MONITORING AND EVALUATION GUIDELINES FOR WOMEN'S ECONOMIC EMPOWERMENT PROGRAMS

JUNE 2015

REPORT PREPARED BY JAMES KNOWLES

This resource was prepared as a companion to the report, "Measuring Women's Economic Empowerment", published by the United Nations Foundation and the ExxonMobil Foundation. It has benefited from the contributions of Mayra Buvinic, Alba Bautista, Emily Pryor and Katy Johnson of the United Nations Foundation and Noa Gimelli of the ExxonMobil Foundation. Bruce Ross-Larson and Communications Development Incorporated designed and copy edited this document.

CONTENTS

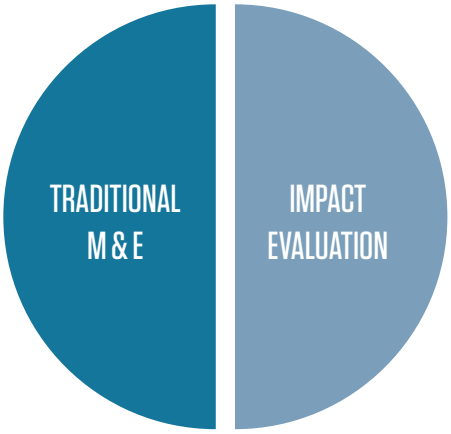
INITIALS

CBA	Cost-benefit analysis
CEA	Cost-effectiveness analysis
IRB	Institutional review board
M&E	Monitoring and evaluation
NGO	Nongovernmental organization
RCT	Randomized controlled trial
WEE	Women's economic empowerment

EXECUTIVE SUMMARY

These guidelines describe the role of monitoring and evaluation (M&E) in programs to empower women economically. That role is assumed to be threefold: to support effective project implementation ("Are we doing things right?"), to determine whether the desired outcomes are being achieved ("Are we doing the right things?"), and to contribute to the global knowledge base on the types of interventions that are most effective in promoting women's economic empowerment ("Do we know what works best?").

The primary audience for the guidelines are implementers and funders of women's economic empowerment (WEE) programs. The purpose is to promote a harmonized approach to M&E in WEE programs, and to provide a common framework for measuring and communicating program outcomes. But the guidelines do not recommend that all WEE programs use the same M&E procedures. Instead, they provide a menu of M&E methods that different WEE program implementers and funders can use to meet their individual needs.



The three main categories of M&E are traditional M&E, impact evaluation, and performance evaluation. The guidelines focus on the first two items. The third is a program (or organization) activity that is typically conducted after the completion of projects to draw lessons from multiple projects that can inform the design of new projects. Traditional M&E (simply "M&E" in the guidelines) is most often conducted by projects to assess how effectively their interventions are being implemented, whether the intended groups are receiving project benefits and at what cost, and to identify and correct any problems discovered as quickly as possible. Impact evaluation estimates the causal effects of an intervention on key outcomes and assesses whether the intervention is obtaining value per dollar spent. Although impact evaluation is a program-level activity, it requires effective collaboration between the evaluators and those implementing an intervention.

OUTCOMES

These guidelines define WEE as an increase in women's productivity, income, and wellbeing. Based on this definition, a set of direct outcomes, intermediate outcomes, and final outcomes is identified for two groups of women: urban women entrepreneurs and business leaders, and rural women entrepreneurs and farmers. The guidelines follow six principles:

1. Given the interdependence of women's economic and social roles, it is important to measure both economic and social outcomes to understand women's economic empowerment.
2. It is also important to measure both individual and household effects, considering the broader context of women's well-being in the household.
3. The what and the how of an evaluation matter equally: "what" refers to the outcomes measured, "how" to the evaluation design.
4. No evaluation is better than a poorly designed evaluation.

5. Not every program can be rigorously evaluated, but we can learn something of value from every program.
6. Complementary qualitative work is important to understand the "why" behind results, which can be quite useful for program staff.

CAUSAL CHAIN AND THEORY OF CHANGE

A first step in developing an effective M&E framework is to set down the *causal chain* linking the activities of an intervention to its outputs and outcomes. The causal chain includes an implementation segment of activities that transform inputs into outputs that lead to outcomes and a results segment of direct, intermediate, and final outcomes. As an intervention moves along the causal chain, the outcomes become less subject to the control of those implementing or benefiting from an intervention. Indeed, final outcomes (or "impacts") may depend on external factors, such as climatic or economic fluctuations, or on characteristics of beneficiaries that are not easy to change, such as social and cultural factors or education levels. Final outcomes may also require considerable time to materialize, or they may be very difficult to measure. In such cases, intermediate outcomes can be proxies for final outcomes.

The causal chain is based on a *theory of change* that describes how the intervention is to deliver the desired results and provides the rationale for the causal sequence from inputs to final outcomes. The theory of change draws on an understanding of the context for implementing the intervention and on previous experience with the same or similar interventions in similar settings. It is often developed with stakeholders, both to draw on their expertise and to ensure that implementation proceeds from a common understanding of the program, its objectives, and how it works. And it makes explicit the assumptions and risks for each link in the causal chain.

INDICATORS

The *outcomes* in a causal chain are typically expressed in broad terms that reflect targeted changes in individual behavior or organizational performance. A critical step in developing an effective M&E framework is to identify *indicators* (variables) that can measure the outcomes. Unlike outcomes, indicators are neutral about the direction of change and do not incorporate targets. Instead, baseline and target values of each indicator are provided separately to indicate clearly the desired direction of change (or that the baseline value is to be maintained).

Indicators can be quantitative or qualitative. *Quantitative indicators* are directly measurable and have either cardinal values indicating relative quantities (such as household income, hours worked during the past week, score achieved on a test) or categorical values (such as sex, housing characteristics). *Qualitative indicators* are often subjective and have categorical values that may or may not indicate an ordinal ranking (such as very satisfied, somewhat satisfied, not at all satisfied). Quantitative indicators are much preferred to qualitative indicators because they are easier to "verify" (confirm by independent measurement) and because they are directly comparable between individuals. However, some important outcomes

relevant to women's economic empowerment can be measured only subjectively, with the responses indicating only a relative ranking (such as overall satisfaction with life, self-confidence, stress).

Selecting suitable indicators for a given intervention can be challenging. One important criterion is that the number of indicators should be limited. As a general rule, no more than one indicator should be used to measure a given outcome. But if the outcome is multidimensional, it is better to use two indicators than to combine two distinct dimensions into a single indicator. In other words, indicators should be simple, clear, and adequate to represent a given outcome. Another important criterion is that indicators should be practical—that is, timely and good-quality data should be available and affordable to measure the indicator. Other criteria for selecting indicators include:

- *Validity*: Does the indicator accurately reflect the outcome it is intended to measure?
- *Sensitivity*: Is the indicator sensitive to change in the outcome measured while being relatively insensitive to other changes?
- *Reliability*: Is the same result obtained when the indicator is re-measured?
- *Verifiability*: Can the indicator's value be corroborated through independent measurement?
- *Precision*: If the indicator is estimated using survey data and is re-estimated with a different set of data, is the expected level of variation acceptable?

The most common pitfall in selecting indicators is to select too many. When this happens, project implementers are likely to spend too much time collecting and processing data and not enough interpreting and acting on the information provided. Another common pitfall is to include indicators for which data are not available or are impractical to collect. Indicators should, however, be selected because they are most relevant to the results being monitored, and not simply because they are easy to measure. It is better to have approximate information about important issues than to have exact information about something trivial. In cases where it may not be feasible or too costly to obtain a direct measure of the desired indicator, a "proxy indicator" can be used in its place—for example, an asset index as a proxy indicator for household income.

DATA SOURCES, METHODS, AND VALUES OF INDICATORS

Data sources for indicators can be either primary (collected by the project) or secondary (collected by another organization for other purposes). Examples of primary data are project records, special surveys, and direct observation. Although using secondary data for some indicators may save money, such data are seldom available for the required time period and location or in exactly the form needed. Data collection methods range from informal and less structured to formal and more structured. Examples of the former include conversations with informed individuals (such as community leaders) and field visits. Examples of the latter include project administrative records, direct observations, and questionnaires. In

general, less formal methods yield less costly and more timely data, but usually at the cost of providing less accurate and less objective information.

All indicators should have *baseline* and target values and a *timeframe*. It is very important to have reliable baseline estimates because the baseline values provide the reference point for later monitoring. Baseline data should always be collected prior to implementation. Targets should be feasible and realistic, based on a careful assessment of the factors that will affect project implementation, including expected funding and other resources. The targets should be acceptable to all stakeholders and should rarely be adjusted during implementation (and only for good reasons). Indicators to monitor implementation (as distinct from results) may also require interim milestones to measure progress. Implementing organizations should avoid "gaming the system" in setting targets—such as setting them so low that they are easily achieved or adjusting them frequently to match actual implementation progress.

The guidelines include recommended indicators for each WEE final or intermediate outcome as well as links to questionnaire modules to collect the data to measure survey-based indicators.

TRADITIONAL M&E

The main purpose of traditional M&E is to improve project implementation and to report meaningfully on implementation progress to stakeholders. It includes both *routine monitoring* and *process evaluation* and is focused on the basic question: Are we doing things right? The routine monitoring of process indicators, called *trend monitoring*, is an indispensable management tool for making informed decisions about project activities and for assessing staff performance. The observed trends in indicators can also provide early warning signals to project managers when things go wrong. If the reasons are not immediately obvious, a process evaluation can be conducted to provide more detailed information about why an intervention is not achieving its planned outputs and direct outcomes as well as possible remedies. Data used for routine monitoring should be collected, processed, and analyzed frequently (monthly or at least quarterly).

In addition to providing timely feedback to project managers, routine monitoring can be an effective tool to inform and motivate project staff toward achieving results and to maintain accountability to stakeholders. Routine monitoring is also a very important input to an effective impact evaluation. The information it provides may enable project managers to take appropriate corrective actions during implementation to enable an intervention that is off-track to achieve its targeted results. In other cases, the monitoring data may indicate that the intervention is having unexpected results (such as unexpected effects or effects on unexpected groups) that need to be reflected in the impact evaluation. This may necessitate collecting additional information in a follow-up survey.

Routine monitoring may also indicate that some assumptions in the theory of change are not validated in practice—for example, that all who are offered an

intervention will take it up. In such cases, the analytical methods for an impact evaluation may need to be modified. In extreme cases, the monitoring data may indicate that an intervention failed and must be repeated. If an impact evaluation was planned, this information could avoid wasting evaluation resources on the impact evaluation of a failed intervention.

The most common pitfall in routine monitoring is to fail to conduct it during the early stages of an intervention, when the information provided would be most useful to determine whether the intervention is on-track or off-track and, if the latter, that the intervention needs to be adjusted. Another common pitfall is to collect too much information without much understanding of how to use it effectively. All too frequently, routine monitoring is not implemented because implementers and stakeholders do not understand its potential value.

Process evaluations are an important M&E activity complementary to routine monitoring. They are usually conducted during the early stages of a project to provide a deeper understanding of why planned outputs and direct outcomes are not being achieved or when there are specific operational concerns. The main sources of information for process evaluations are project records and reports, direct observations by the evaluators, and interviews with project managers, project staff, and beneficiaries. The analysis in connection with a process evaluation is usually more detailed than is possible with trend monitoring. A process evaluation may require going back to the original project records to obtain more detailed information, and it may in some cases require collecting additional data such as participatory appraisals, mini-surveys, focus groups, key informant interviews, beneficiary/nonbeneficiary interviews, exit interviews, community interviews, direct or participatory observations, expert opinions, case studies, and literature reviews. Process evaluations are usually conducted by project staff, often with the assistance of local consultants.

IMPACT EVALUATION

Impact evaluations assess the causal effects of an intervention, including any unintended effects, positive or negative, either on the targeted beneficiaries or other groups. They should also assess the intervention's costs in relation to its effects (its cost-effectiveness), and in some cases, its costs in relation to the estimated monetary value of its effects (cost-benefit analysis). Impact evaluations usually focus on intermediate and final outcomes because, unlike direct outcomes, they are most often affected by factors external to the intervention.

The central question of an impact evaluation is not only what happened after the intervention was implemented (the focus of results monitoring) but what would have occurred *in the absence of the intervention—the counterfactual*. Constructing a credible estimate of the counterfactual is the central challenge in impact evaluation. Although many different methodologies can be used to estimate the counterfactual, the most credible estimates usually come from a carefully designed and effectively implemented *randomized controlled trial* (RCT).

RANDOMIZED CONTROLLED TRIALS

Because RCTs (and credible impact evaluations more generally) are expensive, the guidelines do not recommend that every intervention should conduct an impact evaluation. The guidelines suggest that impact evaluations will be most useful for existing interventions that are candidates for scaling up or for a new intervention considered particularly promising in its potential to be a least-cost solution to an important problem. Impact evaluations are least useful when the effects of an intervention have already been established through several credible impact evaluations in similar settings (such as the effects of an effective immunization program on child survival). Because the findings of a credible impact evaluation can add substantially to the global knowledge base of which interventions are most effective in which settings, it is often possible to find supplementary sources of funding to support an impact evaluation.

There are many opportunities for conducting RCTs, including situations in which access to all or part of an intervention, or the timing of access, can be randomized. One of the most common opportunities occurs when an intervention must be phased in over time, due to logistical or resource constraints. In this case, random selection of those to receive the intervention initially is a practical and ethical choice. The outcomes for those who will receive the intervention in a second phase provide a credible counterfactual as long as there is enough time between the two phases for effects to materialize.

In some cases, RCTs are impractical. One example is when an intervention affects outcomes that are particularly sensitive and therefore difficult to measure precisely (such as an intervention to reduce domestic violence). Another is when an intervention must be implemented uniformly at a national level (such as changing macroeconomic policy). Some large infrastructure investments (a bridge or port) may also be impractical to evaluate with an RCT.

RCTs can also be unethical in some circumstances. A basic practice is that access to an intervention should never be denied solely for the purpose of an evaluation. But strict adherence to this practice can still be consistent with RCTs. When positive results come from an RCT, it may lead to additional funding to make the intervention more widely accessible. When negative results come from an RCT, it may avoid wasting scarce resources on an ineffective intervention, and the knowledge gained may lead to the development of a more effective intervention.

INTERACTION BETWEEN AN RCT AND THE INTERVENTION

An RCT may complicate and increase the cost of the intervention it is designed to evaluate—for example, by requiring the intervention to cover a wider geographical area or by requiring data to be collected as well in a control group. Accordingly, the guidelines focus on where an RCT is most likely to interact with the intervention it is evaluating. Conducting an RCT can be divided into three time periods: the pre-implementation or design period, the implementation period, and the post-implementation period.

The *pre-implementation* period, arguably the most important for an RCT, may require as long as one year to complete. During this period, the research team and the implementing team need to work together closely to design both the intervention and the RCT. Key design decisions concern: the level of randomization, the location of the intervention, the way the intervention will be implemented, the monitoring system to support both the intervention and the RCT, and the types of additional data that will be collected to support the RCT.

Baseline data collection and *randomization* (random assignment of the experimental units to treatment and control groups) must also be completed during the pre-implementation period. It is critical that implementation (including meetings of the implementing team with local political leaders or potential beneficiaries to discuss the intervention) not begin until after the baseline data have been collected and all experimental units have been randomly assigned to treatment or control groups.

The *implementation* period is also critical to the success of an RCT. Its length will vary, depending on the nature of the intervention. The main RCT activities during the implementation period that interact with the intervention are site visits by the research team, analysis of the routine monitoring data, collection of cost data, and process evaluations. Careful analysis of the monitoring data by the research team is particularly important. In its absence, the research team cannot determine whether the absence of impact, if encountered, is due to flaws in the intervention or to flawed implementation. Careful analysis of the monitoring data may also indicate problems of noncompliance, spillovers, attrition, or evaluation-driven effects that could complicate the impact evaluation.

The *post-implementation* period is also important, though it involves less interaction with the intervention than in the other periods. It will typically take six months to one year. The main RCT activities that interact with the intervention during the post-implementation period are collecting follow-up data, analyzing cost data, and disseminating the RCT's findings. Follow-up data, both qualitative and quantitative, should not be collected until the effects of the intervention have had time to materialize. If qualitative data are collected, it may be useful to collect them prior to fielding a follow-up survey—for two reasons. First, they may indicate whether the expected effects have materialized. Second, they may suggest questions that should be added to the questionnaire—for example, to provide additional evidence of any unexpected effects.

1. INTRODUCTION AND OVERVIEW

Most WEE programs provide women with skills and resources to increase their productivity and income. The long-term goal is to help women fulfill their economic potential and improve their well-being and that of their families and communities. When women are economically empowered, communities and nations benefit. Yet, there has been a crucial gap in knowledge about the most effective interventions that directly advance women's economic opportunities and how to measure them effectively. In 2012, the United Nations Foundation and the ExxonMobil Foundation launched a project to address this gap by identifying which development interventions most improve women's productivity and earnings. A group of 35 researchers worked on 17 different review and empirical studies that investigated practical, implementable projects aimed at women's economic advancement. The findings from that research were compiled in a report and Roadmap for Action that outlines what works best to improve women's productivity and earnings, for whom, and where.[1]

The role of monitoring and evaluation (M&E) in WEE programs is to support effective implementation of the projects ("Are we doing things right?"), determine whether the desired results are being achieved ("Are we doing the right things?"), and contribute to the global knowledge base on the types of interventions that are most effective in promoting the well-being of women, their families, and their communities ("Do we know what works best?"). In 2014, five researchers from the Roadmap project and several staff of the two foundations joined in a follow-on project to identify outcome measures for WEE programs. The objective was to identify a common set of widely applicable measures for two categories of women: urban women entrepreneurs and business leaders, and rural women entrepreneurs and farmers.[2] In July 2014, the researchers reached consensus on a set of outcomes for WEE programs (box 1.1) and on the six principles to guide their M&E activities:[3]

1. Given the interdependence of women's economic and social roles, it is important to measure both economic and social (well-being) outcomes to understand women's economic empowerment.
2. It is also important to measure both individual and household effects, considering the broader context of women's well-being in the household.
3. The what and the how of an evaluation matter equally: "what" refers to the outcomes measured, "how" to the evaluation design.
4. No evaluation is better than a poorly designed evaluation.
5. Not every program can be rigorously evaluated, but we can learn something useful from every program.
6. Complementary qualitative work is important to understand the "why" behind results, which can be useful for program staff.

1 United Nations Foundation and ExxonMobil Foundation. *A Roadmap for Promoting Women's Economic Empowerment* (2013) (www.womeneconroadmap.org).
2 While women entrepreneurs and farmers are the principal beneficiaries of WEE programs, the principles and concepts outlined in these guidelines are also broadly applicable to interventions benefiting women wage workers.
3 United Nations Foundation and ExxonMobil Foundation, "Measuring Women's Economic Empowerment," Washington DC (June 2015).

Box 1.1: WEE outcomes

Final outcomes: income/expenditures; assets/savings (and control over them); hours worked for pay; stress/life satisfaction; gender roles and norms; and self-esteem/ self-confidence. Measurements of these outcomes demonstrate that program goals are being achieved. Not every program can measure all of these outcomes.

Intermediate outcomes: management practices, technology adoption and effective use, and engagement/participation in community activities (or business networks). Since final outcomes often depend on factors beyond a woman's control, the intermediate and direct outcomes selected should be within her control.

Direct outcomes: Program take-up and retention; program-specific outcomes (e.g., learning new technology, acquiring new information, developing skills). These can be tailored to individual interventions and are more easily measured in the near term to ensure that the intervention was taken up as planned.

Source: UN Foundation and ExxonMobil Foundation, "Measuring Women's Economic Empowerment," Washington DC (June 2015).

These guidelines describe the role of monitoring and evaluation (M&E) in WEE programs.[4] The primary audience is implementers and funders of women's economic empowerment (WEE) programs. The purpose of the guidelines is to promote a harmonized approach to M&E within WEE programs. But the guidelines do not recommend that all programs use the same M&E methods. Instead, the guidelines provide a menu of M&E methods that different program implementers and funders can use to meet their individual M&E needs, which can vary due to such factors as the size of the program and the suitability of the supported activities to the various M&E methods. In some cases, M&E may be limited to helping a program implementer and funder learn as much as possible from the experience in implementing one or more activities. In other cases, particularly when there is interest in scaling up an existing activity or in piloting a promising new intervention, a more systematic impact evaluation may be called for. This flexibility in adapting M&E to the needs of individual programs, while simultaneously serving the interests of the program, is a key feature of the guidelines. (Principle 5: "Not every program can be rigorously evaluated, but we can learn something useful from every program.") These guidelines provide an overview of M&E methods believed to be most suitable for use in WEE programs.

The guidelines are intended to be concise and easy to follow, yet comprehensive and informative. Although technical terms are used frequently in the guidelines, they are defined as clearly and simply as possible when first introduced (or are displayed in italics to indicate that they are defined in the glossary). The references at the end of the main text provide deeper discussions of the topics in the guidelines.

4 The guidelines have drawn on many sources, including those listed by topic in Key Sources in the References section.

Monitoring and evaluation (M&E) covers a wide range of activities. The activities included in programs' M&E tend to vary with the characteristics of those programs, including their size, scope, and complexity, their expected outcomes, and the nature of the organizations (for example, NGOs or national governments). The M&E activities discussed here are believed to be most suitable for WEE program implementers. Similarly, the M&E terminology in these guidelines is adapted to the special circumstances of WEE programs. Table 1.1 shows the main categories of M&E, the M&E activities included in each type, the *results* focused on, the most common data sources, and the organization responsible for each category of M&E.

TABLE 1.1. MAIN CATEGORIES OF M&E

CATEGORY	PURPOSE	M&E ACTIVITIES	RESULTS	DATA SOURCES	RESPONSIBILITY
1. Traditional M&E	Improve the implementation of ongoing projects	Routine monitoring	Inputs, activities, outputs and direct outcomes	Routinely collected project data	Project
		Process evaluation		Project data, qualitative data, mini-surveys	Project and consultants
2. Impact evaluation	Learn whether an intervention has a causal effect on the targeted outcomes	Evaluation design	Strategy for estimating the counterfactual	Secondary data, qualitative data	Program (almost always with third-party researchers
		Data collection	Outcomes	Surveys, project data, qualitative data	
		Analysis	Estimates of causal effects, cost-effectiveness, and benefits in relation to costs		
3. Performance evaluation	Strengthen program effectiveness through lessons from past projects	Process evaluation	Inputs, activities, outputs, and direct outcomes	Project data, special studies	Program (usually with third-party evaluators)
		Performance monitoring	Intermediate and final outcomes	Surveys	

These guidelines focus on the first and second categories of M&E in table 1.1. Traditional M&E activities are most often conducted by program implementers to assess how effectively their interventions are being implemented, whether project benefits are being received by the intended groups and at what costs, and to identify and correct any problems discovered as quickly as possible. Program-level impact evaluations estimate the causal effects of an intervention on key outcomes as well as to assess whether the intervention is obtaining value per dollar spent

Activities in the first category are referred to in the remaining sections of these guidelines simply as "M&E," and activities in the second, as "impact evaluation."[5]

The guidelines do not focus on performance evaluation (the third category of M&E in table 1.1) because it is usually conducted at the program level after projects are completed and thus does not directly affect program participants.[6] The main purpose of performance evaluation is to improve the design of new projects. Impact evaluation, in contrast, affects all stages of an intervention's implementation, including its design. The discussion of impact evaluation in these guidelines focuses on the ways that it interacts with the design and implementation of an intervention rather than on the "how to" of conducting a credible impact evaluation.

