monitoring scale-up of health practices and interventions. Chapel Hill (NC): MEASURE Evaluation; 2013.	This guide is intended to provide governments, donors, country organizations, and implementing partners with an approach to monitoring the scale-up process in health.
В	Davies R, Dart J. The "Most Significant Change" (MSC) technique: a guide to its use. 2005 April [cited 2016 Aug 1]. Available from: http://www.mande.co.uk/docs/MSCGuide.pdf	This document not only provides an in-depth overview of the MSC technique and a guide to implementation, but also provides many thoughtful considerations for putting it to practice within an existing framework.
С	de Pinho H. Participant guidelines. Systems tools for complex health systems: a guide to creating causal loop diagrams. NY: Columbia University. 2015 Feb [cited 2016 Aug 1]. Available from: http://www.who.int/alliance- hpsr/resources/publications/CLD_Course_ Participant_Manual.pdf	Using simple steps, this course guide demonstrates how to use systems thinking to develop causal loop diagrams (participant and facilitator manual available).
D	Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW Methods for the economic evaluation of health care programmes. 4th ed. Oxford, UK: Oxford University Press; 2015.	This textbook explains different types of economic evaluations within the health sector and the data requirements and methods required for each.
E	FXB village toolkit and planning guide. FXB International; 2015 [cited 2016 Aug 1]. Available from: https://usa.fxb.org/wp- content/uploads/FXB_Toolkit.pdf	This toolkit describes the organization's approach for monitoring and measuring the progress of the families served by the program. The program consists of a series of "FXB villages," each of which is a structured, three-year, holistic program that supports 80 to 100 households with a package of essential health, education, and financial supports.
F	Gertler PJ, Martinez S, Premand P, Rawlings LB, Vermeersch CMJ. Impact evaluation in practice. World Bank; 2011 [cited 2016 Aug 1]. Available from: http:// siteresources.worldbank.org/EXTHDOFFICE/ Resources/5485726-1295455628620/ Impact_Evaluation_in_Practice.pdf	The World Bank provides comprehensive guidance on implementing impact evaluations for development practitioners and policymakers.
G	Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. New York (NY): Oxford University Press; 1996.	This book provides an expert-panel consensus approach for the conduct of cost-effectiveness analysis of health and medical interventions in the United States.

н	Guest G, Namey EE, Mitchell ML. Collecting qualitative data: a field manual for applied research. Thousand Oaks (CA): Sage; 2013.	This boo collectio participa The boo qualitati
I	Lance P, Guilkey D, Hattori A, Angeles G. How do we know if a program made a difference? A guide to statistical methods for program impact evaluation. Chapel Hill (NC): MEASURE Evaluation; 2014.	With a fo various evaluati
J	Masters WA, Webb P, Griffiths JK, Deckelbaum RJ. Agriculture, nutrition, and health in global development: typology and metrics for integrated interventions and research. Ann. N.Y. Acad. Sci. 2014; 1331:1–12.	Specific and nutr and met other's r design o
К	Moreland S, Paxton A. Framework for population, health, and environment programs. Washington, DC: Futures Group, Health Policy Project; 2015 September [cited 2016 Aug 1]. Available from: http:// www.healthpolicyproject.com/pubs/1877_ HPPPHEFrameworkFinal.pdf	In respo by which funded H interact and to sl multi-se generali to any in
L	Patton MQ. Utilization-focused evaluation. 4th ed. Thousand Oaks (CA): Sage Publications; 2008.	This infl conduct and app
Μ	Patton MQ. Developmental evaluation applying complexity concept to enhance innovation and use. New York (NY): Guilford Press; 2010.	Develop particula facilitate staff, the
N	Preskill H, Gopal S, Mack K, Cook J. Evaluating complexity: propositions for improving practice. FSG; 2014 [cited 2016 Aug 1]. Available from: http://www.fsg.org/ publications/evaluating-complexity	This brie change, describe better so on ways

bk is a practical how-to manual for qualitative data on. It covers the three primary qualitative methods ant observation, in-depth interviews, and focus groups. ok also explains other methods that can be used in ive inquiry such as visual and graphical techniques.

ocus on statistical methods, this guidance presents methods when approaching the question of ing impact.

ally focused on the integration of agriculture, health, rition, this paper proposes a typology of interventions trics of integration to help researchers build on each results, facilitating integration of methods to inform the of multi-sector interventions.

nse to the need for a standard analytical framework h to evaluate integrated PHE programs, the USAID-Health Policy Project developed this tool to define the cions among interventions in each of the three sectors show the synergies that can result from an integrated, ector approach. The Health Policy Project built a ized computer model and framework that can be applied ntegrated PHE program.

luential evaluation text offers detailed advice on ting program evaluation through varied methods roaches.

omental Evaluation is an evaluation approach that is arly suited to complex or uncertain environments. It es real-time, or close to real-time, feedback to program us facilitating a continuous development loop.