5 The term "impact evaluation" refers in these guidelines to a study designed to attribute observed changes in one or more outcomes (including intermediate outcomes) to an intervention. The term sometimes also refers to studies that focus on the measurement of change(s) in final outcomes ("impacts") without attempting to attribute the observed change(s) to an intervention. See White (2010).

6 For a detailed discussion of performance evaluation and its role in strengthening programs, see USAID, *Evaluation: Learning from Experience* (January 2011).

2. CAUSAL CHAIN AND INDICATORS

A causal chain (sometimes referred to as a "results chain") links the activities of an intervention to its results, which include both outputs and outcomes.[7] The causal chain for a typical WEE intervention includes both an implementation segment and a results segment (figure 2.1). In the implementation segment, inputs are transformed by activities into outputs that lead to results. The results segment includes direct outcomes, intermediate outcomes, and final outcomes. Direct outcomes are most closely linked to the intervention's outputs, are frequently measured by the project, and are used to monitor how well the intervention is being implemented and the extent to which its outputs are being used by its intended beneficiaries. Intermediate outcomes occur prior to final outcomes and are believed to be causally related to final outcomes. Because final outcomes may take some time to materialize and are often more exposed to the influence of external factors beyond the control of the intervention or its beneficiaries (such as economic or climatic fluctuations affecting income) intermediate outcomes are important and are sometimes even the main focus of evaluations rather than final outcomes. Final outcomes (sometimes referred to as "impacts" in the evaluation literature) represent the intervention's ultimate objectives. If possible, they should be measured more than once because they may materialize only with a delay or be short-lived.

Traditional M&E usually focuses on the implementation segment of the causal chain, whereas impact evaluation focuses mainly on intermediate and final outcomes. But it usually gives some attention to outputs and direct outcomes to confirm that the intervention was implemented as designed. As one moves along the causal chain, the intervention has less control over the "results." Even direct outcomes can be affected by other factors, such as the education levels and motivation of the women participating in training courses.

THEORY OF CHANGE

The causal chain is based on a *theory of change* that describes how the intervention is to deliver the desired results and provides the rationale for the causal sequence from inputs to final outcomes.[8] The theory of change draws on an understanding of the context in which the intervention will be implemented and previous experience with the same or similar interventions in other settings. It is often developed in collaboration with stakeholders, both to draw on their expertise and to ensure that program implementation can proceed from a common understanding of the program, how it works, and its objectives.

The theory of change also makes explicit the assumptions and risks for each link in the causal chain. For example, one implicit assumption in figure 2.1 is that the inputs identified in the causal chain will be mobilized in the right place at the right time. If this does not happen the activities will be delayed or in extreme cases, fail to occur. Another implicit assumption is that trainers have the skills and

7 The term "results" usually refers to outcomes (or to outcomes and outputs), but it is sometimes used more broadly, as it is used here, to refer to all levels of the causal chain, including inputs and activities.
8 Gertler, Martinez, Premand, Rawlings, and Vermeersch 2011.

appropriate training materials and venues to provide effective training and that the trainees have the capacity and motivation to learn. These are important, but strong, assumptions that require validation in many settings. If any of them are incorrect, the direct outcome of "knowledge and skills acquired" may not be achieved.

If pre- and post-training scores are reported as a monitoring outcome, they can provide rapid feedback to those implementing the intervention on whether the training is producing the expected results. If the expected results are not being obtained, appropriate action can be taken to identify and correct the "weak link(s)" in the chain (box 2.1). In extreme cases, the monitoring data may indicate that the intervention was not implemented as planned and must be repeated. If an impact evaluation was planned, this information could avoid wasting resources on an impact evaluation of a failed intervention.

Box 2.1: Weak links in the causal chain of the Bangladesh Integrated Nutrition Project

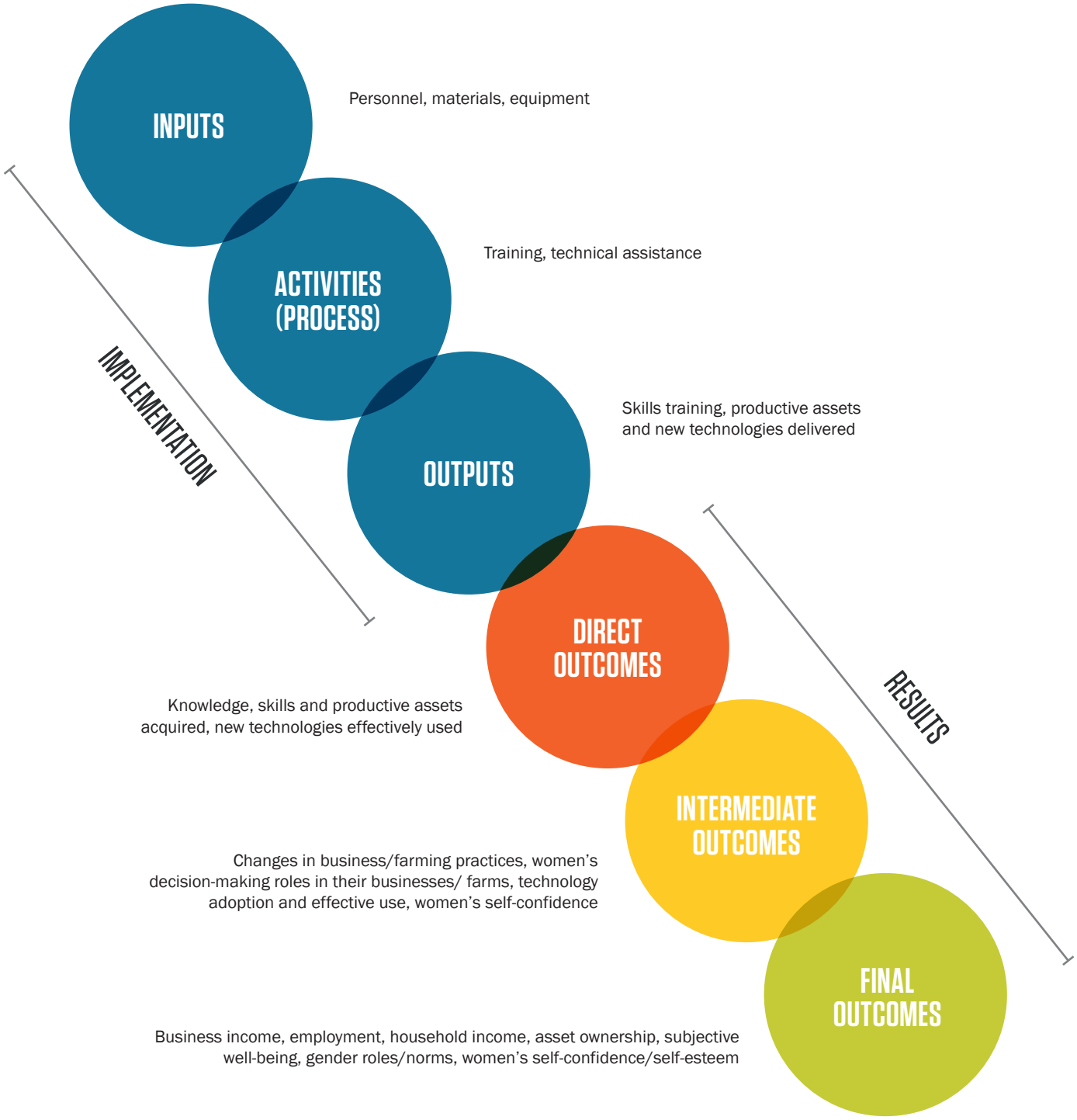
The Bangladesh Integrated Nutrition Project, a World Bank-supported pilot project modeled after a project in India, was widely perceived to have been successful. The project consisted of both nutritional counseling and supplementary feeding, with the main impact expected from the counseling. But the Bank's own evaluation found no evidence of project impact on nutritional status in the Bangladesh project. This surprising finding turned out to be due to four weak links in the theory of change underlying the project's causal chains.

- First, the project targeted the mothers of young children. In Bangladesh, however, mothers are rarely the sole decisionmakers for the health and nutrition of their children. It is men who do the shopping, while for women in joint households, the mother-in-law is the main decisionmaker about matters pertaining to women and children.
- Second, most community nutrition counselors, who were responsible for identifying malnourished children to include in the project, could not read growth charts, so project resources were in many cases directed to the wrong children.
- Third, there was considerable leakage in the supplemental food provided, whether to other household members or through food consumed in place of a meal that would have been eaten anyway.
- Fourth, although women who entered the project were better informed about nutrition, the gap between knowledge and practice was substantial, again due to the role of mothers-in-law.

Discovering these weak links early on would have saved a lot of money and could possibly have led to a more effective project.

Source: White 2008.

FIGURE 2.1. CAUSAL CHAIN FOR A THEORY OF CHANGE FROM INPUTS TO FINAL OUTCOMES



ROLE OF INDICATORS

Indicators measure the extent to which a given result occurs during (or after) implementation. The results expected from an intervention are usually expressed in general terms. Examples are "knowledge and skills acquired" (direct outcome), "improved business practices" (intermediate outcome), and "enhanced well-being" (final outcome). An "indicator" is a measure of an outcome or of a lower-level result (inputs, activities, outputs). Its definition includes the scale for measuring (such as number, percentage, or ratio). Indicators should be neutral about the desired direction of change. Thus "increased income" is not an indicator (it is an outcome), whereas "household consumption per capita" would be a possible indicator to measure "increased income." Similarly, indicators should not refer to milestones, targets, benchmarks or baseline values, which should be listed separately. If a given result is too complex to be represented by a single indicator, it is better to use two indicators than to combine what are essentially two distinct dimensions of a result into a single indicator. The indicators to measure inputs, activities, and outputs are usually based on data collected by the project. Indicators of direct outcomes may also be based on data collected by the project, such as assessments of the competency achieved by trainees. But intermediate outcomes and final outcomes are usually based on population data collected less frequently by a third party to ensure both their quality and credibility.

Indicators can be quantitative or qualitative. Quantitative indicators refer to measurable quantities (such as number of jobs created, age, sex, years of schooling completed, housing characteristics). Qualitative indicators are also observable but are often based on self-reported subjective judgments expressed as categorical values that may or may not indicate an ordered sequence. An example is a trainee's satisfaction with a training course, which might be coded initially as "highly satisfied," "fairly satisfied," "partially satisfied, or "not at all satisfied," but ultimately combined to obtain a simpler indicator, such as the percentage of trainees that are either "highly or fairly satisfied." One important disadvantage of qualitative indicators is that it is difficult to verify them independently because they tend to vary depending on who is doing the measurement. Another is that the values may not be comparable across different groups or over time. So, although it is sometimes desirable to include qualitative indicators in a monitoring framework, they should be used sparingly.

SELECTING INDICATORS

Selecting appropriate indicators is a critical aspect of M&E. The first step is to state clearly the results desired at each step in the causal chain. Overly broad or general statements of results should be avoided. Instead, the results should focus on the specific areas where improvements are needed. The result statement should be clear about the type of change desired. For example, is it the level of knowledge or actual behavior? Is it an absolute change, a relative change, or no change (maintaining an existing level)? The result statement should also indicate the specific groups targeted by the intervention. For example, is it all people in a particular geographic area, or a specific age-gender or socioeconomic group?

An initial list of indicators for each result can be developed as a brainstorming exercise that may include consultations with experts in the substantive program area and/or the participation of stakeholders. A set of basic criteria, frequently used to select among alternative indicators, can then narrow the list (table 2.1).[9] In some cases, the selection process can be made more formal by constructing a table (matrix) with criteria in the first column and the alternative indicators for a given result listed in the first row. A numerical score (such as 1-5) for each criterion can then be assigned to each indicator reflecting the extent to which the indicator satisfies the criterion. If some criteria are considered more important than others, the scores can be weighted to reflect this before they are added up. The scores in each column should give an overall sense of each indicator's relative merit and may help in arriving at the final selection of indicators.

TABLE 2.1. DESIRABLE CHARACTERISTICS OF INDICATORS

CRITERIA	
Simple	Is the information conveyed by the indicator easily understandable and appealing to the target audience?
Clear	Is it clear what the indicator is attempting to measure? Does the indicator attempt to measure only one result?
Valid	Does the indicator accurately reflect the result it is intended to measure?
Sensitive	Is the indicator sensitive to change in the result while being relatively insensitive to other changes?
Reliable	Can data be collected using scientifically defensible methods that produce consistent estimates in repeated measurements
Practical	Are good-quality and timely data available and affordable to measure the indicator?
Precise	If the indicator is estimated with survey data and is re-estimated with a new set of data, is the expected level of variation in the value acceptable?
Verifiable	Can the indicator's value be corroborated through re-measurement by another evaluator?
Targeted	Is the group targeted by the result clearly reflected in the indicator (gender, age, socioeconomic status)?
Objective	Is the indicator objective (directly observable) or is it subjective (based on someone's opinions)
Adequate	Does the indicator adequately represent the result, or does it only reflect one aspect of the result?

The objective is to come up with a limited number of indicators representing the results at each link in the causal chain, keeping in mind that no indicators are perfect. A common pitfall is to select too many indicators so that it becomes impractical to report on them at regular intervals. In this case, too much time is likely to be spent managing the system that produces the data, with not enough

9 Other criteria have been proposed in a more structured way in the M&E literature, including CREAM— Clear (precise and unambiguous), Relevant (appropriate to the subject at hand), Economic (available at reasonable cost), Adequate (provide a sufficient basis to assess performance) and Monitorable (amenable to independent validation)—and SMART—Specific, Measurable, Achievable, Relevant, and Trackable.

time spent using the data to improve implementation. Large volumes of data can also be confusing to users. Particularly to be avoided are multiple indicators of essentially the same result. It is sometimes recommended that an index be constructed that is based on several related indicators and that values of the *index* be monitored instead of each indicator separately.[10]

Another common pitfall is selecting indicators for which data are not available or are impractical to collect. Indicators should, however, be selected because they are most relevant to the results being monitored, and not simply because they are easy to measure. It is better to have approximate information about important issues than to have exact information about something that is trivial. Consideration should also be given to the cost of improving the information currently available to measure one or more key indicators. In cases where it may not be feasible or too costly to obtain a direct measure of the desired result, a *proxy indicator* can be used. For example, it is very difficult to obtain an accurate estimate of household incomes in a typical rural setting. Accordingly, household consumption or an *asset index* is often used as a proxy indicator for household income.

It is important that indicators have a clear meaning. For example, the indicator "Percentage of trainees who achieved a passing score," is ambiguous because what constitutes a "passing score" may vary widely. In such cases, it is better to report the percentage of trainees who achieved a specific minimum score (such as 60%). It is also important not to use indicators that are difficult to interpret. An example is "the number of reported episodes of domestic violence." If the indicator increases, it is unclear whether this is due to increased domestic violence or to increased *reporting* of domestic violence. A less ambiguous indicator in this case might be "the percentage of reported episodes of domestic violence that the relevant authorities have dealt with effectively."

Due in part to the difficulty in arriving at a suitable list of indicators, some programs apply a set of predetermined indicators to all projects. There are both pros and cons for using predetermined indicators. They can reduce M&E costs at the program level and facilitate the process of comparing similar interventions or projects. But they may not be appropriate for all interventions, lessening "ownership" because it does not allow key stakeholder participation in the process of selecting indicators. One possible compromise is to use a uniform set of recommended indicators for final and intermediate outcomes, with program-specific indicators for lower-level results.

Because monitoring is a process that must be repeated at regular intervals in order to gain a clear picture of progress, it is desirable that both the list of indicators and their definitions remain fixed during the period of project implementation. It is equally important that the records or forms to collect the information to measure indicators not change during project implementation. Arriving at the final set of indicators may take some time and involve more than one try. Still, two important

10 Perhaps the best-known development-related index is the UN's Human Development Index (HDI), but there are many other examples.

questions should be considered before changing or dropping an indicator. Has the indicator been tested enough to know whether it is effective? And is it providing information that is useful to project managers and stakeholders?

3. WEE OUTCOMES AND RECOMMENDED INDICATORS

The outcomes in this section were selected to reflect both economic and social objectives, based on the distinct dimensions of the assumed WEE program goal, which includes increasing the income and productivity of women and enhancing their well-being and that of their families and communities. The intertwining of economic and social roles in women's lives means that barriers to either dimension can prevent progress on the other. Given this interdependence, "It is important to measure both economic and social (well-being) outcomes to understand women's economic empowerment" (Principle 1).

Women's family roles may influence business choices and the returns to those businesses. Microenterprises owned by women are often interlinked with other household activities, and decisions on sector, time spent, and growth objectives are made with both spheres in mind. Where interventions cause the business to scale up, they may also result in more stress in juggling household responsibilities, offsetting any potential gains in profits or revenue. Potential gains in profit or business growth may also be compromised by family or community pressures to share windfall cash, pressure women are exposed to more than men in poor households in developing countries.

In these cases, it should be left to the judgment of evaluators and researchers how well economic outcomes convey real increases in women's economic empowerment. For instance, increases in business productivity that maintain the level of business profits may be more easily interpreted as increasing economic empowerment than increases in business productivity with reduced business profits. Qualitative data about the types of women participating in the intervention, and their goals for running a business, will be useful to collect alongside profit and revenue information to address this measurement issue.

The fact that women are embedded in households makes it difficult to separate and measure the effects of programs in one domain when they spill over into another domain, as happens in households and families everywhere, but especially in rural households in developing countries. While this challenge is applicable in theory to all family members, it is particularly an issue for women because of the strong interdependence between their economic and family roles. It is therefore "important to measure effects at both the individual and household levels" (Principle #2).

RECOMMENDED INDICATORS FOR WEE FINAL OUTCOMES

Some WEE final outcomes differ between urban women entrepreneurs and business leaders—and rural women entrepreneurs and farmers. For urban women, the incomes of individuals from their self-employment activities are important final outcomes, along with their conditions of employment. For rural women, however, it is impractical in most cases to separate their individual income-earning activities from those of the overall household. Accordingly, household income is a more suitable final outcome for rural women. The other WEE final outcomes are equally applicable to both urban and rural women, including: individual assets, happiness/

stress, gender roles and norms (focused on decisionmaking in the household), self-confidence, and self-esteem—all measures of wellbeing. Although some of these outcomes are subjective and difficult to measure, they are an important dimension of women's economic empowerment.

Self-confidence refers to a woman's ability to interact freely—and if necessary with confidence and assertiveness—with individuals outside her household and immediate peer group (such as government officials, business owners). It also refers to her willingness to assume risk and her ability to manage her affairs on her own. Although listed here as a final outcome, self-confidence is also clearly an important intermediate outcome, particularly for urban and rural women entrepreneurs. Self-esteem refers to a woman's own assessment of how much she is able to contribute to, and is a valued member of, her family and community—and whether she thinks she should have equal rights with men.

Table 3.1 lists the WEE final outcomes with their recommended indicators. (appendix 1 provides examples of questionnaire modules to collect the data to measure the indicators.) Surveys are the data source for all final outcome indicators, which are population-based. The questionnaire modules for the indicators appropriate for M&E purposes (as indicated in the table) can be administered directly to the individual WEE beneficiaries (there is no need to conduct a household survey). But some of the final outcome indicators are intended for use only in an impact evaluation, and the corresponding questionnaire modules would need to be administered in a household survey.

Accurate measurement of urban women's income from self-employment is challenging for four reasons.

- First, measuring business income requires the respondent to recall figures on sales and costs and assumes she has this information readily at hand (Woodruff and McKenzie 2012). Particularly in microenterprises, where business income is often mixed with household income, women business owners may not have a good idea of their income in any given period.
- Second, baseline and endline measurements may not be comparable if the intervention itself changes the quality of responses through new knowledge and improved business management practices.
- Third, business income often fluctuates considerably in the short term, while collecting data on income over a longer period is likely to increase measurement error.
- Fourth, a common obstacle to accurate measurement of business income is a lack of trust between the interviewer and the respondent. Respondents are likely to underreport income to unfamiliar persons out of fear, for example, that the information provided will lead to higher taxes. If the trust between respondents and interviewers improves during the intervention, this may compromise the comparability of baseline and endline measurements.

TABLE 3.1. RECOMMENDED INDICATORS FOR WEE FINAL OUTCOMES

OUTCOME	INDICATOR	DATA SOURCE	UNIT OF OBSERVATION	USE
Urban women entrepreneurs and business leaders				
Business income	Woman's business profits	Survey	Individual	M&E, impact
	Woman's business revenue (sales)	Survey	Individual	M&E, impact
Employment	Number of employees in the woman's business	Survey	Individual	M&E, impact
	Average monthly hours worked for pay by woman	Survey	Individual	M&E, impact
	Average monthly income earned per hour worked for pay by woman	Survey	Individual	M&E, impact
Rural women entrepreneurs and farmers				
Household income	Total household consumption per capita	Survey	Household	Impact
	Household consumption per capita of selected items	Survey	Household	M&E
	Household asset index	Survey	Household	Impact
	Household savings	Survey	Household	Impact
Both urban and rural women				
Individual assets	Net value of woman's financial assets	Survey	Individual	Impact
	Value of woman's bank and financial accounts	Survey	Individual	M&E
	Value of woman's physical assets	Survey	Individual	Impact
	Value of woman's motor vehicle	Survey	Individual	M&E
	Value of woman's mobile phone	Survey	Individual	M&E
Satisfaction with life	Woman's overall satisfaction with life	Survey	Individual	M&E, impact
	Woman's stress level	Survey	Individual	Impact
Gender roles and norms	Woman's roles in household decisionmaking	Survey	Household	Impact
Self-confidence	Woman's overall self-confidence	Survey	Individual	Impact
	Woman's willingness to assert herself	Survey	Individual	M&E, impact
	Woman's willingness to take risk	Survey	Individual	Impact
Self-esteem	Woman's self-esteem	Survey	Individual	Impact

Note: See appendix 1 for recommended questionnaire modules to use in collecting data on final outcome indicators.

Under these conditions, it is important to use business revenue (sales) as well as profit as an indicator of urban women's income from self-employment. It is also important to collect data on intermediate outcomes directly linked to business income, such as improvements in business practices.

Urban women entrepreneurs' employment indicators include the number of employees in the woman's own business, the average monthly number of hours worked for pay, and the average monthly income earned per hour worked (productivity). The number of hours worked may vary substantially from one month to the next. So, the recommended questionnaire module collects data on both the number of hours worked during a typical month and the number of months worked per year. The interpretation of the resulting data can be complex. For example, in contexts where unemployment and underemployment are problems, an increase in the number of hours worked may be a positive outcome for women entrepreneurs. But in other contexts, the main problem may be low productivity (low income per hour worked), and an increase in the number of hours worked may signal a further fall in productivity.

Among rural women entrepreneurs and farmers, individual incomes are difficult to measure because so many activities are conducted at the household level. Because it is generally infeasible to measure household income directly in the rural areas of most developing countries, a common practice is to use household consumption or an asset index as a proxy indicator of household incomes (Deaton and Grosh 2000; Filmer and Scott 2012). Both measures do about as well over the medium to longer term. Many impact evaluations collect data on household consumption, even though it is time-consuming and requires careful training of interviewers. For M&E (as distinct from impact evaluations), it may be possible to collect data on only a few consumption items that are closely related to household incomes. Examples include the consumption of meat, fish, fruit, vegetables, and meals purchased away from home.

The data to construct a reliable asset index are much easier to collect. Asset indices are usually based on readily observable and verifiable indicators of housing characteristics (such materials used in walls, floors, and roofs) and of household ownership of consumer durables. Asset indices have been used as proxies for household income (or "socioeconomic status") in large scale surveys such as the Demographic and Health Surveys and UNICEF's Multiple Indicator Cluster Surveys. However, different indices may need to be used in urban and rural areas if the types of housing and household durables differ substantially between them. It is currently unclear how sensitive asset indices are to short-term fluctuations in household income (especially to downward fluctuations).[11] Examples of questionnaire modules to obtain the data to measure household consumption or to estimate an asset index are in appendix 1.

11 Work by the Gender, Agriculture, and Assets Project (GAAP) has shown that gendered asset ownership measures are responsive to program interventions with moderate duration of three to five years (Quisumbing and others 2014).

Household savings can also be an indicator of changes in household income in rural areas. Savings is also an important WEE outcome in its own right because increased savings can lead to increased household investment in productive assets, help smooth household consumption in the event of emergency expenditures, and increase women's empowerment if women have control over at least some of the increased savings. Although household savings is defined as the difference between household income and consumption, it cannot be measured reliably in this way (due to the infeasibility of measuring the two components reliably). But household savings can be estimated by comparing estimates of the value of a complete inventory of household physical and financial assets, including land, structures, livestock, consumer and producer durable goods, and financial assets and liabilities at two points in time. Measuring asset values at a point in time is much easier than measuring changes in asset values over a period (Kochar 2000). In rural households, most assets are physical (rather than financial), with the largest categories being land, housing, livestock, consumer and producer durables, and valuables such as gold and jewelry.

The indicators for individual assets include several referring to the net value of a woman's individually owned financial and physical assets. They are equally applicable to urban and rural women. As table 3.1 indicates, some of these indicators are most suitable for an impact evaluation, others for M&E.

The indicators of the remaining final outcomes are more subjective and based on women's responses to questions about their feelings and decisionmaking roles in the household. Although such indicators may not be as reliable as more objective measures of income or assets, objective measures of these important outcomes are not currently available. The questionnaire modules proposed in appendix 1 for collecting data on these indicators have been widely used in international survey programs.

RECOMMENDED INDICATORS FOR WEE INTERMEDIATE OUTCOMES

Some WEE intermediate outcomes also differ between urban women entrepreneurs and business leaders and rural women entrepreneurs and farmers. The intermediate outcomes for urban women entrepreneurs and business leaders include: business practices, value of business training, and gender roles and norms (focused on decisionmaking in the women's own businesses). These same indicators are also applicable for rural women entrepreneurs. The intermediate outcomes for rural women farmers focus on agricultural practices, value of access to new or improved agricultural technology, and gender roles and norms (focused on decisionmaking in women's or family farms). The remaining WEE intermediate outcomes apply equally to both urban and rural women, including: technology adoption and effective use (focused on mobile phones), women's self-confidence, gender roles and norms (focused on sharing of housework), and participation in community, business, or farmer's groups.

Table 3.2 lists the WEE intermediate outcomes with their recommended indicators. (appendix 1 provides examples of questionnaire modules to collect the data to

TABLE 3.2. RECOMMENDED INDICATORS FOR WEE INTERMEDIATE OUTCOMES

OUTCOME	INDICATOR	DATA SOURCE	UNIT OF OBSERVATION	USE
Urban and rural women entrepreneurs				
Business practices	Adoption of recommended business practices	Survey	Individual	M&E, impact
Value of business training	Willingness to pay for general business training	Survey	Individual	Impact
	Willingness to pay for specialized technical business training	Survey	Individual	Impact
Gender roles and norms	Women's decisionmaking role in own business	Survey	Individual	M&E, impact
Rural women farmers				
Agricultural practices	Woman's adoption of recommended agricultural practices	Survey	Individual	M&E, impact
Value of access to new/ improved agricultural technology	Woman's willingness to pay for access to new or improved agricultural technology	Survey	Individual	Impact
Gender roles and norms	Woman's decisionmaking role in her own or family farm	Survey	Individual	M&E, impact
Both urban and rural women				
Technology adoption and effective use	Woman's intensity of mobile phone use	Survey	Individual	M&E, impact
Women's self-confidence	Woman's overall self-confidence	Survey	Surveys	Impact
	Woman's willingness to assert herself	Survey	Individual	M&E, impact
	Woman's willingness to take risk	Survey	Individual	Impact
Gender roles and norms	Sharing of housework between spouses/partners	Survey	Household	M&E, impact
Participation in community, business or farmer groups	Woman's participation in all types of groups	Survey	Individual	M&E, impact
	Woman's participation in mainly women's groups	Survey	Individual	M&E, impact

Note: See appendix 1 for recommended questionnaire modules to use in collecting data for intermediate outcome indicators.

measure the WEE intermediate outcome indicators.) Surveys are also the data source for all intermediate outcome indicators, which are population-based. The questionnaire modules for the indicators appropriate for M&E purposes (as indicated in the table) can be administered directly to the individual WEE

beneficiaries (there is no need to conduct a household survey). But some of the intermediate outcome indicators are intended for use only in an impact evaluation, and the corresponding questionnaire modules would need to be administered in a household survey.

Improved business practices is a key intermediate outcome for both urban and rural women entrepreneurs and urban business leaders. (Appendix 1 presents two alternative questionnaire modules for collecting the data to calculate the indicator on the adoption of recommended business practices.) The first module is relatively brief and suitable for M&E, the second is longer and more suitable for an impact evaluation. Both modules focus on the two practices most often emphasized in training courses: the need to keep business and household finances separated, and the importance of keeping good business records (Woodruff and McKenzie 2012). Other important business practices in both modules include: preparing or updating a business plan, applying for a loan, and marketing efforts (such as advertising, visiting customers).

One limitation of recommended business practices as an intermediate indicator is that it is based on self-reporting, and there may be a tendency of business owners to report that they adhere to business practices they were told to follow during training. There is also a risk that changes in business practices may have no effect on such final outcomes as business revenues or profitability. In Sri Lanka, for example, business management training improved the management practices of women business owners, but the changes in management practices induced by training did not increase profits and sales (de Mel and others 2014).

Other intermediate outcomes for urban and rural women entrepreneurs and urban business leaders are the value of business training and gender roles and norms (focused on women's decisionmaking role in their own business). The value of business training is assessed through questions to elicit women's willingness to pay for both general and technical business training. Women's decisionmaking roles in their own business are assessed using questions about their individual roles in key business decisions and activities.

The adoption and effective use of new or improved agricultural technologies is a key intermediate outcome for rural women farmers. Any technology or new input that can save women's time is particularly important (Doss, Bockius-Suwyn, and D'Souza 2012). Among farmers, the take-up and effective use of technologies such as improved seeds, fertilizer, and improved livestock breeds is often critical, along with the selective use of pesticides and herbicides. The take-up and effective use of improved harvesting, storage, and processing technologies can also contribute to farm incomes. Other important intermediate outcome indicators (for which questionnaire modules are provided in appendix 1) are: value of access to new or improved agricultural technology (assessed through willingness to pay), and gender roles and norms focused on women's decisionmaking role in their own or family farms. Increasing women's role as decisionmakers in their own or family farms

will make it possible for them to use new resources, knowledge, and skills more effectively.

Intermediate outcomes appropriate for both urban women entrepreneurs and business leaders and rural women entrepreneurs and farmers are women's adoption and effective use (focused on mobile phones), women's self-confidence (also a final outcome), gender roles and norms (focused on sharing housework between spouses/partners), and women's participation in community, business, and farmer groups. Mobile phone use can increase women's access to banking services, facilitate payments and money transfers, and improve access to market price information. Increased self-confidence can make it easier for women to interact effectively with customers, middlemen, agricultural extension workers, loan officers, and government officials in ways that can help their businesses and farms. It can also make them more effective participants in community activities, business associations, and agricultural cooperatives. Increased sharing of housework between spouses/partners can provide women more time to use their new resources, knowledge, and skills in their businesses and farms. Increased engagement and participation by women in community activities, business associations, and farmer groups and cooperatives can further boost their self-confidence and self-esteem and give them access to new opportunities and larger networks. Participation in cooperatives can increase prices for farmers by bypassing middlemen.

POSSIBLE INDICATORS FOR WEE DIRECT OUTCOMES

Direct outcomes, more easily measured in the near term, assess whether interventions are being implemented and taken up as planned. These should be tailored to individual programs. The WEE direct outcomes are likely to include intervention take-up and retention, and intervention-specific outcomes such as acquiring new productive assets, learning new technologies, acquiring new information, and developing skills. Table 3.3 lists some typical WEE program direct outcomes with examples of possible indicators for urban women entrepreneurs and business leaders as well as rural women entrepreneurs and farmers. The data for these indicators should be available from project records, but it is useful to triangulate the values of the indicators through periodic surveys of beneficiaries.

DATA SOURCES FOR INDICATORS

Data sources for indicators can be either primary (collected by the project) or secondary (collected by another organization for other purposes). Examples of primary data are project administrative data, special surveys, interviews, and direct observation. Examples of commonly used secondary sources are government administrative data or national multipurpose surveys. Although using secondary data for some purposes may save money, it is seldom available for the required time period or in exactly the form needed. The description of data source(s) should include information on the data collection method, unit of analysis (individual, household, facility, community), and the sampling procedures for each indicator.

TABLE 3.3. EXAMPLES OF POSSIBLE INDICATORS FOR WEE DIRECT OUTCOMES

OUTCOME	INDICATOR	DATA SOURCE	UNIT OF OBSERVATION	USE
Intervention take-up and retention	Percent of women offered an intervention who take it up	Project records	Individual	M&E, impact
	Percent of initial participants who continue to participate after one year	Project records	Individual	M&E, impact
	Percent of women participating in savings schemes	Project records	Individual	M&E, impact
Acquiring productive assets	Number of women receiving seeds, fertilizer, livestock, equipment or other productive assets	Project records	Individual	M&E, impact
	Number of business owners/farmers receiving loans or grants	Project records	Individual	M&E, impact
Learning new technology	Percent of women farmers participating in extension trainings	Project records	Individual	M&E, impact
Acquiring new information	Percent of business training events attended	Project records	Individual	M&E, impact
	Scores on pre- and post-training tests	Project records	Individual	M&E, impact
Developing skills	Percent of trainees certified as competent in skills taught	Project records	Individual	M&E, impact

Note: Project records are supplemented with periodic surveys of beneficiaries.

Address the following questions when selecting an appropriate data source for each indicator:

- Who will collect the data (project staff, the government, a third party)?
- How often will the data be collected (weekly, monthly, quarterly, yearly)?
- What is the cost of collecting the data, and are any special difficulties involved?
- Who will analyze the data (project staff, a third party)?
- Who will report the data (the project, a third party)?
- Who will use the data (project managers, project staff, other stakeholders, funders)?

Data collection methods range from informal and less structured methods to formal and more structured methods. Examples of the former include community meetings, conversations with concerned individuals, field visits, reviews of project records, key informant interviews, participant observation, mapping, focus group

interviews, and expert judgment. Examples of the latter include direct observation, questionnaires, diaries, one-time surveys, *panel surveys*, censuses, and citizen report cards. No single data collection method is best for all indicators. The choice should depend on what is to be measured, relative costs, and the importance of having timely and accurate data. In general, less formal methods yield less costly and more timely data, but usually at the cost of providing less objective and less *reliable* information.

Data to monitor implementation should be collected, processed, and analyzed frequently (such as monthly or at least quarterly). Doing so enables managers to use the data to discern real trends in time to make any necessary adjustments during implementation. Data to monitor higher-level results can be collected less frequently. More credibility is usually associated with survey data collected, analyzed, and reported by a third party. But this is relatively expensive and is more often done in the case of higher-order results, such as intermediate and final outcomes. When collecting survey data, it is important to:

- *Keep it numerical*: Quantitative measures with comparable units allow interpersonal comparisons that are key to any program evaluation. For instance, "How many hours do you work per week?" yields comparable answers, whereas "Do you work full-time or part-time" does not, as the definition of "full-time" or "part-time" might differ for different people.
- *Keep it easy*: Most respondents have limited education, creating challenges for the measurement of probabilities or percentages. Visual and manual instruments can help reduce noise. Two examples that are known to work are "smiley scales" to measure satisfaction and "a bag of beans" to elicit responses on the allocation of scarce resources, for example, time devoted to different activities, or earnings by month.[12]
- *Keep it short*: As fatigue sets in noise increases. Order survey modules so that fun and easy sections alternate with long and tedious sections. Piloting demonstrates which sections respondents like to answer.
- *Keep it consistent*: Use the same unit of measure whenever possible (in a time use survey always ask about hours or minutes), refer to the same time interval when asking recall questions, and make sure multiple-choice options are mutually exclusive and the list is complete.
- *Give a way out*: Always include "don't know" and "refuse to answer" options.
- *Ask sensitive questions toward the end of the interview*: This way, if the respondent refuses to continue the interview, most of the information will have already been collected.
- *Carefully field test questionnaires and train interviewers well*: In particular, culture and language need to be carefully considered when collecting survey data.

12 Respondents are given a set of cards representing different activities (in a time use survey these would be "work," "taking care of children," "cooking," and so on) and a bag of beans that they are asked to allocate to the different cards in proportion to the time they devote to each activity.

Questions obtaining data to measure outcomes should be based to the extent possible on international standard questions. The questionnaires used in established international survey programs can provide guidance on the best way to obtain data, for example, the Demographic and Health Surveys Program (www.dhsprogram.com), UNICEF's Multiple Indicator Clusters Surveys (www.unicef.org/statistics/index_24302.html), and the World Bank's Living Standards Measurement Surveys (Grosh and Glewwe 2000). But these international model questions should be adapted as needed to be consistent with similar questions in national surveys.

BASELINE AND TARGET VALUES OF INDICATORS

Indicators usually require *baseline* and *target* values and a *timeframe*. Process indicators may also need interim *milestone*s to measure progress during implementation.

It is very important to have reliable estimates of the baseline values of all indicators because the baselines provide the starting point for monitoring. In particular, the baseline values should be obtained from the same sources, and using the same methods, as those used during implementation. This may require replacing an original set of baseline values cited in a project design or proposal by more accurate estimates obtained just prior to implementation. Because baseline values refer to the period prior to project implementation, it is important to set up the monitoring system in advance of project implementation. If data are available for several periods prior to project startup, and if there is considerable variation in the values of some indicators from one period to the next, it may be desirable to fix the baseline value as the average value over several recent periods.

The procedures for collecting baseline data for M&E are different from those for collecting baseline data for an impact evaluation. Baseline data for M&E focus on the implementation segment of the causal chain; those for an impact evaluation, on the higher-level outcomes from the results segment. But this distinction is not hard and fast. Baseline data for M&E may include selected indicators of higher-level results (such as business practices, the adoption of new or improved technologies, or women's ownership of financial or physical assets). An impact evaluation will often collect some baseline data on direct outcomes. One hard and fast rule, however, is that when baseline data on results indicators are collected for M&E, the data are collected only from beneficiaries. There is no need to collect such data from nonbeneficiaries for comparison. That kind of information is useful only in connection with an impact evaluation.

Targets should be feasible and realistic, based on a careful assessment of the factors that will affect project implementation, including expected funding and resource levels. Past trends in indicators, when available, may help in setting realistic targets, even if they refer to a similar intervention in a different geographical area. The experience of other projects with the same or a similar intervention should also be considered when setting targets. Relevant international or regional *standard*s may also help in setting targets if a project objective is to bring performance up to those standards. The targets selected

should be acceptable to all stakeholders (in particular, there must be buy-in by project managers) and should not be adjusted during implementation (although some flexibility may be indicated if fewer resources become available or if other unanticipated obstacles arise). Targets are most often set annually. But shorter term (such as quarterly) interim targets (milestones) may be more suitable for inputs, processes, and outputs—with longer term targets (3–4 years into the future) reserved for final outcomes. Setting targets beyond 3–4 years is generally not advisable because of the future's inherent uncertainty.

Special care should be taken in setting targets for new indicators. Temporary targets may be used until more reliable information becomes available. Alternatively, the target of a new indicator may be a range of values—for example, the target for the "percentage of trainees finding a job within three months" may be set initially at 30-40%. Indeed, not all indicators need targets. It may be beneficial to monitor some indicators initially in the absence of targets, deferring target-setting to a later stage of M&E development when more information is available.

Implementing organizations should avoid "gaming the system" in setting targets, for example setting targets so low that they can be easily achieved or proposing changes in the targets during implementation to match actual implementation progress in the absence of unanticipated changes in funding or other unanticipated obstacles.

4. TRADITIONAL M&E

The main purpose of traditional M&E—referred to in these guidelines simply as "M&E"—is to improve project implementation and to provide information on implementation progress to stakeholders. M&E is essentially focused on the basic question: Are we doing things right? Close and continuous monitoring is essential to detect shortcomings in an intervention's design and implementation. Is the project mobilizing the needed inputs and has it undertaken and completed the planned activities? Have the intended outputs been delivered according to schedule, and are they of acceptable quality? Are the processes used to transform inputs into outputs efficient? Are the expected direct outcomes resulting from the production of outputs? Is take-up sufficiently high by the targeted beneficiaries? It is also essential for the monitoring to be combined with an effective mechanism to feed the findings back into the implementation process so that problems can be discovered early enough to take corrective action (Johansson de Silva, Paci and Posadas 2014).

At a minimum, M&E should include the continual monitoring of process indicators. But it may also include one or more "process evaluations" (that focus on the segment of the causal chain from inputs to direct outcomes). When the regular monitoring of trends in key process indicators sends signals that project implementation is off-course—that the target population is not making use of the services provided, that costs are accelerating, that trainees have not acquired the necessary skills, or that there is resistance to adopting an innovation—the information provided by a process evaluation can help to understand the reasons for the observed trends and suggest possible solutions. M&E focuses on implementation, the segment of the causal chain from input to outputs and usually including direct outcomes. But project-level monitoring may also include periodic reporting of higher level results indicators among beneficiaries if reliable information can be collected economically. This is called "performance monitoring."

ROUTINE MONITORING

The routine monitoring of project implementation in M&E measures progress in project implementation and allows comparisons of planned and actual results. The continual monitoring of process indicators, called "trend monitoring," is an indispensable management tool for making informed decisions about project activities and for assessing staff performance. The observed trends in indicators can provide early warning signals to project managers when things go wrong. Trend monitoring can also inform and motivate project staff toward achieving results. The periodic reporting of indicators is also an important part of maintaining accountability to stakeholders.

The periodic reporting of indicators at regular intervals supports all types of program-level evaluations, including impact evaluations. The information it provides may enable project managers to take appropriate corrective actions during implementation to enable an intervention that is off-track to achieve its desired results. In other cases, the monitoring data may indicate that the intervention is having unexpected results that need to be reflected in the impact evaluation, such

as unexpected outcomes or effects on unexpected groups. This may necessitate collecting additional information in the follow-up survey. The monitoring data may also indicate that certain assumptions made in the *theory of change* were not *validated* in practice—for example, that all who are offered an intervention will take it up. In such cases, the analytical methods used in the impact evaluation may need to be modified. In extreme cases, the monitoring data may indicate that an intervention effectively failed and must be repeated. If an impact evaluation was planned, this information could save evaluation resources from being wasted on an impact evaluation of a failed intervention.

Routine monitoring involves the continual collection of project-level data on inputs, activities (processes), outputs, and direct outcomes to measure a previously agreed list of indicators—and the systematic analysis and reporting at regular intervals of the indicators to project managers and stakeholders. The most common pitfall in routine monitoring is to fail to implement it during the early stages of an intervention, when the information provided would be most useful. In many projects, unfortunately, routine monitoring is not implemented at all because implementers and stakeholders do not understand its potential value. Another common pitfall in routine monitoring is to collect too much information without the ability to use it effectively. Collecting large amounts of data with no analysis will not be very helpful to managers and will not generate the actionable information needed to achieve *efficient* and *effective* implementation. Even in the absence of perfect data, an effectively implemented monitoring system can support useful analysis and provide helpful feedback to project managers.

FOR AN INTERVENTION DISTRIBUTING IMPROVED TECHNOLOGIES

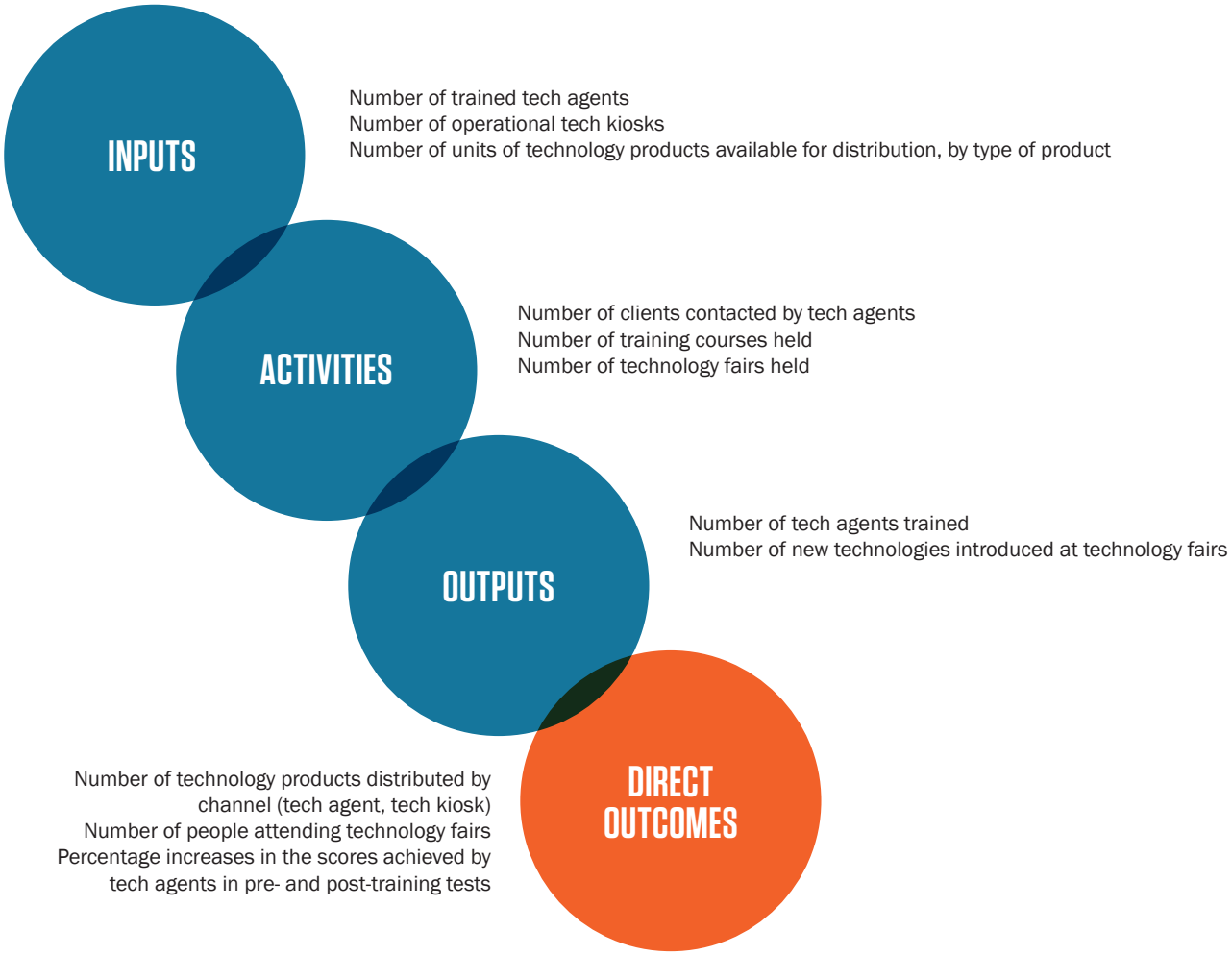
Figure 4.1 provides an example of process indicators that could be used for an intervention designed to empower women through the distribution of improved technologies. This intervention recruits and trains women to serve as "tech agents" to sell the technologies sponsored by the project (water filters, solar lights and batteries) for which the women receive a commission. The intervention also distributes its products through fixed sites ("tech kiosks") and "tech fairs" that introduce new technologies.