This brief "brings together what we know about systems change, complexity, and evaluation in a way that clarifies and describes how the practice of evaluation needs to evolve to better serve the social sector." It provides overall direction on ways to conceptualize, design, and implement evaluations of complex initiatives and/or initiatives that operate in complex settings.

Appendices

# APPENDIX B

ο	Rossi PH, Lipsey MW, Freeman HE. Evaluation: a systematic approach. 7th ed. Thousand Oaks (CA): Sage Publications; 2004.	This text provides a basic background on designing, implementing, and evaluating social programs. It includes many examples, techniques, and approaches, as well as guidance on how to fit evaluations to many different contexts and needs.
Ρ	Stern E, Stame N, Mayne J, Forss K, Davies R, Befani B. Broadening the range of designs and options of impact evaluations. Department for International Development (DFID); 2012 Apr [cited 2016 Aug 1]. Working Paper 38. Available from: https://www.gov. uk/government/uploads/system/uploads/ attachment_data/file/67427/design-method- impact-eval.pdf	This document details impact evaluations in the current landscape, and then considers broader methods and designs that link cause and effect. The authors detail the challenges of complex development programs, and the kinds of questions that can be asked and answered for complex programs through the range of designs for determining causal inference.
Q	UNAIDS. Manual for costing HIV facilities and services. Geneva, Switzerland: UNAIDS. Report. 2011.	This report provides a step-by-step guide and example data formats for those interested in costing HIV services.
R	Weiss CH. Evaluation: methods for studying programs and policies. 2nd ed. Englewood Cliffs (NJ): Prentiss Hall; 1998.	Focused solely on evaluation, this text offers accessible and diverse guidance for evaluation at many different stages drawn from diverse fields of study.
S	Westhorp G. Realist impact evaluation: an introduction. London: Overseas Development Institute, Methods Lab; 2014 Sept [cited 2016 Aug 1]. Available from: https://www.odi.org/ sites/odi.org.uk/files/odi-assets/publications- opinion-files/9138.pdf	This paper examines the role of context in developing impact evaluations, and outlines how to design and conduct an impact evaluation based on a realist approach.
т	White H, Shagun S. Quasi-experimental design and methods. Methodological Briefs: Impact Evaluation No. 8. Florence: UNICEF Office of Research; 2014 [cited 2016 Aug 1]. Available from: https://www.unicef-irc.org/publications/ pdf/brief_8_quasi-experimental%20design_ eng.pdf	This brief details standard quasi-experimental methods for data collection and analysis, along with general good practices and challenges when using quasi-experimental methods.
U	Wilson-Grau R, Britt H. Outcome harvesting. Cairo: Ford Foundation; 2012 [cited 2016 Aug 1]. Available from: http://www. managingforimpact.org/sites/default/ files/resource/outome_harvesting_brief_ final_2012-05-2-1.pdf	This brief serves as an introduction to outcome harvesting, with examples for possible application.

# Planning an integrated agriculture and health program and designing its evaluation: Experience from Western Kenya

Donald C. Cole<sup>a,b,\*</sup>, Carol Levin<sup>c</sup>, Cornelia Loechl<sup>d</sup>, Graham Thiele<sup>e</sup>, Frederick Grant<sup>g</sup>, Aimee Webb Girard<sup>f</sup>, Kirimi Sindi<sup>h</sup>, Jan Low<sup>h</sup>