The two core activities of this intervention are distributing technology products, either through the outreach sales efforts of tech agents or directly through sales at tech kiosks, and providing training in basic business skills to tech agents. The indicators directly linked to these two core activities are listed in figure 4.1. The most important process indicator is probably the "Number of units of technology products distributed" (a direct outcome). Project managers will want to monitor closely the trends in this indicator over time. If it dips or surges in a given reporting period, they will need to determine why. Another important indicator is "Percentage increases in the scores achieved by tech agents in pre- and post-training tests" (also a direct outcome), which measures the effectiveness of the training activity.

Trend monitoring does not have to be limited to the primary indicators listed in figure 4.1. It may also be useful to monitor "composite indicators" constructed

from two or more of the indicators listed in the figure. For example, it is reasonable to expect that the "Number of units of technology products distributed" would be directly related to the two input indicators "Number of trained tech agents" and the "Number of operational tech kiosks" available in each reporting period. In this case, it may be more useful to monitor trends in composite indicators such as "Number of units of technology products distributed by tech agents per trained tech agent" and the "Number of units of technology products distributed at tech kiosks per operational tech kiosk."[13] Monitoring trends in the composite indicator "Number of clients contacted by tech agents per trained tech agent" might also be useful. The identification of appropriate composite indicators is an example of how the output of a routine monitoring system can be further developed into an effective management information system.

FIGURE 4.1. PROCESS INDICATORS FOR AN INTERVENTION SUPPORTING THE DISTRIBUTION OF NEW TECHNOLOGIES

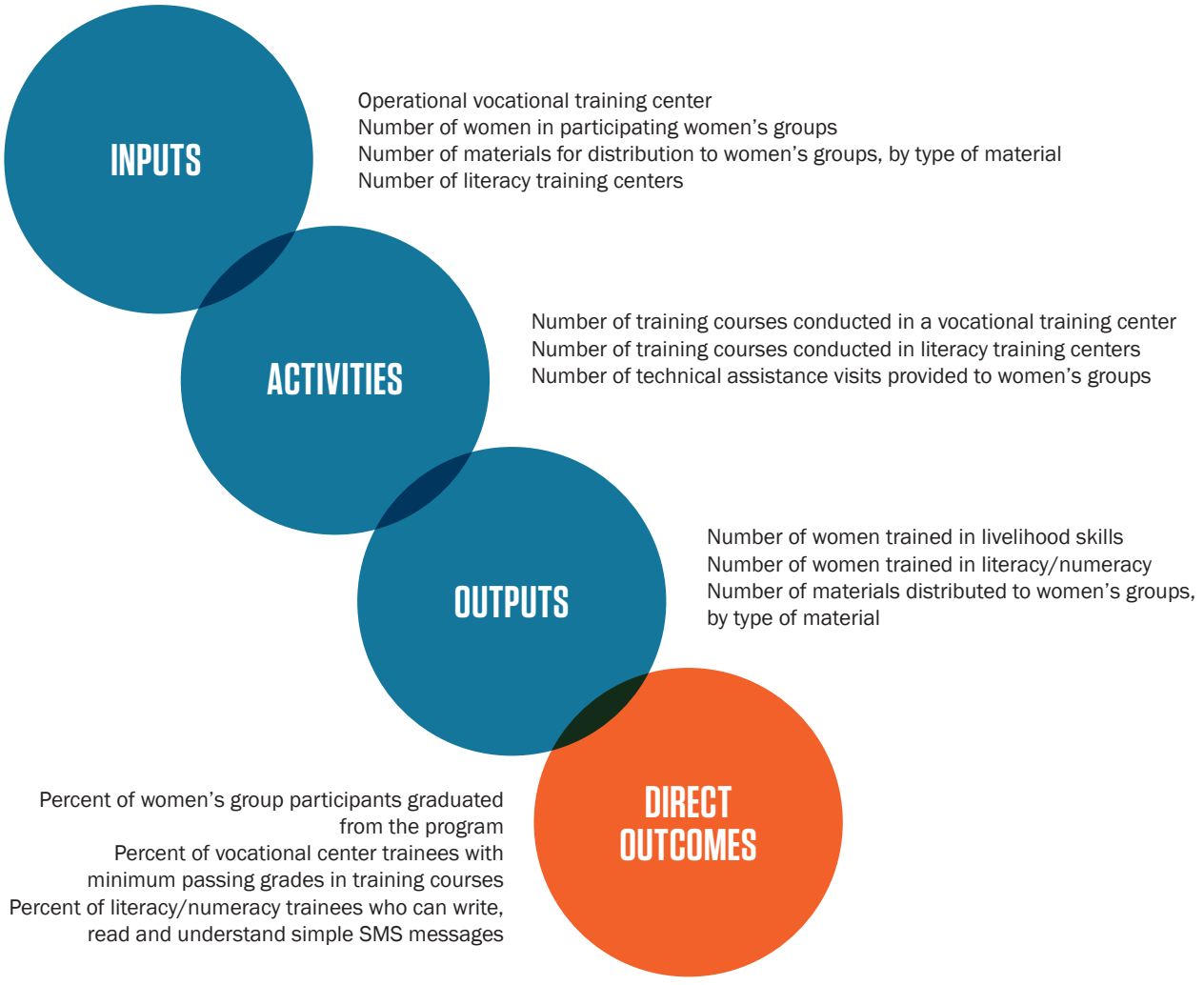


13 These two composite indicators could also be turned into a single composite indicator "Number of units of technology products distributed per sales point" with the number of sales points defined as a weighted sum of tech agents and tech kiosks, with weights based on the average number of units of technology products distributed by each type of sales point.

FOR AN INTERVENTION SUPPORTING RURAL LIVELIHOODS

A second example is a project supporting rural livelihoods. The two main interventions of this project are livelihood and literacy/numeracy training for the members of women's groups participating in the project—and technical, material, and financial support for the women's groups. Illustrative process indicators directly linked to these interventions are listed in figure 4.2. The most important training-related indicator is probably the "Number of women trained by the program" (an output indicator), while the most important technical, material, and financial support indicator is probably "Number of materials distributed to women's groups, by type of material" (an output indicator). Project managers would be expected to monitor trends in both over time. Another important training-related indicator is "Percent of women with minimum passing grades in training courses" (a direct outcome), which measures the effectiveness of the project's training. Again, composite indicators calculated from two or more of the indicators in figure 4.2 might support more effective trend monitoring. An example might be "*Percentage of available materials* distributed to women's groups."

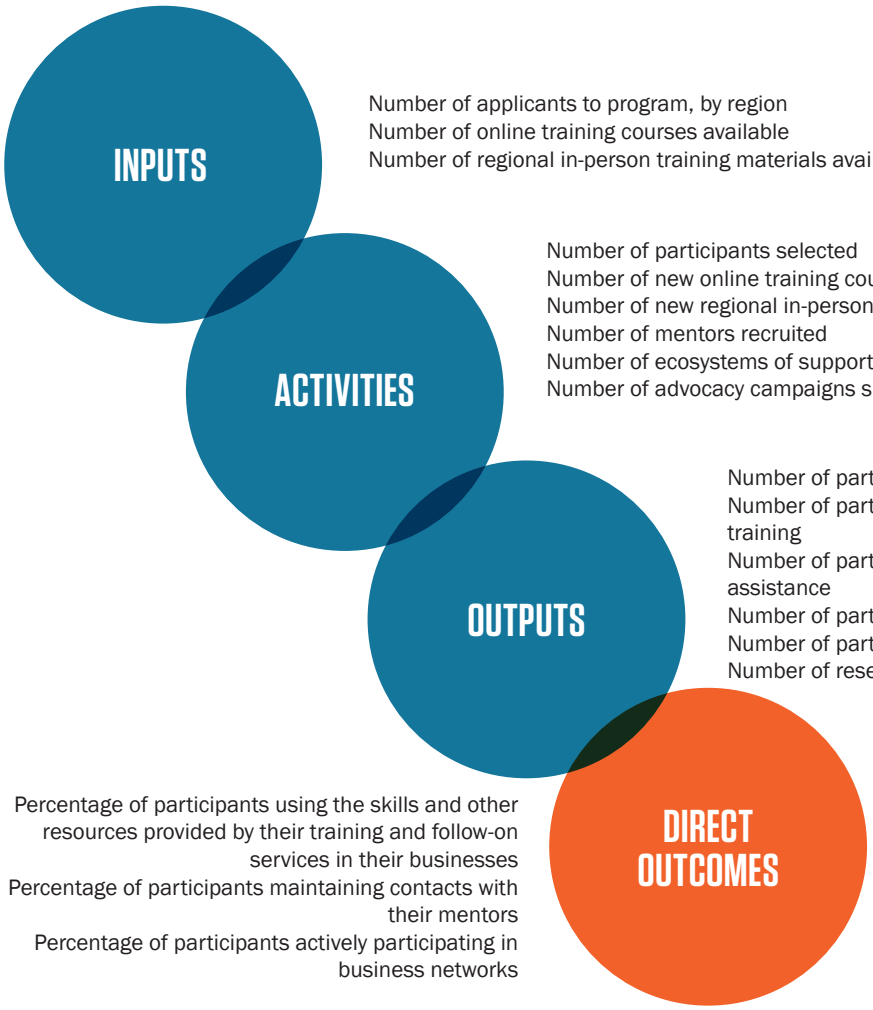
FIGURE 4.2. PROCESS INDICATORS FOR A PROJECT SUPPORTING RURAL LIVELIHOODS



FOR AN INTERVENTION SUPPORTING PROMISING ENTREPRENEURS

A third example is a program to identify and support promising women entrepreneurs in Sub-Saharan Africa, Middle East and North Africa, and Latin America/Central America. A set of targeted interventions would cultivate women's business leadership by fostering their growth aspirations, building their confidence, and integrating them into networks that can expand their businesses by reaching new buyers and markets. The program uses eight interventions to achieve its objectives: tailored mentoring using established business leaders; a year-long fellowship program with online preparatory coursework; regional in-person practical training; tailored technical assistance services to address gaps in participants' action plans; support for regional businesswomen's networks; mapping the ecosystem of support for participants' businesses; advocacy for a more enabling entrepreneurial environment for women; and research on women-focused business networks and on access to finance. Figure 4.3 lists illustrative process indicators for this program.

FIGURE 4.3. ILLUSTRATIVE PROCESS INDICATORS FOR AN ECONOMIC EMPOWERMENT AND ENTREPRENEURSHIP PROGRAM



PROCESS EVALUATIONS

Process evaluations complement routine monitoring. They provide more in-depth understanding of the project implementation process from inputs to direct outcomes. And they are usually conducted in the early stages of a project to learn why planned outputs and direct outcomes are not being achieved or when there are specific operational concerns. They usually focus on the processes of converting inputs into outputs and outputs into direct outcomes and answer such questions as the following (Rubio 2012):

- Is the intervention being implemented according to its design?
- What activities are involved in delivering a product or service?
- Are the operational procedures appropriate to ensure the timely delivery of products or services of acceptable quality?
- Are operational procedures *efficient*, or are there possibly better ways to implement the project?
- What is the level of compliance with the operations manual?
- Are there adequate resources (money, equipment, facilities, personnel with the necessary skills) to ensure the timely delivery of products or services of acceptable quality?
- Are adequate systems in place (financial, management, information) to support project operations?
- Are project beneficiaries receiving products and services of acceptable quality?
- Are project beneficiaries generally satisfied with the processes and with the services they are receiving?
- Are there any operational bottlenecks?
- Is the project reaching the intended beneficiaries? Are project outreach activities adequate to ensure the desired level of participation?

Managers can use the information from process evaluations to determine whether they need to make any midcourse corrections to improve project *efficiency* and *effectiveness*.

The most common sources of information for process evaluations are project records and reports, interviews with project managers, project staff, and beneficiaries, and direct observations by the evaluators. The analysis in connection with a process evaluation is usually more detailed than is possible with trend monitoring. A process evaluation may require going back to the original project records to obtain more detailed information, and it may in some cases require the collection of additional data using such data collection methods as mini-surveys, focus groups, key informant interviews, beneficiary/nonbeneficiary interviews, exit interviews, community interviews, direct or participatory observation, expert opinion, case studies, and literature reviews.[14]

In the technology distribution intervention (figure 4.1) a process evaluation might look closely at an indicator such as the number of clients contacted per tech agent.

14 These data collection methods are discussed in Imas and Rist (2009).

This might involve drawing a sample of tech agents and using available project records to prepare a database that would include the number of clients contacted each month by each sampled tech agent together with the individual characteristics of the tech agents (age, education, other occupation, previous work experience, scores on training tests) and selected characteristics of the geographical areas where they work (total population, population density, characteristics of the road network). Analysis of these data by evaluators might lead to insights about how to select tech agents more effectively, which tech agents may need additional training, or which geographical areas are more suitable for expansion. If the analysis indicates that test scores are an important determinant of tech agent performance, project managers may decide to focus more on improving the effectiveness of tech agent training. A process evaluation might also analyze trends in the number of technology products distributed to see if they show any relationship to the timing of technology fairs. For example, is there any evidence that the number of products distributed increases in areas where tech fairs have been held, and is the increase temporary or permanent? Alternatively, is there evidence of increased distribution of technology products after a tech kiosk is established in the area, or do the data indicate that it merely "cannibalizes" the sales of tech agents?

In the project supporting rural livelihoods (figure 4.2), a process evaluation might look closely at the factors related to the indicators of direct outcomes achieved by the project (the percentage of women's group participants graduated from the program and indicators of the skills they have acquired) and how these are related to the characteristics of the individual women participants as well as the characteristics of the group to which they belong. Again, this would probably require constructing a special database from the available project records that would include not only the characteristics of individual women participants but also data on the inputs provided to each women's group (the number or value of the types of materials provided to each group, the types of training provided, the value of credit provided, and the number of technical assistance visits provided). By correlating project outcomes to these individual and women's group characteristics, evaluators are likely to obtain findings that would help project managers fine-tune their activities to increase project *efficiency* and *effectiveness* over time. A process evaluation might also look at take-up rates of the various interventions (literacy and livelihoods training, participation in the savings scheme), identify any constraints to take-up, and suggest possible solutions.

In the program identifying and supporting promising women entrepreneurs (figure 4.3), a process evaluation might focus on the learning outcomes of the training provided through the program. A basic assessment could be done online, possibly supplemented by a more in-depth assessment during the in-person regional training courses. If done online, it would be possible to assess the learning outcomes of past cohorts of fellowship recipients to see whether learning outcomes are sustained or possibly even enhanced with the passage of time. The individual results might be related both to the personal characteristics of the participants (education, age, region) and to their individual outcome indicators as tracked by the program—such as the percentages of participants who report that their business

practices have improved. The findings might be used to improve the program's selection criteria or its training methods.

The preceding are examples of the kind of quantitative analysis that can be done in a process evaluation using administrative data collected by the project. In many cases, process evaluations will require the assistance of evaluation specialists, particularly for the preparation of special databases and for multivariate analysis. But this type of assistance can usually be obtained in-country.

There is also frequently a role for qualitative analysis in a process evaluation. *Participatory appraisal* is an evaluation methodology that is well-suited to process evaluations. It involves self-assessment, collective knowledge production, and cooperative action. The stakeholders in a development intervention participate substantively in identifying evaluation issues, designing the evaluation, collecting and analyzing data, and taking action as a result of the evaluation findings (Imas and Rist 2009). Participatory appraisal emphasizes the voices and decisions of program participants when analyzing implementation difficulties or program effects, or when information is needed on stakeholders' knowledge of project goals or their views on progress.

Participatory appraisal usually employs *rapid assessment techniques*, which are simpler, quicker, and less costly than other conventional data collection methods, to collect both qualitative and quantitative data. It can provide a quick turnaround to see whether projects are basically on track (Rubio 2012). But the credibility and *reliability* of participatory appraisals (and of their rapid assessment techniques) may be more open to question because of their susceptibility to individual bias and preconceptions and a lack of quantitative data that can be easily *validated* (Bamberger 2012). The groups conducting a participatory appraisal may be dominated by a few vocal people. The participants may defer to politically powerful, wealthier, or more educated group members. Or a group facilitator may direct the group toward certain conclusions.

5. IMPACT EVALUATION

Impact evaluations focus on specific interventions supported by a project and address questions such as (Rubio 2012):

- Has the welfare of the intervention's targeted beneficiaries' improved, and are these improvements attributable to the intervention or would they have occurred anyway?
- Are there any unintended *effects*, positive or negative, of the interventions either on the intervention's intended beneficiaries or on other groups?
- Are the intervention's costs justified by its estimated effects?
- Is the intervention the least-cost intervention among those that obtain the same effects?

Impact evaluations usually focus on intermediate and final outcomes and are necessary because these outcomes are usually affected by many factors other than the intervention. The central question of an impact evaluation is not only what happened after the intervention was implemented (the focus of *performance monitoring*) but what would have occurred *in the absence of the intervention*.

Impact evaluations tend to be expensive, require careful planning beginning at the intervention design stage, and may affect the way an intervention is implemented, including its cost. So they should not be done for all interventions. The first question, then, is when should an impact evaluation be done and when not. The simplistic answer is when its expected benefits exceed its expected costs. Its expected benefits are likely to be highest for existing interventions that are candidates for scaling up, both within the WEE program supporting the evaluation and potentially within similar programs supported by other organizations. An impact evaluation's expected benefits may also be high for a new intervention that is considered particularly promising for its potential to be a least-cost solution to an important problem. Even in such cases, however, certain preconditions increase the likelihood of a successful impact evaluation of a pilot intervention (box 5.1).

The expected benefits from an impact evaluation will be lower for interventions that have already benefited from several careful impact evaluations in similar contexts. For example, an intervention providing immunizations to infants to reduce infant mortality does not normally need an impact evaluation because several have already been conducted in a variety of settings and have demonstrated that immunization, when implemented correctly, is an effective (and cost-effective) intervention against infant mortality.

The expected costs of an impact evaluation will also vary according to the intervention's characteristics and the level of personnel and other costs in the area where the evaluation is to be conducted. An impact evaluation in India, for example, would tend to cost less than a similar one in Sub-Saharan Africa. Because the findings of a credible impact evaluation can add substantially to the global knowledge base of which interventions are most effective in which settings, it is often possible to find supplementary sources of funding to support an impact

Box 5.1: Preconditions for a successful impact evaluation of a pilot intervention

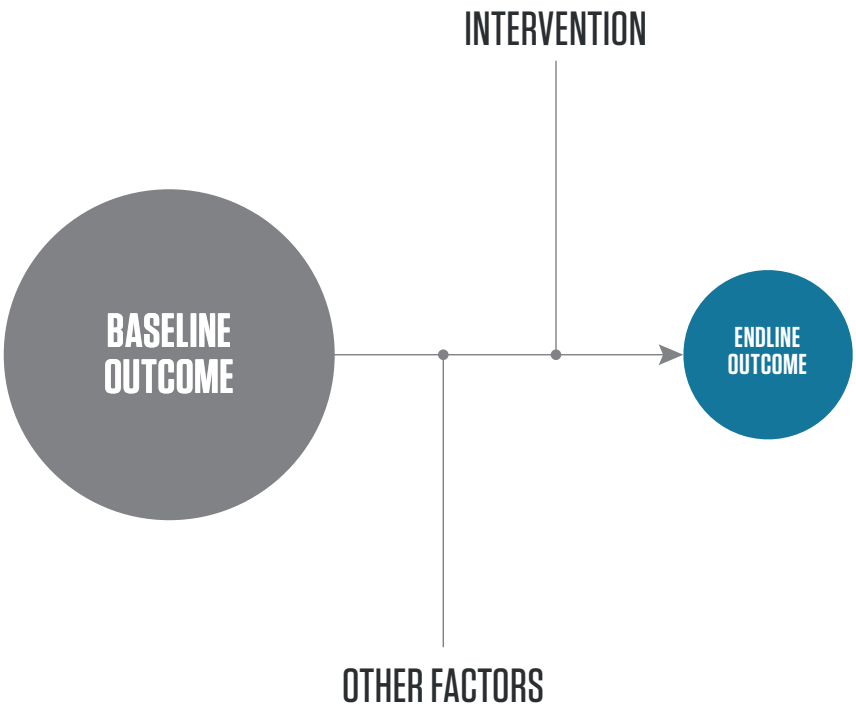
- The overall objectives of the intervention need to be realistic and aligned with available time and resources.
- Interventions need time to show impacts; measuring endline results prematurely can be very misleading.
- Rigorous quantitative impact evaluations without adequate time and resources for planning, implementation, and evaluation are pointless. Alternative methods, such as a good qualitative study, may then provide a more effective option.
- Quantitative data collection needs to be supported by qualitative information, and a wide range of process and outcome indicators should be used to obtain a complete picture of results and to hedge (with qualitative data) against potential data collection and other measurement failures.
- An adequate monitoring system is essential to ensure effective implementation, by providing early warnings when it is necessary to adapt the program to unforeseen circumstances—and effective evaluation, by providing critical information on intermediate changes and processes.

evaluation. For example, highly skilled university researchers can often find their own funding to support their participation in an impact evaluation.

BASIC CONCEPTS

The key to an effective impact evaluation is being able to measure the causal effects of a development intervention on one or more outcomes of interest. An "effect" is an observed change in an outcome that is a direct or indirect consequence of the intervention, whether intended or unintended. It is tempting to conclude that any observed change in an outcome between the pre-project baseline and the post-project endline can be attributed to the intervention. However, this is not the case. Change in an outcome between the *baseline* and the *endline* (as represented by the large horizontal arrow) is determined not only by the "intervention" but also by "other factors" (figure 5.1). If it were possible to measure all of the "other factors" accurately, it would be possible to obtain a reasonably accurate estimate of the intervention's effects in many cases. But only some "other factors" can be easily measured (weather, prices). Some "other factors" that may affect the intervention's outcomes are not easily measured (such as cultural and genetic factors, ambition, individual preferences/tastes, and the individual commitment of local officials). Since most outcomes are plausibly affected by these *unobservable* factors, it is not apparent from such a before-and-after comparison how much of any observed change in the outcome is due to the project and how much is due to the unobservables. Impact evaluations based on before-and-after comparisons may be credible in unusual circumstances, but their credibility depends on the validity of the assumption that the outcomes of interest are unaffected by all other factors.

FIGURE 5.1. IMPACT MEASURED BY BEFORE-AND-AFTER VALUES OF AN INTERVENTION'S OUTCOMES

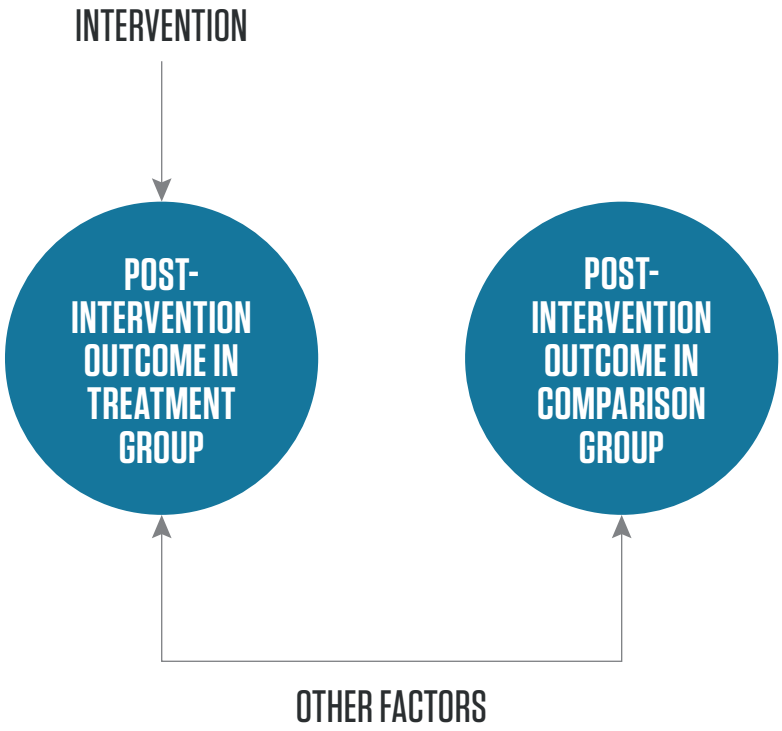


It is tempting to estimate the counterfactual by comparing the outcome of a group that did not receive the intervention (called the "comparison group" or "control group") with that of the group that did (called the "treatment group"). Sometimes the comparison group is selected on the basis of its perceived similarity to the treatment group (similar location, similar level of development, facing similar need), and sometimes special statistical techniques are used to identify a comparison group with the same or similar observable characteristics to those in the treatment group. But impact cannot usually be estimated credibly by comparing post-intervention outcomes in the two groups. The post-intervention differences in the outcomes of both groups are affected by other factors, some of which may be *unobservable* (figure 5.2). It generally is not feasible to identify the separate effects of the intervention and of other factors on the group that received the intervention.

Unobserved differences between treatment and comparison groups are more likely if the people in the treatment group take the initiative to receive the treatment (for example, women who apply to participate in a training course) and even more likely if participation is based on selection criteria used as part of the intervention. Alternatively, those implementing an intervention may decide where to implement it on the basis of criteria likely to be correlated with the outcomes (for example, locating an intervention in an area where it is most likely to succeed). In either case, any difference in the outcome between the treatment and comparison group will reflect not only the effect of the intervention but also the effect of the unobserved characteristics that affect participation or access. Estimating the

intervention's impact as the difference in the outcome between the two groups would in this case yield a "biased" estimate (an estimate that does not reflect the true effect). This type of bias, called "selection bias," is a common problem in many impact evaluations.

FIGURE 5.2. IMPACT MEASURED BY COMPARING POST-INTERVENTION OUTCOMES IN A TREATMENT AND A COMPARISON GROUP

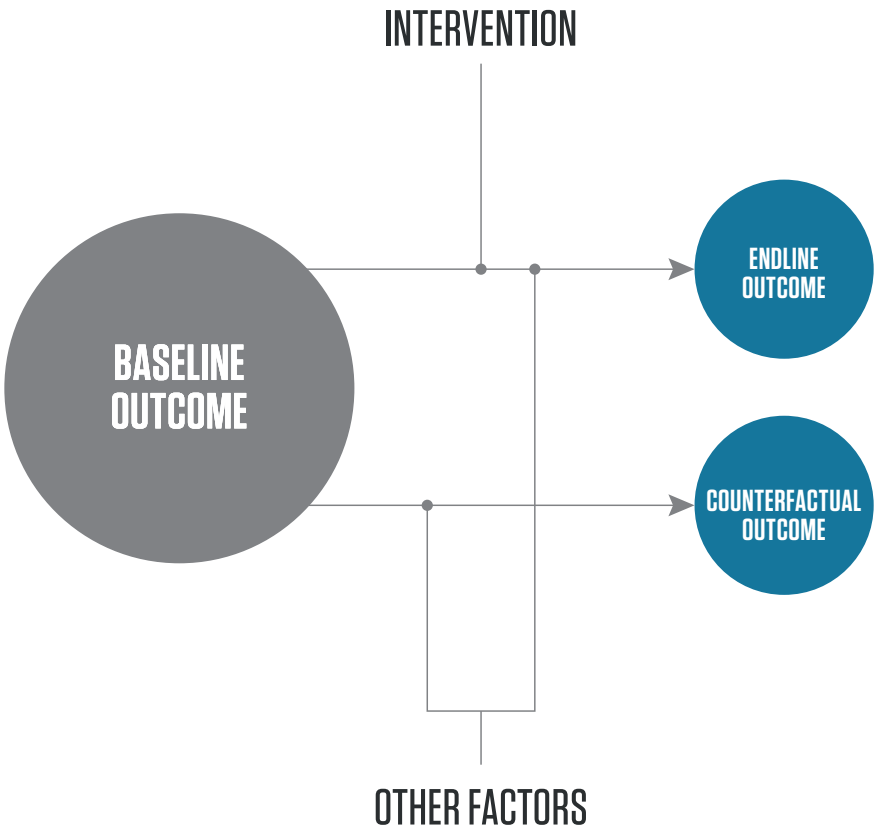


What is usually needed to measure an intervention's impact is an estimate of what the endline outcome would be if the project had not occurred. This is called the *counterfactual*. If the counterfactual is known, it is possible to estimate impact as the difference between the actual endline outcome and the counterfactual outcome (figure 5.3). Constructing a credible estimate of the counterfactual is the main challenge of impact evaluation. Unless an impact evaluation can provide a credible estimate of the counterfactual outcome(s), it is probably not worth doing. (Principle 3: The what and the how of an evaluation matter equally: "what" refers to the outcomes measured, "how" to the evaluation design; Principle 4: No evaluation is better than a poorly designed evaluation.) There are several ways to obtain credible estimates of counterfactual outcomes.[15] But among them, most effective and practical in most WEE programs is a "randomized controlled trial" or RCT.

An RCT is a *prospective evaluation* of a development intervention in which the individuals that will receive the intervention (the treatment group) are selected

15 For a description of several alternative approaches, see Gertler, Martinez, Premand, Rawlings and Vermersch (2011).

FIGURE 5.3.IMPACT MEASURED AS THE DIFFERENCE BETWEEN THE ENDLINE AND COUNTERFACTUAL OUTCOMES OF AN INTERVENTION



randomly, while those not selected serve as the "control group" of individuals not receiving the treatment (at least initially).[16] In the absence of the intervention, the treatment and control groups in an RCT would be expected to have the same outcomes, with any observed differences due only to *sampling error*. And in a correctly implemented RCT, there is no *selection bias* (box 5.2). RCTs have been used for many years in agricultural and medical research (as in drug trials). They have also been used extensively to evaluate domestic social programs in the United States. Their use in international development was rare until recently. Although initially resisted as impractical and possibly even unethical in some circumstances, RCTs have become increasingly common in developing countries during the past 10 years, with growing acceptance.[17]

A classic RCT has six steps (Imas and Rist 2009):

1. Formulate a hypothesis.
2. Obtain baseline measures of the outcomes of interest.

16 Some authors use the term "comparison group" instead of "control group" even in the context of an RCT. However, the term "comparison group" is traditionally applied only to *quasi-experimental designs*. See, for example: http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20188242~menuPK:415130~pagePK:148956~piPK:216618~theSitePK:384329,00.html

17 "For impact evaluations, experimental methods generate the strongest evidence. Alternative methods should be utilized only when random assignment strategies are infeasible." USAID Evaluation Policy (2011). For a dissenting view, see Angus Deaton (2009).

3. Randomly assign subjects to intervention (treatment) and nonintervention (control) groups.
4. Implement the intervention in the treatment group.
5. Obtain one or more follow-up measures of the outcomes of interest.
6. Calculate the differences in the outcomes between the treatment and control groups and test for their statistical significance.

Random assignment ensures that differences in observed and unobserved "other factors" between the treatment and control groups are due only to sampling error.

WHEN IS AN RCT PRACTICAL, AND WHEN NOT?

OPPORTUNITIES FOR CONDUCTING AN RCT

There are many opportunities for conducting RCTs (table 5.1). These include situations in which access to all or part of an intervention, or the timing of access,

Box 5.2: Selection bias in impact evaluations of business training

Governments and donors have spent many millions of dollars on entrepreneurship training programs around the world, despite the absence of convincing evidence of their impact (Woodruff and McKenzie 2012). The impact of a large business training program in the United States (the GATE project) was evaluated recently by an RCT in which 4,197 applicants interested in starting or improving a business were offered best-practice business training by 14 different government and non-governmental organizations at seven sites in three states. Follow-up surveys were conducted at 6, 18, and 60 months after random assignment.

A recent study of the findings failed to find any significant impact of business training on most business outcomes including: sales, profits, employment, business size, work satisfaction, the survival of pre-existing businesses, or household income (Fairlee, Karlan and Zinman 2013). Although the study did find a short-term impact on business ownership among those unemployed at baseline, the effect disappeared after 18 or 60 months.

The study also obtained nonexperimental estimates of the "effects" of business training on several outcomes by restricting the analysis to the control group (which received no business training through the GATE project but some individuals obtained entrepreneurship training from other sources). The results indicate misleadingly that the training had significant positive "effects" on business ownership (at 6, 18, and 60 months), on monthly business sales (at 18 and 60 months), on whether the business has any employees (at 6, 18, and 60 months), and on household income (at 6, 18, and 60 months), even when controlling for a large number of covariates. These results suggest that the use of nonexperimental data to estimate the impact of business training would be subject to a strong upward selection bias.

can be randomized. Several of the opportunities listed in table 5.1 would be likely to occur with the interventions in figures 2-4. For the intervention supporting the distribution of new technologies (figure 4.1), expansion of the intervention into a new regency (the administrative level immediately below a province in Indonesia) could provide an opportunity to conduct an RCT, particularly if the villages to be covered could be randomly phased in over time. In this case, the villages (or tech agents serving a group of villages) to receive the intervention in the second phase could be used to estimate the counterfactual during the first phase. This is an example of the "admission in phases" opportunity for randomization in table 5.1. As long as the time between phases is long enough to observe the effects of the intervention on the outcomes of interest, the RCT should be able to obtain an *unbiased estimate* of the effects of the intervention.

Alternatively, a "randomized trial" (a randomized experiment without a control group) can be conducted to evaluate different ways of implementing an existing intervention. For example, a randomized trial (RT) could evaluate the effects of alternative incentive schemes for tech agents in the intervention supporting the

TABLE 5.1. OPPORTUNITIES TO CONDUCT RCTS

OPPORTUNITY	DESCRIPTION
New program design	When a problem has been identified but there is no agreement about which solution to implement. The research team works with the implementers from the outset to design an intervention that is suitable for field testing
Admission in phases	When logistical and resource constraints prevent all potential beneficiaries from receiving an intervention immediately so that they can be randomly assigned to receive it in phases over time
Admission cutoffs	When an intervention uses a minimum (maximum) cutoff point to determine eligibility and those just below (above) the cutoff can be randomly assigned to eligibility in order to evaluate the desirability of lowering (raising) the cutoff point
New interventions	When a new intervention is being pilot tested
New services	When an existing project adds a new intervention
New people	When the target group of an existing intervention is expanded
New locations	When the geographical area served by an existing intervention is expanded
Oversubscription	When there are more people interested in an intervention than can be served
Undersubscription	When not everyone eligible for an intervention takes it up
Rotation	When an intervention's benefits or burdens are to be shared by rotation

Source: Adapted from Glennerster and Takavarasha (2012).

distribution of new technologies (figure 4.1). In this case, there would be multiple treatment groups (sometimes referred to as "treatment arms"). The main outcome would be the number of technology products distributed by tech agents (a direct outcome), and the values for each treatment group would be compared with each other or with a "control group" of tech agents using the current system of incentives (or no incentives). RTs might also be used to identify the effects of different durations of training on the pre- and post-training test scores of trainees or the effect of randomly placed tech kiosks or technology fairs on the numbers of technology products distributed.

These examples of RTs focus on the implementation segment of the results chain (although in some cases the alternative treatments might also have different higher level effects). They are relatively inexpensive to conduct because project records can usually provide data on the direct outcomes (no expensive household surveys are needed). It is increasingly common to use RTs in this way (box 5.6). They may be viewed as a type of "operations research" of using advanced analytical methods to improve performance.

The impact of the project supporting rural livelihoods (figure 4.2) could also be evaluated in an RCT, for example, if the project were expanded to one or more of the remaining prefectures in the same region. Villages (or clusters of villages)

Box 5.6: Randomized trials to improve the performance of savings schemes

Several RTs have been conducted to improve the performance of the savings schemes supported by microfinance providers. In Guatemala, different savings schemes were randomly offered to microfinance borrowers of Guatemala's largest public-sector bank (Atkinson, de Janvry, McIntosh, and Sadoulet 2013). Prompting for savings deposits at the time of loan repayment doubled savings compared with savings accounts without such a reminder, and a savings deposit equal to 10% of the loan payment led to a further doubling of savings.

Similar RTs in Bolivia, Peru, and the Philippines found that including monthly reminders to randomly selected savings account holders, either by mobile phone or by letter, increased savings by 6% compared with account holders not receiving reminders.

Another RT in Chile focused on 2,687 microentrepreneurs who were members of 196 self-help groups (Kast, Meier, and Pomeranz 2012). The groups were randomly assigned to one of the following: a basic savings account, the basic account plus a self-help peer group component, and a basic account earning a higher interest rate. Adding the self-help peer group component to the basic account increased the number of deposits more than three-fold and almost doubled average savings, but the substantially higher interest rate had no effect on most participants.

within the added prefecture(s) would need to be randomly allocated to different treatment phases. The women's groups in the localities that would receive the treatment in the second and any successive phases, could serve as controls until they, too, receive the intervention. Alternatively, an RT may be feasible if the take-up of some interventions is low. In the agriculture project, the literacy training, the preparation of business plans or the project-supported savings plans could be used to evaluate the effects of using various forms of "encouragement" to boost participation (the "undersubscription" opportunity in table 5.1). Such encouragement designs are feasible when there are enough resources to treat everyone but take-up is low (for example, if there is unused capacity in the literacy training intervention).

The encouragement provided should address the main constraints to take-up, which might be identified through a *process evaluation*. Examples might be reimbursement of bus fares to attend literacy training sessions or periodic reminders to make deposits in a savings account. It is critical that the encouragement have a positive effect on take-up (that take-up rates are higher in the treatment than in the control group) and that it not be too large (particularly if it involves financial incentives) so that it could affect outcomes other than take-up. In an encouragement design, the treatment group receives both the intervention and the encouragement, while the control group receives only the intervention. In this case, the RT would measure the effect of the intervention on those who take it up due to the encouragement. The RT would not be able to measure the effect of the intervention on those who had already taken it up. That would require an RCT in connection with a more costly project expansion, as described above.

The limited number of fellowships awarded annually in the program identifying and supporting promising women entrepreneurs (figure 4.3) might prevent the use of an RCT to measure the program's impact. But there may be opportunities for RTs to be used for operations research within the program. For example, an RT might estimate the effects of alternative mixes of training, such as two periods of regional in-person training sessions instead of only one. But the limited number of fellowships awarded in each year might in this case require that the RT be conducted for several years in order to obtain reliable estimates of the effect.

LIMITATIONS AND DRAWBACKS OF RCTS

Although RCTs provide the most reliable information about an intervention's impact, they have some inherent limitations and drawbacks.

Not all development interventions are suitable for an RCT. In some cases, the counterfactual is so obvious that no control group is needed. White (2010) cites the example of the impact of a water supply project on the time household members spend collecting water. The counterfactual in this case is clearly the situation of the treatment group before the intervention. Nor is an RCT needed for interventions that have been proven effective by other RCTs in similar settings (immunizing infants to prevent child mortality). RCTs are also unsuitable for interventions designed to affect outcomes that are particularly sensitive and

therefore difficult to measure (such as reducing domestic violence). In addition, any intervention that must be implemented uniformly at a national level is unsuitable for an RCT (such as macroeconomic policy changes). Some infrastructure investments (roads, ports, bridges) are not suitable for RCTs. Isolating the effect of mobile phone technology in an RCT is also usually impractical due to the rapid and nearly universal adoption and use of mobile phones (though evaluating the impact of alternative interventions using an existing mobile phone platform is clearly feasible). In some cases, RCTs may not be practical due to effective limits on the number of subjects that can be randomized in small programs.

There are also some important drawbacks to RCTs in some cases. One is the length of time to conduct some RCTs. RCTs, like all experiments, are *prospective evaluations*. Work begins at the planning stage, continues through implementation, and extends into the post-implementation stage. RCTs are best suited, therefore, to interventions that can be implemented relatively quickly (or at least to interventions that produce measurable results relatively quickly). Otherwise, there can be considerable delays in obtaining the results. The full effects of a randomized infant feeding intervention may take many years to materialize fully. In such cases, sample *attrition* can become a major problem (Maluccio and others 2009). Perhaps most importantly in the context of these guidelines, an RCT may complicate and increase the cost of the intervention it is designed to evaluate (say, by requiring the intervention to cover a larger geographical area or by also requiring data collection in the control group). Although these drawbacks are not arguments against conducting RCTs when they are needed, they suggest that those supporting an RCT should be prepared to allow enough time for impacts to materialize and to provide enough resources to support both the intervention and the RCT (Johannson de Silva, Paci, and Posadas 2014).

It is sometimes suggested that RCTs are unethical because they deprive individuals in the control group of the intervention's benefits. This may indeed be a valid concern in some cases.[18] Access to an intervention should never be denied solely for the purpose of an evaluation (Gertler and others 2011), but strict adherence to this practice can still be consistent with RCTs in many cases. For example, most development interventions cannot be applied immediately to all eligible individuals, whether due to limited resources or to operational constraints. Even if resources are available to make the intervention available to all eligible individuals eventually, it will still need to be phased in over time. In such cases, selecting randomly who among eligible individuals will have initial access to an intervention is arguably the fairest solution. Even if sufficient resources are *not* available to make the intervention available to all eligible individuals eventually, a convincing finding that the intervention is effective (possible only with the RCT) may lead to additional funding becoming available to make the intervention available to all eligible individuals. If the intervention is not found to be effective, the RCT can

18 In addition to the issues discussed in this paragraph, many countries and organizations have set up institutional review boards (IRBs) or ethnics committees to regulate research on human subjects. Because RCTs involve research, any RCT will need to comply with the procedures mandated by such institutions.

prevent scarce resources from being wasted on an ineffective intervention and lead to the development of a more effective intervention.

Another concern is that RCTs are usually conducted in a very limited geographical area and that their results cannot be generalized to other contexts (other locations, other times, other scales).[19] There may indeed be a tradeoff between "internal validity" (the ability of an evaluation to distinguish between correlation and causality) and "external validity" (the generalizability of evaluation results to other contexts). But it is possible to design RCTs to maximize their external validity by conducting the RCT in a representative location with a representative implementing organization—or conducting an RCT in randomly selected locations covering a wider geographical region (including multiple countries in some cases), albeit at increased cost. It may also be possible to combine what is learned from RCTs with what is learned from other types of impact evaluations.

Concerns have also been expressed about the high cost of RCTs. Although RCTs vary widely in their costs, some are indeed very expensive, particularly when they involve an attempt to evaluate multiple *treatment arms*. The main costs of an RCT are the costs of data collection and research support. These tend to be higher with an RCT (and with other types of experiments) than with some other types of impact evaluations because an RCT involves data collection in control groups and activities extending from the pre-implementation phase into the post-implementation phase. However, whether any impact evaluation is too expensive depends on its expected benefits in relation to its costs, as discussed above.

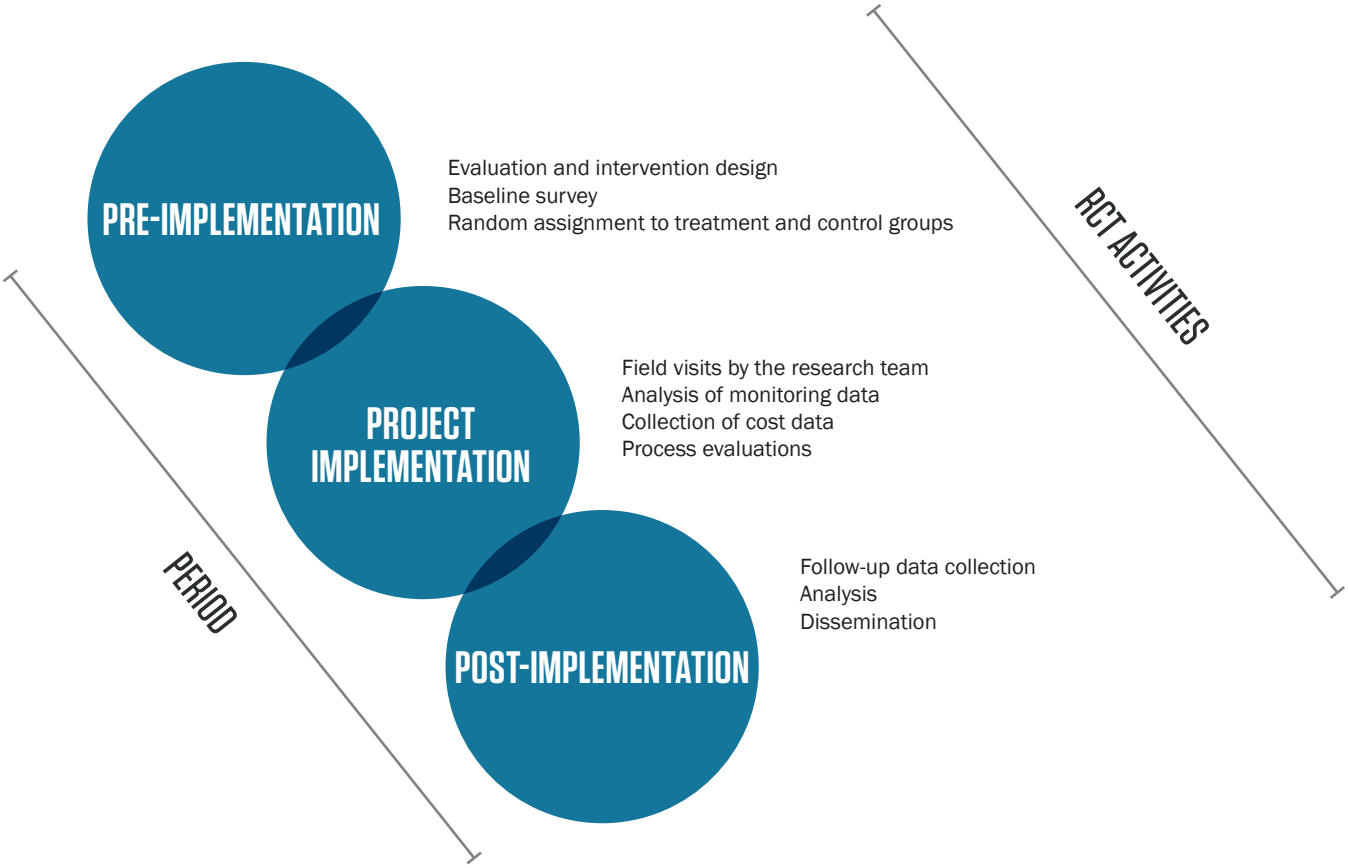
It is sometimes suggested that the money to support RCTs (or any other type of impact evaluation) should be spent instead on the interventions themselves because only the interventions benefit people. This may be true at the project level, ignoring the possibility that the supported interventions benefiting from the proposed reallocation might not be effective. But it is not true at the program level, where decisions about investing in alternative interventions must be made. Such decisions should be made on the basis of what works best in actual field conditions, and the evidence on this is often patchy at best. Moreover, the findings of an impact evaluation are a *public* good. The resulting knowledge can and should be made freely available to other programs so that they can make better investment decisions as well. As the knowledge base of what works best increases, many more people in need around the world will potentially benefit. When decisions about when and how to conduct an impact evaluation are made effectively, the resources used are likely to be among the most productive investments in a development program.