Multi-sectoral programs that involve stakeholders in agriculture, nutrition and health care are essential for responding to nutrition problems such as vitamin A deficiency among pregnant and lactating women and their infants in many poor areas of lower income countries. Yet planning such multi-sectoral programs and designing appropriate evaluations, to respond to different disciplinary cultures of evidence, remain a challenge. We describe the context, program development process, and evaluation design of the Mama SASHA project (Sweetpotato Action for Security and Health in Africa) which promoted production and consumption of a bio-fortified, orange-fleshed sweetpotato (OFSP). In planning the program we drew upon information from needs assessments, stakeholder consultations, and a first round of the implementation evaluation of a pilot project. The multidisciplinary team worked with partner organizations to develop a program theory of change and an impact pathway which identified aspects of the program that would be monitored and established evaluation methods. Responding to the growing demand for greater rigour in impact evaluations, we carried out quasiexperimental allocation by health facility catchment area, repeat village surveys for assessment of change in intervention and control areas, and longitudinal tracking of individual mother-child pairs. Mid-course corrections in program implementation were informed by program monitoring, regular feedback from implementers and partners' meetings. To assess economic efficiency and provide evidence for scaling we collected data on resources used and project expenses. Managing the multi-sectoral program and the mixed-methods evaluation involved bargaining and trade-offs that were deemed essential to respond to the array of stakeholders, program funders and disciplines involved.

#### Available from: http://www.sciencedirect.com/science/article/pii/S0149718916300581

Cole DC, Levin C, Loechl, C, Thiele G, Grant F, Girard AW, et al. Planning an integrated agriculture and health program and designing its evaluation: experience from Western Kenya. Evaluation and Program Planning. 2016 Jun;56:11-22 [cited 2016 Aug 1]. Available from: http://www.sciencedirect.com/science/article/pii/S0149718916300581

#### **ARTICLE INFO**

Article history:

Received 4 May 2015 Received in revised form 5 March 2016 Accepted 7 March 2016 Available online 14 March 2016

Keywords:

Multi-sectoral program Nutrition Maternal and child health Agriculture Micronutrient Vitamin A Bio-fortification Development evaluation

- a Dalla Lana School of Public Health, University of Toronto, Toronto, Canada
- b International Potato Center (CIP), Peru
- c University of Washington, Seattle, USA
- d International Atomic Energy Agency, Vienna, Austria
- e CGIAR Research Program on Roots, Tubers and Bananas, Lima, Peru
- f Rollins School of Public Health, Emory University, GA, USA
- g CIP, Dar es Salaam, Tanzania
- h CIP, Nairobi, Kenya

Appendices

100

# Glossary

**ANNUALIZATION FACTOR** – Similar to depreciation, the annualization factor is used to spread the acquisition cost of a resource, which lasts multiple years across its expected useful life. Unlike depreciation, it also considers the value of the money tied up in the investment that is no longer available for other uses (opportunity cost of the investment).

**AMPLIFIED EFFECTS** – A potential benefit of integration where the act of integrating produces a result that is larger than what would have occurred in vertical programs.

**APPRECIATIVE INQUIRY** – "[A] group process that inquires into, identifies and further develops the best of 'what is' in organizations in order to create a better future. Often used in the organization development field as an approach to large-scale change, it is a means for addressing issues, challenges, changes and concerns of an organization in ways that build on the successful, effective and energizing experiences of its members."30

#### С

**CASE STUDY** – The holistic investigation of a program within its real-life context, focusing on how causes interact to allow an understanding of effects.

**COMPLEX ADAPTIVE SYSTEM** – A set of interrelated, interacting, and interdependent components and subsystems that make up a whole and can adapt, learn, and change over time.

**COMPONENTS** – The pieces that make up a system, including stakeholders, activities, and settings.

30 Preskill H, Catsambas TT. Reframing evaluation through appreciative inquiry. Thousand Oaks (CA): Sage Publications; 2006

**COST ANALYSIS** – Technique for documenting the extent to which any operational benefits occur in integrated programs and the size of these gains.

**COST-EFFECTIVENESS ANALYSIS** – A form of economic analysis that compares the costs and outcomes of two or more courses of action.

**COUNTERFACTUAL** – A control or comparison group that matches the treatment group on as many dimensions and variables as possible.

#### E

**ECONOMIC COSTS** – Costs that consider the opportunity cost of securing a resource. A resource is assigned a unit value equal to what it would cost to obtain the resource in the local market. A donated input would be valued at what the user would need to pay to obtain that resource if it was not being provided to the program at no cost.

**EXPERIMENTAL DESIGN** – CRandomly allocating interventions to a sample of participants either individually or as part of a larger group or cluster (such as a community).

#### E.

FINANCIAL COSTS – Costs that consider the transfer of funds required to secure the use of a resource. Resources for which no transaction occurs (donations) would be assigned a unit cost equal to zero.

FORMATIVE RESEARCH – Research that informs program content, design, and operations prior to implementation. The primary task is to determine the key problem or nexus of problems to be addressed, and opportunities and ideas for addressing them.