19 There are important exceptions, including the RCT conducted in Mexico to assess the impact of a large conditional cash transfer program and several education RCTs in India (Glennerster and Takavarasha 2013).

HOW DOES AN RCT INTERACT WITH THE INTERVENTION BEING EVALUATED?

Conducting an RCT can be divided into three periods: pre-implementation, implementation, and post-implementation.[20] Figure 5.4 shows the main RCT activities conducted during each period that interact with the intervention being evaluated.

FIGURE 5.4. MAIN POINTS OF INTERACTION BETWEEN AN RCT AND THE INTERVENTION



PRE-IMPLEMENTATION PERIOD

The pre-implementation period, arguably the most important period for an RCT, may require as much as one year to complete. During this period, both the RCT and the intervention itself are designed. It is highly desirable that the implementing team and the research team collaborate closely and effectively on the design of both the intervention and the RCT. Key design decisions concern the level of randomization, the location of the intervention, the way the intervention will be implemented, the monitoring system to support both the intervention and the RCT, and the types of additional data that will be collected to support the RCT. Baseline data collection and "randomization" (random assignment of the experimental units to treatment and control groups) must also be completed during this period. Meetings of the

20 A more detailed general discussion of RCTs, including many topics not discussed here, can be found in Glennerster and Takavarasha (2013).

implementing team with local political leaders or potential beneficiaries to discuss the intervention must not begin until after the baseline data have been collected and all experimental units have been randomly assigned to treatment or control groups (box 5.8).

One of the most important decisions for the research team is the level of random assignment. Random assignment can occur at the individual/household level or at a higher level (such as a village or a women's group). Often the choice is dictated by the nature of the intervention. For example, if an intervention is targeted to women's groups (such as training or providing materials), it is most practical to assign all members of a group (rather than individual members) to treatment or control. This is called a "cluster design." In some cases, the choice of randomization level may involve a compromise between the technical or political-administrative requirements of the intervention—including the way the intervention is delivered and the technical requirements of the RCT (such as its ability to detect an effect when one exists). The level of randomization may also affect the cost of both the intervention and the RCT. When randomization is at the cluster level, the cost of both the intervention and the RCT usually increases due to the need to cover a wider geographical area.

Other important design-related decisions concern where and how the intervention will be delivered. If possible, it should be delivered in an area that is at least broadly representative of the larger area in which the intervention, if successful, is likely to be "rolled out." In this case, "representative" refers both to the extent of the problem to be addressed by the intervention (for example, the extent to which women are not currently empowered economically) and to the broader characteristics of the population that may be expected to affect the intervention's effectiveness (such as education, income, roads, other infrastructure, degree of urbanization, cultural characteristics). A thorough understanding of the context of an intervention is important both to its design and to the design of an impact evaluation (White 2009). Secondary data from a previous census or survey may be available to help in the choice of location. If not, some data may need to be collected as part of an initial assessment during the pre-implementation period. The intervention should also be implemented in a way that is consistent with the way it would be implemented in a "roll-out." Using more highly qualified and compensated staff to implement an intervention than are likely in its roll-out would impair the *external validity* of the RCT.

The design of an appropriate monitoring system is another key point of interaction between the research team and the implementing team during the pre-implementation period. The monitoring system should be designed for two main purposes. First, it should provide feedback to the implementing team and the research team on the progress of implementation and on the validity of the assumptions underlying the intervention's causal chain (the *theory of change*). For example, to what extent has the intervention been taken up by the target population? Have the participants received the planned intervention, and has it resulted in the expected direct outcomes? Second, the monitoring system

should provide information to the research team about possible problems during implementation that can affect the ability of the RCT to obtain a valid estimate of the intervention's effects.

Another important design-related decision is the type of additional data that need to be collected to support the RCT. The findings of RCTs are usually based mainly on quantitative data obtained from baseline and one or more follow-up household surveys. Baseline data must be collected before implementation (follow-up surveys are usually conducted during the post-implementation period). If any implementation activities begin before baseline data are collected, there is a risk that the baseline data will reflect adjustments in behavior made by the target population in response to, or even in anticipation of, the intervention, making it difficult to measure the intervention's causal effects accurately. The main roles of baseline data in an RCT are to check whether randomization was successful, to obtain baseline values of outcomes and covariates for use in estimating effects, and to guide the collection of qualitative data. In some RCTs (such as those with a small number of *experimental unit*s), baseline data can also help in achieving more efficient *random assignment*s.

Beginning implementation only after baseline data have been collected may pose problems for the implementing team. However, rushing the process of conducting a baseline survey can have disastrous consequences. Visits to the field are an essential input for the design of good baseline questionnaires (box 5.7). If the questionnaires are administered in a different language, they should be re-translated into the original language to check that the meanings have not changed. The draft questionnaires should be carefully field tested and appropriately revised after each field test. It is usually recommended that at least three field tests be conducted, with about two weeks for each field test and subsequent revisions to the questionnaire. Both baseline and follow-up data should always be collected under the supervision of the research team, in close cooperation with (but not the participation of) the implementing team. To ensure that the data are credible, the implementing team should not be directly involved in conducting sample surveys. And because conducting sample surveys requires highly specialized skills, interviewers and supervisors should be experienced and receive careful training.

There is also an important role for qualitative data in an RCT (White 2008, 2010). Qualitative data can be collected through a variety of methods, including focus groups, key informant interviews, and beneficiary/nonbeneficiary interviews. The qualitative data may be helpful in understanding why certain anticipated outcomes did not result from a given intervention (or why unexpected ones did occur). (This is an illustration of principle 6: Complementary qualitative work is important to understand the "why" behind results.) It is also useful to be able to *triangulate* findings using both qualitative and quantitative data (Imas and Rist 2009). In other cases, qualitative indicators may provide more accurate measures than quantitative indicators of certain outcomes. Although questions eliciting subjective responses and "open-ended questions" (without predetermined responses) can also be included in a survey, there is less opportunity to probe an initial response

Box 5.7: The importance of context

Howard White (2008) relates his experiences from field visits in connection with two studies. The first was a study on rural electrification in Laos. Initial analysis of a baseline survey using a regression model with distance from the nearest major road and from provincial headquarters and average community income as explanatory variables was unable to explain much of the variation in whether communities were connected to the electricity grid. When he made a field visit to an off-grid village with electricity supplied by solar panels, he began to understand why his regression model did not work. The off-grid village was only a short distance from the major road and the provincial capital and was not particularly poor. But it was an island, as were many of the other off-grid villages he visited on the same trip. He had omitted a key variable from his model—that is, whether the village was on an island. Moreover, the baseline survey had not collected this information in its community questionnaire.

In another field visit in connection with a study on basic education in Ghana, he visited the best and worst schools in one district of the Volta Region. Morning visits to the best schools provided no surprises. In the afternoon, however, he visited the worst school, where the average score on the standardized math test was zero. The school was three hours' walk from the main road, with the last hour along a single-track path through the hills. Four teachers had been posted to the school, but only one had taken up his post. At one point this teacher explained to his visitors, "I would like to put posters on my walls but I have no posters. In fact, as you can see, I have no walls." In fact, there were no walls. The World Bank had financed the construction of a three-classroom structure consisting of a concrete platform, steel girder uprights, and a metal roof. No walls were provided because the communities were expected to provide them. But poorer communities could not afford to complete the structures, leaving their schools with no walls or furniture.

effectively in a survey. It is also relevant to consider the data collection setting. Respondents may react differently to an enumerator than to a specially trained social scientist. The most useful data collection strategy may be to begin with qualitative research, to use the findings to develop the baseline survey, and then to use the survey data to draw a more representative sample for a second round of qualitative research (White 2008).

It is also highly desirable to collect cost data in connection with an RCT. Without cost data, an RCT is limited to measuring one or more treatment effects. Significant effects by themselves do not provide evidence that the intervention is worth replicating or scaling up. Such a decision should also reflect the cost of the resources to implement the intervention. Cost data can also be useful during implementation in process evaluations to investigate whether the intervention is being provided efficiently. For example, a process evaluation can analyze how *unit costs* may vary by location, provider, or scale and make adjustments to improve efficiency during implementation.

The random assignment of experimental units to treatment or control groups must be completed before any implementation activities begin. Otherwise, there is a risk that prior commitments will be made to experimental units that they will receive the intervention. For example, if the implementing team begins discussing the intervention with local leaders in advance, that may raise expectations about the intervention and lead to commitments by the implementing team that affect the research team's ability to assign experimental units to treatment or control groups. This is a particular problem when randomization is at the cluster level and the number of clusters is small (box 5.8).

IMPLEMENTATION PERIOD

The length of the implementation period will vary, depending on the nature of the intervention. The main RCT activities during the implementation period that involve interaction with the intervention are site visits by the research team, analysis of the routine monitoring data, collection of cost data, and process evaluations.

Site visits by the research team to the project sites should be at regular intervals and closely coordinated with the implementing organization. It is often recommended that a member of the research team remain on-site during implementation. If this is not practical, site visits should be at least quarterly and timed to coincide with the availability of the monitoring data. The research team should use the site visits to confirm that the monitoring system is being implemented as planned and that the agreed upon data are being collected and are readily available for analysis (including the information to estimate the intervention's costs). If there have been delays in project implementation, often the case, the research team should determine how they are likely to affect the RCT and revise the evaluation timetable accordingly. Site visits should be made if possible both to the treatment and control areas (if randomization is at the cluster level) to ensure that the site visits have a neutral effect on the behavior of treatment and control groups.

Box 5.8: Random assignment

The Mekong Results-Based Initiative targeted female bamboo handicraft producers in three provinces of Cambodia and Laos, providing them with technical and business training and study visits to Thailand (Knowles, Golla, and Rex 2013). The number of villages producing bamboo handicrafts considered suitable for the project was limited. Unfortunately, preliminary discussions between the implementing team and local officials in Laos resulted in commitments to include 4 of 18 villages as treatment villages in one province and 7 of 15 villages in another province. As a result, these villages could not be randomly assigned to treatment or control groups and had to be excluded from the RCT. This reduced the number of study villages significantly and made it more difficult to obtain significant estimates of project impact.

The research team will use the intervention's monitoring data as an "early warning system." Only with careful analysis of the monitoring data will the research team be able to conclude whether any absence of impact is because of inherent flaws in the intervention or because the intervention was not implemented as planned (Johansson de Silva, Paci, and Posadas 2014). The monitoring data may provide evidence that some of the assumptions in the intervention's theory of change are not valid. For example, the monitoring data may question the assumption that most individuals in the treatment group will take up the intervention, or that all those who do take it up are in the treatment group. If the monitoring data indicate that trainees are not obtaining the expected skills, this may question the assumption that the planned training is appropriate. That information can lead directly to improvements in the intervention to make it more effective, or it may signal the need for the more in-depth information that a process evaluation could provide.

Careful analysis of the monitoring data may also indicate that there are problems of noncompliance, spillovers, attrition, or evaluation-driven effects that can be remedied. These problems are common in any experiment.[21] In some cases, they are caused by design decisions (such as the level of randomization), in others by the way an intervention is implemented.

Noncompliance occurs either when some people in the treatment group do not receive the treatment (box 5.9) or when some members of the control group do receive it. Among several reasons, the most common reason is low take-up in the treatment group and another is that another organization has begun providing the same or a similar treatment to people in the control group.

Spillovers occur when the effects of an intervention are observed among people in the control group, despite their not receiving the treatment. Either positive or negative, they occur most often when treatment and control groups are in close proximity. They may be welcome in some interventions—for example, if the objective is to diffuse information or a new technology. But they make it more difficult to detect a *treatment effect*.

Attrition occurs when some people in the baseline sample drop out of the intervention, relocate, or refuse to respond to some or all questions in a follow-up survey—say, because the questionnaire is too long. The result is missing data (box 5.9). Depending on the characteristics of those who drop out, attrition may bias estimates of the treatment effect.

Evaluation-driven effects occur when participation in an RCT (including both treatment and control groups) can affect people's behavior independent of any effects of the intervention. Common examples include:

1. Hawthorne effects—people in treatment groups work harder.
2. John Henry effects—people in control groups work harder.

21 For a detailed discussion of these problems in the context of an RCT, see Glennerster and Takavarasha (2013).

Box 5.9: Noncompliance and attrition

Noncompliance and attrition are common problems in training programs for women entrepreneurs, reducing sample sizes and the power of statistical tests. For example, close to half of female microentrepreneurs in Lima, Peru, who were offered business training and technical assistance tailored to their businesses did not take up the program due to travel time requirements and lack of child care (Valdivia 2013). The response rate in the evaluation of the Goldman Sachs 10,000 women initiative, covering over 100 cohorts from 11 countries representing more than 3,000 women entrepreneurs, was initially much better (Babson College 2014). But the response rate dropped from 91% at baseline to 58% in the second follow-up survey 18 months later.

3. Resentment and demoralization effects—usually confined to those in the control group, who may resent not having access to the treatment.
4. Demand effects—participants change their behavior in response to their perception of what the research team wants.
5. Anticipation effects—people in the control group change their behavior in anticipation of receiving the treatment in the future.
6. Survey effects—exposure to frequent surveys changes the subsequent behavior of participants.

Discovering these problems early on may make it possible to adjust the intervention to improve the RCT's "statistical power" (its ability to detect a treatment effect when there is one).

In addition to analyzing the monitoring data, the research team should also work closely with the implementing team to collect information on the cost of the intervention. The cost data need to be both accurate and complete (Dhaliwal, Duflo, Glennerster, and Tulloch 2011). A common shortcoming of the cost data collected in evaluations is that they do not cover all inputs. Most often neglected are volunteer labor, donated materials, the time and travel costs of participants, and capital goods.

Analysis of the monitoring data may raise questions that one or more process evaluations can address most effectively. These will typically involve the collection of qualitative data. For example, if take-up is low, it may be possible to use focus groups of eligible beneficiaries to investigate the underlying reasons. The findings may lead to modifications of the intervention to increase take-up. Or, if training is found not to be effective, a special study may be needed to assess the knowledge and effectiveness of trainers.

POST-IMPLEMENTATION PERIOD

The post-implementation period involves less interaction with the intervention than in the other periods and typically takes six months to one year. The main RCT activities that interact with the intervention during the post-implementation period

Box 5.10: Cost-effectiveness analysis and cost-benefit analysis

Cost-effectiveness analysis (CEA) involves comparing the estimated total cost of an intervention with an estimated average treatment effect (such as the cost per job created). To facilitate CEA comparisons across interventions, the definition of costs must be the same across interventions and the cost estimates must be comparable, particularly if they are obtained from different countries (Dhaliwal, Duflo, Glennerster, and Tulloch 2011). Because costs and effects often occur at different times with different interventions, both must be converted ("discounted") to present values using an appropriate *discount* rate. The cost-effectiveness of interventions that incur the costs relatively quickly while the effects are spread out over time will be very sensitive to the discount rate used.

One limitation of CEA is that it can only be used to compare interventions with the same outcomes. Another limitation is that it can be misleading when comparing alternative interventions with multiple outcomes.

Cost-benefit analysis (CBA) can address both of these limitations. The estimated effects are assigned a monetary value and added together to provide an estimate of the intervention's "benefits." These are then compared with an estimate of its costs to determine whether the benefits justify the costs. Because benefits and costs are likely to occur at different times, future streams of both benefits and costs are also converted (discounted) to present values using a discount rate.

An intervention is generally considered to be attractive if the present value of its benefits exceeds the present value of its costs for a plausible range of discount values. This implies that the total value of the benefits obtained from the intervention are greater than the value of the resources used to obtain them. The main challenge with CBA is to assign a monetary value to effects involving outcomes that are not traded in markets (such as "improved well-being"). Special methods are available, however, to value such outcomes, including *contingent valuation* and *willingness* to pay.

include collecting follow-up data, analyzing cost data, and disseminating the RCT's findings.

There is no point in collecting follow-up data until the effects of the intervention have had enough time to materialize, which in some cases may be before implementation is complete. But it is more typical to wait until implementation has been completed. If implementation has been delayed (which is common), this is likely to delay the collection of follow-up data. Follow-up data may be qualitative and quantitative. If qualitative, it may be useful to collect them prior to fielding a follow-up survey, for two reasons. First, they may indicate whether the expected effects have materialized. Second, they may suggest that some questions should be added to the follow-up questionnaire to learn about unexpected effects.

The research team needs to work closely with the implementing team in analyzing the cost data. Whereas the survey data are usually collected and provided to the research team by a third party, estimates of the intervention's costs are usually based on information from the implementing team. These data will need to be checked and verified, usually in the field, with the assistance of the implementing team. The cost estimates will be used, together with estimates of the average treatment effect, in either cost-effectiveness analysis or cost-benefit analysis (box 5.10)

Ideally, the research team should return to the field to present its findings to project staff and local stakeholders. The research team needs to plan dissemination activities with the implementing team and other stakeholders. The feedback received during dissemination events provides an important input to the impact evaluation, perhaps explaining any surprising findings. Once the dissemination activities have been completed, the research team can finalize its report. To exploit the global *public* good of impact evaluations, dissemination of the report should extend beyond the WEE program supporting the evaluation to include such additional media channels as journals and websites specializing in impact evaluations.

6. MAIN CONCLUSIONS

These Guidelines discuss both traditional M&E and program-level impact evaluation in the context of programs designed to empower women economically. In the case of traditional M&E, which focuses on the linkages between inputs, processes, outputs and direct outcomes, the Guidelines call both for strengthening routine monitoring and for exploiting opportunities for process evaluations, particularly those focusing on the analysis of project-level data in order to increase the efficiency and effectiveness of project implementation. Even the most effective traditional M&E should not absorb more than 3-5% of a project's resources and can be expected to yield resource savings of at least an equivalent magnitude.

- Projects themselves have the main responsibility for implementing traditional M&E effectively, although in some cases they will need the assistance of local third parties to develop and maintain monitoring databases and to conduct process evaluations. The most critical function of traditional M&E is routine monitoring. Common pitfalls in routine monitoring include: (i) attempting to track too many indicators, (ii) selecting indicators that are difficult to measure, (iii) failure to implement the monitoring system as planned, and (iv) failure to use the information provided by the monitoring system to improve implementation.

- These Guidelines recommend that some impact evaluations be conducted, particularly in connection with the piloting of innovative interventions that look particularly promising or for existing interventions that are expected to be scaled up in the future. In such cases these Guidelines recommend that the impact evaluation be based on a randomized controlled trial (RCT). Although recognizing that RCTs can be expensive, they may require considerable time to complete and can affect the way projects are implemented, the Guidelines argue that these costs are justified in some cases. Moreover, the benefits of carefully selected and effective impact evaluations can extend well beyond the program supporting the evaluation by contributing to the rapidly growing knowledge base on which interventions are most effective and cost-effective in achieving the program's goals in different settings. The *public good* nature of impact evaluation provides a rationale for its support, much as it provides a rationale for the *performance evaluation* activities such as the Roadmap and Metrics discussed in section 1. Moreover, the costs of impact evaluations can (and should) be shared with other organizations that are committed to developing the knowledge base of effective and cost-effective development interventions.

The Guidelines focus on the points at which an RCT interacts most closely with the intervention being evaluated and that therefore require effective collaboration between the research team and the team implementing the intervention. The most important interaction occurs during the pre-implementation period when it is critical for the research team and the implementing team to work closely and effectively on the design of both the intervention and the RCT. It is also critically important that any implementation activities (including discussions with local officials and targeted beneficiaries) not begin until after baseline data have been collected and until after the study subjects have been randomly assigned to treatment and control groups.

It is very important for the research team to make regular visits to the project sites during implementation to confirm that the monitoring and cost data are available as planned, to analyze the monitoring data to confirm that the intervention is being provided as designed and to work with the implementing team to address any problems that may compromise the RCT.

REFERENCES

KEY SOURCES

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M.J. Vermeersch. 2011. *Impact Evaluation in Practice*. Washington DC: The World Bank.

Glennerster, Rachel and Kudzai Takavarasha. 2013. *Running Randomized Evaluations: A Practical Guide*. Princeton, NJ: Princeton University Press.

Imas, Linda G. M. and Ray C. Rist. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations.* Washington DC: The World Bank.

Rubio, Gloria M. 2012. "Key Choices for a Menu of Evaluations." In Gladys Lopez-Acevedo, Philipp Krause and Keith Mackay, eds, *Building Better Policies: The Nuts and Bolts of Monitoring and Evaluation Systems*. Washington DC: The World Bank.

United Nations Foundation and ExxonMobil Foundation, "Measuring Women's Economic Empowerment," Washington DC (June 2015).

USAID. 2011. *Evaluation: Learning from Experience*. USAID Evaluation Policy (January).

OTHER REFERENCES

Atkinson, Jesse, Alain de Janvry, Craig McIntosh, and Elisabeth Sadoulet. 2013. "Prompting Microfinance Borrowers to Save: A Behavioral Experiment from Guatemala." *Economic Development and Cultural Change* 62(1):21–64.

Bamberger, Michael, Vijayendra Rao, and Michael Woolcock. 2010. "Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development." Policy Research Working Paper 5245. Washington DC: The World Bank (March).

Babson College. 2014. "Investing in the Power of Women: Progress Report on the Goldman Sachs 10,000 Women Initiative." Wellesley, MA: Babson College.

Deaton, Angus S. 2009. "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development." Working Paper 14690, National Bureau of Economic Research, Cambridge, MA (January).

Deaton, Angus and Margaret Grosh. 2000. In Margaret Grosh and Paul Glewwe, eds, *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*. Washington DC: The World Bank.

De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2014. "Business Training and Female Enterprise Start-Up, Growth and Dynamics: Experimental Evidence from Sri Lanka." *Journal of Development Economics* 106: 199–210.

Dhaliwal, Iqbal, Esther Duflo, Rachel Glennerster, and Caitlin Tulloch. 2011. "Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education." Abdul Latif Jameel Poverty Action Lab (August) (www.povertyactionlab.org).

Dos, Cheryl, Zoe Bockius-Suwyn, and Shereen D'Souza. 2012. "Women's Economic Empowerment in Agriculture: Supporting Women Farmers." Report prepared for the UN Foundation (November).

Fairlee, Robert W., Dean Karlan, and Jonathan Zinman. 2013. "Behind the GATE Experiment: Evidence on Effects of and Rationales for Subsidized Entrepreneurship Training." Processed, University of California at Santa Cruz (September).

Filmer, Deon and Kinnon Scott. 2012. "Assessing Asset Indices." *Demography* 49:359–392.

Grosh, Margaret and Paul Glewwe, eds. 2000. *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*. Washington DC: The World Bank.

Johansson de Silva, Sara, Pierella Paci, and Josefina Posadas. 2014. *Lessons Learned and Not Yet Learned from a Multicountry Initiative on Women's Economic Empowerment*. Washington DC: The World Bank.

Karlan, Dean, Margaret McConnell, and Sendhil Mullainathan. 2011. "Getting to the Top of Mind: How Reminders Increase Saving." Department of Economics, Yale University (January).

Kast, Felipe, Stephan Meier, and Dina Pomeranz. 2012. "Under-Savers Anonymous: Evidence on Self-Help Groups and Peer Pressure as a Savings Commitment Device." National Bureau of Economic Research (NBER) Working Paper No. 18417 (September).

Knowles, James C., Anne M. Golla and Helene Carlsson Rex. 2013 "Impact of a Bamboo Handicrafts Intervention in Cambodia and Laos." Processed, Bangkok, Thailand (April).

Kochar, Anjini. 2000. In Margaret Grosh and Paul Glewwe, eds, *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*. 2000. Washington DC: The World Bank.

Mallucio, J.A., J. Hoddinott, J.R. Behrman, R. Martorell, A.R. Quisumbing, and A.D. Stein. 2009. “The Impact of Improving Nutrition During Early Childhood on Education Among Guatemalan Adults.” *Economic Journal* 119 (April):734–763.

Ravallion, Martin. 2012. “Poor, or Just Feeling Poor?: On Using Subjective Data in Measuring Poverty.” Policy Research Working Paper 5968. Washington DC: The World Bank (February).

White, Howard. 2008. “Of Probits and Participation: The Use of Mixed Methods in Quantitative Impact Evaluation.” Network of Networks on Impact Evaluation (NONIE) Working Paper No. 7 (January).

White, Howard. 2009. “Theory-based Impact Evaluation: Principles and Practice.” Working Paper 3, International Initiative for Impact Evaluation (June).

White, Howard. 2010. “A Contribution to Current Debates in Impact Evaluation.” *Evaluation* 16(2):153–164.

Woodruff, Christopher and David McKenzie. 2012. “Improving the Productivity and Earnings of Women Owned and/or Managed Enterprises in Developing Countries: What Works?” Report prepared for the United Nations Foundation (December).

GLOSSARY

The Glossary draws heavily on other sources, including the OECD/DAC *Glossary of Key Terms in Evaluation and Results Based Management* and the glossaries in Glennerster and Takavarasha (2013), Gertler and others (2011), Kusek and Rist (2004), and Imas and Rist (2009). Many of the definitions refer to "interventions." But they also apply to "projects" (consisting of one or more interventions),"programs" (consisting of multiple projects), and "policy changes." Terms in italics are defined elsewhere in the Glossary.

TERM	DEFINITION
Activity	Refers to the actions taken to transform *inputs* into *outputs* (such as training courses)
Anticipation effect	Tendency of people in a *control group* to alter their behavior because they expect to receive an *intervention* in the future
Appraisal	An overall assessment of the *relevance*, feasibility, and potential *sustainability* of an *intervention* prior to a decision on funding (sometimes referred to as a *prospective evaluation* or an *ex-ante evaluation*)
Asset index	A proxy measure of household income that is usually based on housing characteristics and on the ownership of consumer durables
Attribution	Ascribing a *causal link* between observed (or anticipated) changes in outcomes and a specific *intervention*
Association	See *correlation*
Attrition	Inability to collect follow-up data for some of the individuals in the *baseline* sample
Balance	The extent to which key *outcomes* and/or *covariates* vary between *treatment* and *control groups* following *random assignment*
Baseline	The time period prior to *implementation*
Before-and-after design	A nonexperimental design in which the *effects* of an intervention are inferred by comparing baseline and *endline* measurements of *outcomes* (also referred to as a *pre-post design*)
Benchmark	Reference point or standard against which *performance* is assessed, usually based on what has been achieved in the past by other comparable organizations implementing similar *interventions* (sometimes used interchangeably with *target*)
Beneficiaries	The individuals, groups, or organizations, whether targeted or not, that benefit directly or indirectly from an *intervention* (sometimes referred to as an intervention's *reach*)
Benefit-cost analysis	See *cost-benefit analysis*
Benefits	The monetary value of the *effects* of an *intervention*
Bias	The intentional or unintentional distortion of data in collecting, analyzing, or reporting
Biased estimate	Does not converge toward the true value as the *sample* size increases

TERM	DEFINITION
Cardinal data	Indicate relative quantities (in contrast to *ordinal data* that indicate only rank)
Categorical data	Fit into one of multiple nonoverlapping categories (such as sex, religion, country of origin, or housing characteristics)
Causal chain	The causal sequence of an *intervention* that indicates the sequence of steps to achieve the desired *results*, beginning with *inputs*, moving though *activities* and *outputs*, and leading directly to various levels of outcomes (sometimes referred to as a *results framework*)
Causal link	Relationship between two variables in which a change in the value of one results in a change in the value of the other
Census	Data collected for the total *population* (as distinct from a sample *survey*)
Cluster	A group of individuals with some similar characteristics (such as residents of a village or students attending the same school)
Cluster design	An *RCT* in which all individuals within a *cluster* are *randomly assigned* to the *treatment* or *control group*
Comparison group	Used to identify the *counterfactual* in the absence of a *randomly assigned control group* (sometimes used interchangeably with *control group*)
Compliance	Adherence to a planned *treatment*
Composite indicator	Constructed from two or more *indicators*
Control group	People in an experiment who are *randomly assigned* not to receive an *intervention* so that they can be used to identify the *counterfactual* (sometimes referred to as a *comparison group*)
Contamination	Usually refers to *noncompliance* but also sometimes including *spillover* (no longer used widely)
Content analysis	Systematic analysis of *qualitative data*
Contingent valuation	A survey-based economic technique for the valuation of nonmarketed resources, such as environmental preservation or the impact of contamination
Control variable	See *covariate*
Correlation	Tendency of two variables to move together in such a way that a change in one signals a likely change in the other without implying a *causal link* between the two variables (sometimes referred to as an *association*)
Costs	The monetary value of the *inputs* used by an *intervention*
Cost-benefit analysis	Comparison of the *cost* per unit of *benefits* obtained from an *intervention*
Cost-effectiveness analysis	Comparison of the *cost* per unit of *effect* of a set of interventions achieving the same effect
Cost-output analysis	Analysis of the factors related to variations in *unit* costs

TERM	DEFINITION
Counterfactual	The situation or condition that would hypothetically prevail in the absence of an *intervention*.
Covariate	Variable that is used in a regression model to increase the *precision* of an estimate of the *treatment effect* in an RCT (also referred to as a *control variable*)
Demand effects	Tendency of people in an experiment to alter their behavior in response to their perception of what the researchers want to observe
Design	Detailed plan for an *intervention* or an *evaluation*
Development objective	An *outcome* at a higher level even than a *final outcome* (such as reduced poverty or rapid and sustained economic growth). Sometimes referred to as a *goal*.
Direct outcome	*Outcome* (usually observed at the population level) that is directly affected by the *activities* conducted by an *intervention*, such as knowledge and skills acquired or productive assets used (sometimes combined with *outputs*)
Discount rate	A rate (such as 5 percent) used to convert a future value to an equivalent present value (equal to the interest rate when there are no distortions in credit markets)
Effect	Change in an *outcome* that can be directly or indirectly attributed to an *intervention*, whether positive or negative, intended or unintended
Effectiveness	The extent to which an *intervention* achieves its objectives
Efficiency	Whether an *intervention* uses the least-cost combination of *inputs* to achieve its *objectives*
Eligible population	People who are eligible to receive an *intervention* (also referred to as the *target population*)
Endline	The time period after *implementation* ends
Estimate	Statistical term referring to the value of a *statistic* obtained from a *sample*
Estimation	Statistical term referring to the process of obtaining an *estimate*
Evaluation	Systematic and objective assessment of the value of an ongoing or completed project, including its *design*, *implementation* and *results* in order to improve its *effectiveness* or to inform future project *designs* (sometimes referred to as a review)
Evaluation-driven effects	Effects of an experiment due to evaluation-related factors other than the *intervention*
Ex ante evaluation	See *prospective evaluation*
Experiment	See *randomized controlled trial* (sometimes referred to as a *true experiment*)
Experimental design	See *randomized controlled trial*
Experimental unit	Unit (individual or *cluster* of individuals) *randomly assigned* to a *treatment* or *control* group in an *RCT*
External evaluation	An *evaluation* conducted by an organization or individuals with no ties to the organizations funding or implementing the *intervention*

TERM	DEFINITION
Externality	See *spillover*
External validity	The applicability of the *findings* of an *evaluation* to a wider context (such as the entire eligible population in a country or region)
Final outcome	Ultimate *outcome* of an intervention (the last step in the intervention's *causal chain*, also frequently referred to as *impact*). Final outcomes include increases in productivity, income, and well-being
Finding	A factual statement based on evidence obtained from an *evaluation*
First principal component	The linear combination of a *standardized* set of variables that explains the maximum amount of their total variation
Follow-up survey	A survey conducted after *implementation* has begun
Formative evaluation	See *process evaluation*
Goal	See *development objective*
Grant	One of several possible ways to finance a *project* (sometimes used interchangeably with *project*)
Hawthorne effects	Tendency of people in a *treatment group* to try harder
Impact	See *final outcome*
Impact evaluation	Use of analytical methods to obtain estimates of an *intervention's effects*, its effects in relation to its *costs* (*cost-effectiveness analysis*) or its *benefits* in relation to its costs (*benefit-cost analysis*). Sometimes referred to as a *summative evaluation*
Implementation	The segment of the *causal chain* from *inputs* to *outputs* (sometimes including *direct outcomes*)
Implementation monitoring	See *routine monitoring*
Independent evaluation	An *external evaluation* by an organization or individuals free of the control or influence of the organizations funding and implementing the *intervention* being evaluated
Index	Summary measure of a set of related *indicators*, such as the Human Development Index
Indicator	A *quantitative* or *qualitative variable* to measure an *outcome* or other phenomenon (sometimes referred to as a *measure* or *metric*)
Inputs	The financial, human, and material resources used by an *intervention* to produce *outputs*
Institutional review board	A committee that reviews research proposals to ensure that they comply with ethical guidelines
Intermediate outcome	A condition that is susceptive to change in the short- to medium-term and that is believed to be instrumental to achieving *final outcomes* (such as improved business practices or gender empowerment)

TERM	DEFINITION
Internal evaluation	*Evaluation* by the organization funding or implementing the *intervention* (also referred to as *self-evaluation*)
Internal validity	Ability of an *impact evaluation design* to rule out factors other than the *intervention* as a cause of observed *results*
Intervention	A set of *activities* designed to produce one or more targeted *results*
John Henry effects	Tendency of people in a *control group* to try harder
Logical framework	Matrix that links the *activities*, *outputs*, *results*, *purpose*, and *goal* of an *intervention* in a causal hierarchy, and including the assumptions and risks that may affect *performance* (also referred to as a *logframe*)
Logframe	See *logical framework*
Measure	See *indicator*
Mean	A measure of the central tendency of the values in a *population*
Metric	See *indicator*
Measurement error	The difference between the observed value of an *indicator* and the true (unobserved) value
Mid-term evaluation	See *process evaluation*
Milestone	Interim *target*
Monitoring	Continuous tracking and reporting at regular intervals of the *indicators* of an *intervention's* progress in achieving expected *results*
Monitoring of results	See *performance monitoring*
Noncompliance	Failure to follow a *treatment protocol*
Nonresponse	See *attrition*
Observable	Factor for which data are available in a given data set (in contrast to an *unobservable*). Examples are age, sex, and education
Operations research	Advanced analytical methods to improve the *implementation* of an *intervention*
Ordinal data	Data that can be placed on a scale that signifies rank order
Outcome	Benefits (usually measured at a population level but sometimes at an organizational or community level and usually expressed only in broad terms) that are expected to be obtained directly or indirectly from an *intervention*. Most outcomes entail behavioral or organizational change and are typically affected by factors other than an *intervention*
Outputs	The products or services resulting directly from an *intervention* (such as training of potential entrepreneurs or resources delivered)

TERM	DEFINITION
Panel survey	Survey that involves re-interviewing the same sample
Participatory appraisal	An *evaluation* in which *stakeholders* play a key role in its *design* and implementation and in the interpretation of its *findings*
Participatory evaluation	See *participatory appraisal*
Performance	Extent to which an *intervention* achieves its planned results
Performance evaluation	A program-level *evaluation* designed to learn lessons from a completed *project*, including what the project has achieved and not achieved, how it has been implemented, obstacles encountered, and whether expected *results* have been achieved (usually without attempting to attribute the results to the project). Performance evaluations typically apply criteria such as *relevance*, *efficiency*, *effectiveness*, *impact*, and *sustainability*. Sometimes referred to as a *results-based evaluation*
Performance indicator	A measure used to assess whether an *intervention* is achieving its intended *results*
Performance monitoring	Monitoring of both *results* and *implementation* as distinct from *process monitoring* (sometimes referred to as *monitoring of results*)
Phase-in design	An *evaluation design* in which people are *randomly selected* to have access to an *intervention* at different times
Pilot	Small-scale innovative intervention, often including an *impact evaluation* within a fixed *timeframe* with the intention to scale up the *intervention* if the results of the *impact evaluation* are positive
Population	Total number of units from which a *sample* is drawn
Population-level data	Data referring to people and that are usually collected through household surveys. "Population" may refer more narrowly to the population of a limited geographical area (such as one or more districts) or to a population segment targeted by a project (such as women or youth)
Precision	The extent to which the value of an *indicator* varies from one *sample* to another (related to *sampling error*)
Pre-post design	See *before-and-after design*
Primary data	See *project-level data*
Process	The *implementation* segment of a *causal chain*
Process evaluation	*Evaluation* to assess whether an *intervention* is being implemented as planned (also referred to as a *formative evaluation* or *mid-term evaluation*)
Process monitoring	See *routine monitoring*
Program	A group of *projects* with the same or similar *development objective*
Project	A complementary package of *interventions* that are designed to achieve the same or similar *final outcomes*
Project-level data	Data collected by a *project* during *implementation* (also referred to as *primary data*)

TERM	DEFINITION
Prospective evaluation	An *evaluation* that is designed in advance of *implementation*
Proxy indicator	An *indicator* that is highly *correlated* with another indicator and that is used instead of the other indicator because it is easier to measure
Public good	A good or service whose consumption by one person does not reduce the amount available for others
Purpose	See *final outcome*
Qualitative data	Data involving descriptions or subjective assessments, that can be observed but not precisely measured (such as data on relationships and behavior)
Qualitative indicator	An indicator with categorical values (such as satisfaction, ethnicity, sex, or housing characteristics) that may be based on *qualitative data* and that may or may not imply an ordered sequence (a ranking)
Quantitative data	Data that can be precisely measured (such as age, height, sex, or housing characteristics)
Quantitative indicator	An indicator with *cardinal values* (such as income or age) or categorical values (such as sex or housing characteristics)
Quasi-experimental design	An experimental design in which a *comparison group* is nonrandomly selected or constructed after the fact to provide an estimate of the *counterfactual*
Random assignment	Placement of units in *treatment* or *control groups* on the basis of a *random ordering*
Random order	An ordering in which each item bears no relationship to the items that either precede or follow it
Randomized controlled trial (RCT)	An experimental *design* in which beneficiaries are *assigned randomly* to *treatment* and *control groups*. Considered to provide the most reliable estimates of the *counterfactual* when implemented correctly
Random selection	Selection of items from a list of items in *random order*
Randomized trial (RT)	An experimental design in which *subjects* are *assigned randomly* to two or more *interventions* without any *control group*
Randomize	To arrange items in *random order*
Random sample	A sample in which every member of the parent *population* has a known nonzero chance of being selected
Rapid assessment techniques	Simple, quick, and less costly informal techniques for the collection of both qualitative and quantitative data, such as community meetings, self-directed focus groups, popular theater, community radio, wealth rankings, and mini surveys
Reach	See *beneficiaries*
Recall data	Data collected by asking people to provide information about past events or conditions existing during a previous period of time

TERM	DEFINITION
Relevance	Extent to which an *intervention* addresses an important need and is consistent with *program* priorities. Also can refer to the extent to which an *indicator* measures something important
Reliability	Consistency and dependability in an indicator; that is, whether the same value is obtained in repeated measurements
Resentment and demoralization effects	Tendency of people in an experiment (usually in a control group) to alter their behavior because they are denied access to the *intervention*
Respondent	A person interviewed in a *survey*
Response rate	Percentage of the intended *sample* for which data are actually collected
Result	Term usually referring to both the *outputs* and *outcomes* in an *intervention's causal chain* (although some evaluators do not include *outputs* among results, while others use "result" for all levels of the *results chain*)
Results chain	See *causal chain*
Results framework	See *causal chain*
Results-based evaluation	See *performance evaluation*
Results-based monitoring and evaluation	See *performance monitoring*
Retrospective evaluation	An evaluation designed after implementation has ended
Routine monitoring	See *process monitoring*
Sample	Subset of a *population* for which data are collected
Sample mean	The sum of the observations in a sample divided by the number of observations. A measure of the central tendency of values in a sample, an *unbiased estimate* of the *population mean*
Sample size	Number of observations in a *sample*
Sample standard deviation	A measure of the average dispersion of sample values from the *sample mean*
Sampling error	The size of the error caused when using a *sample* to estimate a *population* characteristic (such as the mean) instead of a *census*
Secondary data	Data collected other than by an implementing organization
Selection bias	Distortion of evidence about the effects of an *intervention* due to systematic differences in the characteristics of the group of people receiving and intervention and those not receiving the intervention
Self-evaluation	See *internal evaluation*

TERM	DEFINITION
Self-reported data	Data based on responses to direct questions (what was your agricultural income last month?) instead of being obtained indirectly from detailed and presumably more reliable information (such as data on production of individual crops and related expenses)
Spillover	Occurs when the *effects* of an *intervention* diffuse to members of the *control group* (sometimes referred to as externalities)
Stakeholders	Agencies, organizations, groups, or individuals (including *beneficiaries*) who have a direct or indirect interest in an *intervention*
Standard	Values of a set of related *indicators*, *benchmarks*, or *indices* that indicate minimally accepted levels of *performance* in a given context
Standard deviation	A measure of the average dispersion of values around the *mean* of a *population*
Standardized value	Values ranging between zero and one obtained by subtracting the population *mean* or the *sample mean* from the original variable and dividing by the population *standard deviation* or the *sample standard deviation*
Statistic	Statistical term referring to a measure calculated from *sample* data (such as *sample mean*)
Statistical power	The ability of an *experiment* to detect an *effect* when in fact there is one
Statistically significant	An estimate that is inconsistent with a given hypothesis (such as when an intervention has no effect on an outcome) with a stipulated level of probability
Summative evaluation	See *impact evaluation*
Survey effects	Any change in behavior in a *treatment* or *control group* as the result of being *surveyed*
Sustainability	Continuation of the *benefits* of an *intervention* after it has been completed and funding terminates.
Target	A specified amount of change in an *indicator* that is to be achieved over a specified time frame (also referred to as a *benchmark*)
Target population	See *eligible population*
Theory-based evaluation	Evaluation that assumes the existence of the *causal links* displayed in a *causal chain* and their underlying *theory of change*
Theory of change	Theory describing how an intervention leads to desired *results*
Third party	Organization or individual without any formal ties to organizations funding or implementing an *intervention*
Timeframe	The period of time and frequency for measuring and reporting indicators
Traditional monitoring and evaluation	*Monitoring* and *evaluation* focused on the *implementation segment* of the *causal chain*
Treatment arms	Variations of an *intervention* that are evaluated in an *experiment*

TERM	DEFINITION
Treatment effect	The *effect* of an *intervention* on the *treatment group*
Treatment group	Individuals in an *experiment* whose members have access to the *intervention*
Trend monitoring	Analyzing successive values of an *indicator* to discern trends
Triangulation	Comparing the findings from more than one source of information (such as those from *qualitative* and *quantitative data*) on the same *outcome*
Unit cost	Total cost of an *intervention* divided by the number of units produced or served
Unbiased estimate	An *estimate* that obtains the correct value on average, as when the *sample mean* is an unbiased estimates of the population *mean*
Unobservable	*Variable* for which data are not available in a given data set but that may exert important effects on *outcomes* (such as genetic characteristics, individual tastes, or ambitions)
Validate	To corroborate (or verify) on a sound and authoritative basis
Validity	Extent to which an *indicator* correctly and adequately measures what it is intended to measure
Verifiable	Susceptible to independent validation—say, through remeasurement by another evaluator
Willingness to pay	Maximum amount an individual is willing to sacrifice to procure a good or service or to avoid something undesirable

APPENDIX 1. SUGGESTED QUESTIONNAIRE MODULES FOR MEASURING WEE OUTCOME INDICATORS

I. FINAL OUTCOMES

A. URBAN WOMEN ENTREPRENEURS AND BUSINESS LEADERS

1. BUSINESS INCOME (FINAL OUTCOME)

Two alternative questionnaire modules are suggested for measuring individual business income. Module #1 collects data on business income as part of a set of relatively simple questions collecting data on an individual's income and employment. Module #2 collects more detailed data on business income.

Indicator: Woman's business profits

Definition: Difference between business revenue (Q16 in Module #1) and business costs (Q17 in Module #1) or directly from the response to Q5.7 in Module #2 or indirectly from responses to Q5.5 and Q5.6 in Module #2.

Indicator: Woman's business revenue (sales)

Definition: Q16 in Module #1 or is calculated from responses to Q5.6 in Module #2

MODULE #1

Source: Adapted from Oriana Bandiera, "Evaluating skills and capital transfers programs targeted to women not in stable employment (young and/or ultrapoor)," Note prepared for Metrics Meeting (April 2014)

	QUESTION	RESPONSE
Q1	Are you engaged in any income generating activity (IGA)?	Yes.....1 No.....2 > Next module
Q2	How many months of the year are you engaged in all of your IGAs?	(Months)
Q3	In a typical month, how many days do you work in all of your IGAs?	(Days)
Q4	In a typical day, how many hours do you work in all of your IGAs?	(Hours)
Q5	Are you self-employed and/or working for an employer for wages or a salary in your IGAs?	Employer only.....1 Self-employed only.....2 > Q12 Both.....3

	QUESTION	RESPONSE	
Q6–Q11 refer to your main job as a wage or salary worker			
Q6	How many employees are in your employer's firm/ organization?	2-10.....1 11-50.....2 50+.....3	
Q7	How often are you paid?	Daily.....1 Weekly.....2 Every two weeks.....3 Monthly.....4	
Q8	Are you paid in cash or in-kind?	Cash.....5 In-kind.....6 Both.....7	
Q9	How much were you paid the last time you were paid?	(Local currency)	
Q10	How much does this job usually pay?	(Local currency)	
Q11	If you were offered more work at the same daily rate, would you accept?	Yes.....1 No.....2	> Next module if Q5=1 (no self-employment)
Q12–Q19 refer to your main self-employment activity			
Q12	How many female HH workers do you work with?	(Number)	
Q13	How many male HH workers do you work with?	(Number)	
Q14	How many female hired workers do you work with?	(Number)	
Q15	How many male hired workers do you work with?	(Number)	
Q16	What were your business revenues between (state 12 months before interview) and now?	(Local currency)	
Q17	What were your business costs between (state 12 months before interview) and now?	(Local currency)	
Q18	Would you like to expand this business?	Yes.....1 No.....2 > Next module	
Q19	What are you lacking?	Credit.....1 Trustworthy workers...2 Childcare.....3 Time.....4 Other.....5	

MODULE #2

Source: Kenya Female Enterprise Survey (2013), Baseline Questionnaire (version 10) (http://microdata.worldbank.org/index.php/catalog/1985)

5.5 Business expenses during the last month: Please report the amount you have spent on each of the following categories of business expenses during last month

Interviewer: include only business and not household expenses; do not include wages the owner pays herself as an expense

	ITEM	COST (LOCAL CURRENCY UNIT)
1	Purchase of materials and items for resale	
2	Purchase of electricity, water, gas and fuel	
3	Market fee (e.g. for use or city tax on stall space)	
4	Interest paid on loans	
5	Wages and salaries for employees	
6	Rent for land or buildings	
7	Taxes	
8	Other expenses, including equipment rental, telephone, transportation	
9	**Total expenses in the last month**	

5.6 Business sales (revenue):		RESPONSE
a	What were the total sales of your business YESTERDAY? *Write zero if the business was closed yesterday. Include sales on credit and value of any sales in a barter transaction.*	(Local currency)
b	What were the total sales of your business in the LAST WEEK? *Write zero if the business was closed in the last week.*	(Local currency)

5.6 Business sales (revenue) Continued:		
c	In a typical week in the last month, can you tell me approximately how much you would sell on each day of the week? *[Read options. Write zero for days the business is usually closed, 999 for don't know]*	
	DAY OF WEEK	**TYPICAL DAILY SALES (LOCAL CURRENCY)**
i	Monday	
ii	Tuesday	
iii	Wednesday	
iv	Thursday	
v	Friday	
vi	Saturday	
vii	Sunday	
d	Now consider the different months of the year. Write 100 in the month or months in which the sales of your business are highest. Then for the other months, write the percentage of the best month's sales that you would typically get in that month. For example, if your best month of sales is February, write 100 in February. Then if you typically sell only half this much in March, write 50 for March. Write zero for months you don't sell anything.	
	MONTH	**TYPICAL MONTHLY SALES AS % OF HIGHEST MONTH**
1	January	
2	February	
3	March	
4	April	
5	May	
6	June	
7	July	
8	August	
9	September	
10	October	
11	November	
12	December	

What was the total income the business earned during last month after paying all expenses including wages of employees, but not including any income you paid yourself. That is, what were the profits of your business during last month? *Note: If you paid yourself a salary, add that back in to your profits.)*	(Local currency) 999. Don't know/ refuse to answer

2. EMPLOYMENT (FINAL OUTCOME)

The three employment indicators can be calculated from responses to questions in Module #1 above.