#### FRACTIONAL FACTORIAL DESIGN -

An experimental design consisting of a specific subset, or fraction, of the possible arms of a full factorial design.

#### FULL FACTORIAL INTEGRATED

**DEVELOPMENT DESIGN** – Research design in which participants are randomized to 1) a control group (no intervention), 2) a single intervention arm for each activity included in the study, or 3) a multi-intervention arm for each permutation of integration. The simplest full factorial design for integrated development evaluation will include four arms: one control, one for the first activity, one for the second activity, and one for the integrated activities. In such a design, if true amplification is achieved, the integrated arm(s) should show a degree of change that is greater than the sum of change among all of the vertical

#### G

arms.

**GOAL** – The desired long-term result of the program.

causal attribution.

**IMPACT** – The "positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended."31

#### **IMPACT EVALUATION** – An evaluation that assesses the outcomes and impact that can be attributed to a particular intervention by comparing outcomes between intervention group(s) and a counterfactual (control group). For the purposes of this document, this is limited to experimental and quasi-experimental evaluation designs, which rigorously determine

31 OECD-DAC. Glossary of key terms in evaluation and results based management. Paris: OECD; 2002.

L

Μ

102

**IMPACT PATHWAY** – A planning tool for evaluating a program that includes many of the same components of a logic model, but also shows the progression or chain of what the project will do, through outputs that achieve outcomes and the desired impact.

#### INCREMENTAL COST-EFFECTIVENESS

**RATIO** – A ratio that summarizes costeffectiveness analysis by calculating the change in total cost divided by the change in the outcome or impact indicator, resulting in a dollar amount per unit change in the indicator.

**INDICATOR CREEP** – The inflation of obligations as new measurements are required to account for expansion of activities.

#### **INTEGRATED DEVELOPMENT** – An

intentional approach that links the design, delivery, and evaluation of programs across sectors to produce an amplified, lasting impact on people's lives.

**LOGIC MODEL** – A planning tool for evaluating the effectiveness of a program that lays out what the program hopes to accomplish and its intended impact. It includes the following basic components: goal, resources/inputs, activities, outputs, outcomes, and impact.<sup>32</sup>

MIXED-METHODS – The combination of at least one quantitative and one qualitative method in a single research or evaluation project.

32 W.K. Kellogg Foundation. Using logic models to bring together planning, evaluation, and action: logic model development guide. Battle Creek (MI): W.K. Kellogg Foundation; 2004 cited 2016 Aug 1]. Available from: http://www.smartgivers.org/ uploads/logicmodelguidepdf.pdf

104

# MOST SIGNIFICANT CHANGE

(MSC) TECHNIQUE – A qualitative participatory method used to evaluate complex interventions by systematically collecting stories of significant change from program participants, and then selecting stories that represent the most significant changes to analyze.

#### 0

**OUTCOMES** – The measurable short- and medium-term result(s) or effect(s) of program activities.

**OUTCOME HARVESTING** – A method that identifies, formulates, analyzes, and interprets the outcomes to answer useable questions. It is useful for detecting intended and unintended results and in complex contexts.

**OUTPUT** – The products, goods, and services that results from an intervention.

**OPPORTUNITY COST** – A value assigned to a resource that reflects the value of the resource in its best alternative use. For example, a person's time spent waiting is valued according to what that person could earn through paid employment but cannot because he or she is waiting.

#### Ρ

**PATTERN MATCHING** – Case study approach in which case study data are compared to several possible patterns to determine which one fits best.

**PERFORMANCE INDICATORS** – The precise measures used to assess success by a program or activity.

#### PERFORMANCE MONITORING PLAN

(PMP) – A document designed to support the measurement of progress in achieving the results specified in the program's logic model and specifying performance indicators, definitions, data sources, reporting frequencies and responsibilities, and targets.

**POTENTIAL FOR SCALE-UP** – Inherent characteristics of the development intervention that predict whether the intervention can be implemented on a broad scale with success.

**PRIMARY DATA** – Information collected directly by a program.

**PROCESS EVALUATION** – A method of assessing and understanding how a program is being implemented, focusing on the program's operations and service delivery.

**PROGRAM MONITORING** – The routine, systematic observation and recording of program implementation and problems using the performance indicators developed and other monitoring processes, including analysis and feedback about the progress of the program to the stakeholders.

#### **PROPENSITY SCORE MATCHING – A**

statistical matching technique used in nonrandomized experiments that attempts to mimic randomization by creating a control that is comparable on all observed measures.

**PROXY INDICATOR** – A stand-in or approximate indicator for an indicator of interest. It may use an alternative definition and/ or be derived from a different data source.

#### Q

**QUALITATIVE RESEARCH** – An inductive method of inquiry that employs open-ended questioning. Qualitative research generates data that are non-numeric in nature, primarily text and images.

#### QUALITY-ADJUSTED LIFE YEAR - A

common measure of effectiveness used in costeffectiveness analyses of health interventions that considers not only life expectancy but also the "quality" of the remaining life years based on the presence of diseases or other conditions.

**QUANTITATIVE RESEARCH** – A method of inquiry that investigates observable phenomena through statistical, mathematical, or computational techniques.

QUASI-EXPERIMENTAL DESIGN – A research design in which exposure to the intervention is not determined by a random process; instead, a counterfactual not exposed to the intervention is constructed in such a way that it matches the treatment group on as many dimensions and variables as possible.

#### R

**RESULT** – A consequence, effect, or conclusion of something.

#### **RIPPLE-EFFECT MAPPING** – Uses elements

of appreciative inquiry, mind mapping, and qualitative data analysis to engage individuals to map the intended and unintended changes a program targets.

by a program.

S

**SCALE-UP** – The process of expanding the reach of a successfully tested practice in order to benefit more people and to develop sustainable and institutionalized programs and policies long-term.

**SUBSYSTEM** – A self-contained group of interrelated, interacting, and interdependent components within a larger system (such as the primary health care system or tertiary education system).

**SUCCESS WITH SCALE-UP** – A dimension assessed under scale-up evaluation that examines the efforts at scaling up in terms of both process and outcomes.

**SYSTEM** – A set of connected components or elements that form a complex whole.

**SYSTEMS THINKING** – A perspective in which a range of approaches and methods that aim to describe and develop an understanding of the underlying structure of a system are used to make inferences about the system, to develop programs that work best in the system, and to most effectively affect change.

**SECONDARY DATA** – Information derived from other sources rather than collected directly

#### SECTOR-SPECIFIC INDICATORS -

Indicators, sometimes standardized, that are either required or recommended for programs to collect and that may vary by funder.

**SYNERGY** – A potential benefit of integration where the act of integrating produces a result that would not have been seen in vertical programs.

#### Τ.

**TARGETS** – The amounts of expected change within a given timeframe.

#### THEORY OF CHANGE (TOC) -

The description of a sequence of events that is expected to lead to a particular desired outcome. It typically includes the context for the initiative, the long-term change the initiative supports and from which it will ultimately benefit, the process/sequence of change anticipated to lead to the desired long-term outcome, assumptions about how these changes might happen, and a diagram and narrative summary that captures the outcomes.<sup>33</sup>

#### V

**VALUE-ADDED INDICATORS** – Indicators that measure a synergy or amplified effects above and beyond what would have occurred in a vertical program.

#### VERTICAL PROGRAM/ACTIVITY -

An activity that is designed, supervised, and executed, either wholly or to a great extent, by a specialized, often sector-specific group with a focused purpose.

#### W

#### WHOLE-SYSTEM-IN-THE-ROOM

**WORKSHOP** – A method that convenes a diverse set of stakeholders from a variety of sectors that work on or are affected by the development problems that the program plans to address.

107

<sup>33</sup> Davies R. Rick on the Road Blog [Internet]. Criteria for assessing the evaluability of a theory of change. 2012 Apr 5 [cited 2016 Aug 1]. Available from: http://mandenews. blogspot.co.uk/2012/04/criteria-for-assessing-evaluability-of. html Vogel I (UK Department of International Development). Review of the use of "Theory of Change" in international development. Review report. 2012 Aug [cited 2016 Aug 1]. Available from: http://rdd.fdid.gov.uk/pdf/outputs/mis\_spc/ dfid\_toc\_review\_vogelv7.pdf



fhi360.org/integrated-development



#### **SEPTEMBER 2016**

FHI 36O is a nonprofit human development organization dedicated to improving lives in lasting ways by advancing integrated, locally driven solutions. Our staff includes experts in health, education, nutrition, environment, economic, development, civil society, gender, youth, research, technology, communication and social marketing — creating a unique mix of capabilities to address today's interrelated development challenges. FHI 36O serves more than 7O countries and all U.S. states and territories.

For more information:

**GREGORY BECK** Integrated Development Director askid@fhi360.org