Indicator: Number of employees in the woman's business
Definition: Sum of responses to Q12-Q15 in Module #1

Indicator: Average monthly hours worked for pay by woman
Definition: Average monthly hours worked for pay (AMHW) can be calculated from the responses to Q2-Q4 in Module #1, i.e., AMHW=Q3*Q4*(Q2/12).

Indicator: Average monthly income earned per hour worked for pay by woman
Definition: Calculated from the responses to Q10, Q16 and Q17 in Module #1 and from AMHW above, i.e., [Q10*(Q2/12)+(Q16-Q17)/12]/AMHW.

B. RURAL WOMEN ENTREPRENEURS AND FARMERS

1. INCOME/EXPENDITURE (FINAL OUTCOME)

Indicator: Total household consumption per capita
Definition: Total household consumption divided by the number of household members. Total household consumption is the sum of three components: (1) food consumption, (2) nonfood consumption and (3) the imputed rental value of consumer durable goods. The first two items include not only purchased consumption items but also items produced at home and items received as gifts. The annual rental value of major consumer durable goods owned by the household is used instead of expenditure on major consumer durable goods because the latter might result in an upward bias of the effect of an intervention on household consumption if it led to the purchase of an expensive consumer durable (e.g., a motorbike).

The accuracy of data on household consumption is clearly related to the degree of detail in the definition in the consumption categories, i.e., use of more categories is generally believed to result in more accurate estimates of consumption. The food and nonfood consumption modules presented below are adapted from the 2009 Cambodia Socio-Economic Survey and were used successfully during the period 2004-2011 to obtain poverty estimates. In these modules, the individual consumption items are grouped into 20 food groups and 11 nonfood groups,

perhaps the least number of consumption categories that can be used to obtain reasonably accurate estimates of total household consumption.

Indicator: Household consumption per capita of selected items
Definition: Sum of household consumption per capita limited to items that are believed to be sensitive to income change and that can therefore serve as good proxy indicators of household income change. Examples that are often used for this purpose include consumption of fish, meat and poultry, fresh vegetables, fruit and purchased meals (i.e. food groups 2, 3, 7, 11, 19 and 20 in the following food consumption module).

FOOD CONSUMPTION

Source: Adapted from Cambodia-Economic Survey (CSES) 2009 "Household Questionnaire"

Respondent: The household member who knows most about food, beverage, and tobacco consumption during the past 7 days

Note: any household expenditure for business purposes should not be reported below

	For each item group try to estimate quantity of items consumed and then how much of the consumed quantity had been purchased in cash and how much was from own production or received as payment in kind for work, or as gift, or free collection.	VALUE OF CONSUMPTION IN LOCAL CURRENCY WRITE "0" IF NOTHING		
		Q1. CASH EXPENDITURE	Q2. OWN PRODUCTION, WAGES IN KIND, GIFTS, FREE DISTRIBUTION (IMPUTED VALUE)	Q3. TOTAL CONSUMPTION (COL. 3 + COL. 4)
	FOOD/BVERAGE/TOBACCO ITEMS	LOCAL CURRENCY	LOCAL CURRENCY	LOCAL CURRENCY
1	2	3	4	5
1	Cereals (rice, bread, corn, wheat flour, rice flour, corn meal, rice cakes, noodles, biscuits, etc.)			
2	Fish (fresh fish, salted and dried fish, canned fish, shrimp, prawn, crab, etc.)			
3	Meat & poultry (beef, buffalo, mutton, lamb, pork, chicken, duck, innards, incl. liver, spleen, dried beef, etc.)			
4	Eggs (chicken egg, duck egg, quail egg, fermented/salted egg, etc.)			
5	Dairy products (fresh milk, condensed or powdered milk, ice cream, cheese, other dairy products, etc.)			

	For each item group try to estimate quantity of items consumed and then how much of the consumed quantity had been purchased in cash and how much was from own production or received as payment in kind for work, or as gift, or free collection.	VALUE OF CONSUMPTION IN LOCAL CURRENCY WRITE "0" IF NOTHING		
		Q1. CASH EXPENDITURE	Q2. OWN PRODUCTION, WAGES IN KIND, GIFTS, FREE DISTRIBUTION (IMPUTED VALUE)	Q3. TOTAL CONSUMPTION (COL. 3 + COL. 4)
	FOOD/BVERAGE/TOBACCO ITEMS	LOCAL CURRENCY	LOCAL CURRENCY	LOCAL CURRENCY
1	2	3	4	5
6	Oil and fats (rice bran oil, vegetable oil, pork fat, butter, margarine, coconut/frying oil, etc.)			
7	Fresh vegetables (onion, shallot, cabbage, spinach, carrots, beans, chilies, tomatoes, etc.)			
8	Tuber (cassava, sweet potato, potato, sugar beet, etc.)			
9	Pulses and legumes (green gram, cowpea, bean sprout, other seeds, etc.)			
10	Prepared and preserved vegetables (cucumber pickles, other pickles, tomato paste, etc.)			
11	Fruit (banana, orange, mango, pineapple, lemon, papaya, water melon, grape, apple, canned and dried fruits, etc.)			
12	Dried nuts and edible seeds (coconut, cashew nut, lotus nut, peanut, gourd seed, other nuts, etc.)			
13	Sugar, salt and spices (sugar, jaggery, salt, chocolate, candy, coriander, red pepper spice, garlic, ginger, soy sauce, fish sauce, monosodium glutamate, etc.)			
14	Tea, coffee, cocoa			
15	Non-alcoholic beverages (canned or bottled soft drinks, mineral water, fruit juice, fruit syrup, etc.)			
16	Alcoholic beverages (beer, wine, whisky, other distilled spirits, etc.)			
17	Tobacco products (cigarettes, mild tobacco, strong tobacco, etc.)			
18	Other food products (ice, flavored ice, other food products, etc.)			

	For each item group try to estimate quantity of items consumed and then how much of the consumed quantity had been purchased in cash and how much was from own production or received as payment in kind for work, or as gift, or free collection.	VALUE OF CONSUMPTION IN LOCAL CURRENCY WRITE "0" IF NOTHING		
		Q1. CASH EXPENDITURE	Q2. OWN PRODUCTION, WAGES IN KIND, GIFTS, FREE DISTRIBUTION (IMPUTED VALUE)	Q3. TOTAL CONSUMPTION (COL. 3 + COL. 4)
	FOOD/BVERAGE/TOBACCO ITEMS	LOCAL CURRENCY	LOCAL CURRENCY	LOCAL CURRENCY
1	2	3	4	5
19	Food taken away from home (meals at work, school, restaurants, snacks, coffee, soft drinks purchased outside home)			
20	Prepared meals bought outside and eaten at home			

NONFOOD CONSUMPTION (EXCLUDING MAJOR CONSUMER DURABLES)

Source: Adapted from Cambodia-Economic Survey (CSES) 2009 "Household Questionnaire"

Respondent: The household member who know most about household non-food expenditure. Include only expenditure for household consumption

Note: any household expenditure for business purposes should not be reported below

	What was your household's consumption of the following items during the indicated time periods?		VALUE (IN LOCAL CURRENCY) WRITE "0" IF NOTHING		
			Q4. CASH EXPEND-ITURE	Q5. OWN PRODUC-TION, WAGES IN KIND, GIFTS, FREE DISTRIBUTION (IMPUTED VALUE)	Q6. TOTAL CONSUMPTION (COL. 4 + COL. 5)
	NON-FOOD ITEMS	TIME PERIOD	LOCAL CURRENCY	LOCAL CURRENCY	LOCAL CURRENCY
1	2	3	4	5	6
1	Housing (house rent, rental value of rent-free housing, rental value of owner-occupied housing, house maintenance and minor repairs) NOT INCLUDING THE COST OF MAJOR REPAIRS OR COST OF NEW HOUSING CONSTRUCTION	Last 1 month			

	What was your household's consumption of the following items during the indicated time periods?		VALUE (IN LOCAL CURRENCY) WRITE "0" IF NOTHING		
			Q4. CASH EXPENDITURE	Q5. OWN PRODUCTION, WAGES IN KIND, GIFTS, FREE DISTRIBUTION (IMPUTED VALUE)	Q6. TOTAL CONSUMPTION (COL. 4 + COL. 5)
	NON-FOOD ITEMS	TIME PERIOD	LOCAL CURRENCY	LOCAL CURRENCY	LOCAL CURRENCY
1	2	3	4	5	6
2	Utilities and fuel (water charges, sewage or waste water disposal, garbage collection, electricity, gas (LPG), kerosene, firewood, charcoal, batteries)	Last 1 month			
3	Medical care (doctors' fees, other medical services, drugs, hospital charges, other medical supplies, etc.)	Last 1 month			
4	Transportation (operation of transport equipment, maintenance and repair of equipment, gasoline and diesel for own transportation, fees for public transport, moving fee, driving lessons, etc.) NOT INCLUDING THE PURCHASE PRICE OF MAJOR PERSONAL TRANSPORTATION EQUIPMENT (e.g., car, truck, motorbike, bicycle)	Last 1 month			
5	Communications (postage stamps, fax, telephone and internet phone charges, phone cards, internet charges etc.) NOT INCLUDING THE PURCHASE PRICE OF MAJOR COMMUNICATIONS EQUIPMENT (computers, laptops, tablets, cell phones)	Last 1 month			
6	Personal care (soap, toothpaste, razor, sanitary napkins, haircut, manicure, etc.)	Last 1 month			
7	Clothing and footwear (tailored clothes, ready-made clothes, rain clothes, underwear, baby clothes, diapers, hats, shoes, boots, etc.)	Last 6 months			

	What was your household's consumption of the following items during the indicated time periods?		VALUE (IN LOCAL CURRENCY) WRITE "0" IF NOTHING		
			Q4. CASH EXPENDITURE	Q5. OWN PRODUCTION, WAGES IN KIND, GIFTS, FREE DISTRIBUTION (IMPUTED VALUE)	Q6. TOTAL CONSUMPTION (COL. 4 + COL. 5)
	NON-FOOD ITEMS	TIME PERIOD	LOCAL CURRENCY	LOCAL CURRENCY	LOCAL CURRENCY
1	2	3	4	5	6
8	Furniture, furnishings and household equipment and operation (curtain, household appliances, cooking utensils, light bulbs, soap and detergents, domestic salaries.) DO NOT INCLUE THE PURCHASE OF MAJOR APPLIANCES (stove, refrigerator, washing machine, air conditioner, sewing machine)	Last 12 months			
9	Recreation (entertainment services, recreational goods and supplies, tourist travel, hotel accommodation, gambling) DO NOT INCLUDE THE PURCHASE OF MAJOR ELECTRONIC EQUIPMENT (TV, stereo, VCR/DVD player, camera, video camera)	Last 12 months			
10	Education (school fees, textbooks, private tutoring charges, etc.)	Last 12 months			
11	Personal effects (costume/gold jewelery, handbags, wallets, wristwatch, clocks, umbrella)	Last 12 months			
12	Miscellaneous items (special occasions such as funeral rituals, weddings, parties, cash gifts, charity, etc.)	Last 12 months			

CONSUMER DURABLES

The information on the ownership of consumer durables is obtained in order to include an imputed annual rental value of consumer durables in household nonfood consumption. This includes two sub-components: (1) the opportunity cost of the capital tied up in ownership of the good (based on the response to Q7), and (2) the annual physical depreciation of the good (based on responses to Q3-Q6). An estimate of the average useful life of each good can be obtained by doubling the reported average age of the goods owned by the sample households.

Source: Adapted from Margaret Grosh and Paul Glewwe, Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study. Washington DC: The World Bank (2000).

Note: the following list of consumer durables may need to be adapted to reflect local circumstances.

Q1 Does your household own any of the following items?

Determine which durables the household owns by asking Q1. For each durable owned, write the description and code provided under Q2 and proceed to ask Q3-Q7 for each item.

ITEM	CODE	YES	NO
Stove	201		
Refrigerator	202		
Freezer	203		
Washing machine	204		
Sewing machine	205		
Air conditioner	206		
Computer, laptop, tablet	207		
Cell phone, smart phone	208		
Television	209		
VCR/DVD Player	210		
Camera	211		
Video camera	212		
Bicycle	213		
Motorcycle	214		
Car, truck	215		
[country-specific item #1, etc]	216		

	Q2. LIST ALLTHE ITEMS OWNED BY THE HOUSEHOLD, THEN PROCEED TO ASK Q.3		Q3. HOW MANY YEARS AGO DID YOU ACQUIRE THIS (ITEM)?	Q4. DID YOU PURCHASE IT OR RECEIVE IT AS A GIFT OR PAYMENT FOR SERVICES? PURCHASE...1 GIFT OR PAYMENT...2 (>>Q.6)	Q5. HOW MUCH DID YOU PAY FOR IT? (>>Q.7)	Q6. HOW MUCH WAS IT WORTH WHEN YOU RECEIVED IT?	Q7. IF YOU WANTED TO SELL THIS [ITEM] TODAY, HOW MUCH WOULD YOU RECEIVE?
	DESCRIPTION	CODE	YEARS		LOCAL CURRENCY	LOCAL CURRENCY	LOCAL CURRENCY
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							

Indicator: Household asset index

The following questionnaire module collects data that can be used to construct an asset index. The responses to most questions are typically used to define a large set of 0-1 indicators referring to specific housing characteristics and to the ownership of specific physical assets. The asset index is usually constructed as the first principal component of the full set of such indicators.

Source: Household Questionnaire, Multiple Indicator Cluster Survey (MICS), UNICEF (October 2013)

	QUESTION		
HC2	How many rooms in this household are used for sleeping?	Number of rooms	__ __
HC3	Main material of the dwelling floor. *Record observation*	Natural floor	
		Earth / Sand	11
		Dung	12
		Rudimentary floor	
		Wood planks	21
		Palm / Bamboo	22
		Finished floor	
		Parquet or polished wood	31
		Vinyl or asphalt strips	32
		Ceramic tiles	33
		Cement	34
		Carpet	35
		Other (*specify*) _____	96
HC4	Main material of the roof. *Record observation.*	Natural roofing	
		No Roof	11
		Thatch / Palm leaf	12
		Sod	13
		Rudimentary roofing	
		Rustic mat	21
		Palm / Bamboo	22
		Wood planks	23
		Cardboard	24
		Finished roofing	
		Metal / Tin	31
		Wood	32
		Calamine / Cement fibre	33
		Ceramic tiles	34
		Cement	35
		Roofing shingles	36
		Other (*specify*) _____	96

	QUESTION			
HC5	Main material of the exterior walls. *Record observation.*	Natural walls		
		No walls	11	
		Cane / Palm / Trunks	12	
		Dirt	13	
		Rudimentary walls		
		Bamboo with mud	21	
		Stone with mud	22	
		Uncovered adobe	23	
		Plywood	24	
		Cardboard	25	
		Reused wood	26	
		Finished walls		
		Cement	31	
		Stone with lime / cement	32	
		Bricks	33	
		Cement blocks	34	
		Covered adobe	35	
		Wood planks / shingles	36	
		Other (*specify*) _____	96	
HC6	What type of fuel does your household mainly use for cooking?	Electricity	01	01>HC8
		Liquefied Petroleum Gas (LPG)	02	02>HC8
		Natural gas	03	03>HC8
		Biogas	04	04>HC8
		Kerosene	05	05>HC8
		Coal / Lignite	06	
		Charcoal	07	
		Wood	08	
		Straw / Shrubs / Grass	09	
		Animal dung	10	
		Agricultural crop residue	11	
		No food cooked in household	95	95>HC8
		Other (*specify*) _____	96	
HC7	Is the cooking usually done in the house, in a separate building, or outdoors? If 'in the house', probe: is it done in a separate room used as a kitchen?	In the house		
		In a separate room used as kitchen	1	
		Elsewhere in the house	2	
		In a separate building	3	
		Outdoors	4	
		Other (*specify*) _____	6	

	QUESTION				
	Does your household have:		N	Y	
HC8	[A] electricity?	Electricity	1	2	
	[B] a radio?	Radio	1	2	
	[C] a television?	Television	1	2	
	[D] a non-mobile telephone?	Non-mobile telephone (fixed line telephone)	1	2	
	[E] a refrigerator?	Refrigerator	1	2	
	[F] country specific items (*Add as necessary*)	Country Specific Item	1	2	
	Does any member of your household own:		N	Y	
HC9	[A] a watch?	Watch	1	2	
	[B] a mobile telephone?	Mobile telephone	1	2	
	[C] a bicycle?	Bicycle	1	2	
	[D] a motorcycle or scooter?	Motorcycle / Scooter	1	2	
	[E] an animal-drawn cart?	Animal-drawn cart	1	2	
	[F] a car or truck?	Car / Truck	1	2	
	[G] a boat with a motor?	Boat with motor	1	2	
	[H] country specific items (*Add as necessary*)	Country Specific Item	1	2	
HC10	Do you or someone living in this household own this dwelling? *If the response is "rent", then ask: do you rent this dwelling from someone not living in this household?* *If "rented from someone else", circle "2". For other responses, circle "6".*	Own Rent Other (*specify*) _____	1 2 6		
HC11	Does any member of this household own any land that can be used for agriculture?	Yes No	1 2		2>HC13

	QUESTION			
HC12	How many hectares of agricultural land do members of this household own? *If less than 1, record "00". If 95 or more, record "95". If unknown, record "98".*	Hectares	__ __	
HC13	Does this household own any livestock, herds, other farm animals, or poultry?	Yes No	1 2	2>HC15
HC14	How many of the following animals does this household have? [A] cattle, milk cows, or bulls? [B] horses, donkeys, or mules? [C] goats? [D] sheep? [E] chickens? [F] pigs? [G] country specific additions (*Add as necessary*) *If none, record "00". If 95 or more, record "95".* *If unknown, record "98".*	 Cattle, milk cows, or bulls Horses, donkeys, or mules Goats Sheep Chickens Pigs Country Specific Addition	 ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___	
HC15	Does any member of this household have a bank account?	Yes No	1 2	

Indicator: Household savings

Definition: The indicator is defined as the sum of the current value of all household assets at a given point in time, including: (1) the household's dwelling, (2) business assets, (3) agricultural assets, (4) credit, (5) other real estate and financial assets, and (6) consumer durables, less the value of household debt from responses to Q9. Estimates at two different points of time (for example, at baseline and endline) can be used to obtain an estimate of household savings during the intervening period.

Sources: adapted from Margaret Grosh and Paul Glewwe, Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study. Washington DC: The World Bank (2000).

DWELLING

Q1. PLEASE TELL ME ABOUT THE DWELLING CURRENTLY OCCUPIED BY YOUR HOUSEHOLD.

	QUESTION	RESPONSE	
Q1_1	Is this dwelling owned by a member of your household?	Yes.....1 No.....2	>Q2
Q1_2	Do you have legal title to the dwelling or any other document that shows ownership?	Yes.....1 No.....2	
Q1_3	Do you have legal title to the land on which the dwelling is located or any other document that shows ownership?	Yes.....1 No.....2	
Q1_4	Who is the legal owner(s) of this dwelling?	Woman only.....1 Spouse only.....2 Woman and spouse jointly.....3 Other person(s).....4	
Q1_5	Could the owner(s) sell this dwelling if they wanted to?	Yes.....1 No.....2	>Q2
Q1_6	If you sold this dwelling today how much would you receive for it?	_____ (*local currency*) Don't know.....999	

BUSINESS ASSETS

Q2. PLEASE TELL ME ABOUT THE INVENTORY BELONGING TO ALL BUSINESSES OWNED WHOLLY OR PARTLY BY THIS HOUSEHOLD.

	QUESTION	RESPONSE	
Q2_1	Do any of the businesses owned by this household have an inventory of raw materials, items requiring further processing, or finished products?	Yes.....1 No.....2	>Q3
Q2_2	What is the total current value of the inventory of all these businesses?	_____ (*local currency*) Don't know.....999	
Q2_3	Is all of this inventory owned by household members?	Yes.....1 No.....2	>Q3
Q2_4	How much of this inventory value belongs to members of this household?	_____ (*local currency*) Don't know.....999	

Q3.PLEASE TELL ME ABOUT THE ASSETS BELONGING TO ALL BUSINESSES OWNED WHOLLY OR PARTLY BY THIS HOUSEHOLD.

NOTE: INCLUDE ONLY ITEMS USED MAINLY FOR BUSINESS PURPOSES

	ITEM	Q3.1. DO ANY OF THE BUSINESSES OWNED BY THIS HOUSEHOLD USE ANY OF THE FOLLOWING ASSETS? YES.....1 NO.....2 >NEXT ITEM	Q3.2. IF THE OWNER(S) OF THESE ASSETS WANTED TO SELL THEM, HOW MUCH WOULD THEY GET TODAY? (LOCAL CURRENCY) DON'T KNOW.....999	Q3.3. HOW MUCH OF THIS WOULD BELONG TO MEMBERS OF THIS HOUSEHOLD? (LOCAL CURRENCY) DON'T KNOW.....999
Q3_1	Land			
Q3_2	Buildings			
Q3_3	Equipment and machinery			
Q3_4	Furniture			
Q3_5	Small or large tools			
Q3_6	Large vehicles (trucks, cars, boats, etc.)			
Q3_7	Small vehicles (bicycles, carts, etc.)			
Q3_8	Other durable goods			

AGRICULTURAL ASSETS

Q4. PLEASE TELL ME ABOUT EACH PLOT OF LAND BELONGING WHOLLY OR PARTLY TO YOUR HOUSEHOLD THAT HAS BEEN CULTIVATED OR USED FOR OTHER AGRICULTURAL PURPOSES BY A MEMBER OF YOUR HOUSEHOLD OR THAT HAS BEEN RENTED OUT TO OTHER HOUSEHOLDS DURING THE LAST TWO CROPPING SEASONS.

	Q4.1. WHAT IS THE AREA OF THIS PLOT? UNIT CODES: SQUARE METERS.....1 HECTARES.....2 [LOCAL UNIT].....3 AMOUNT UNIT CODE	Q4.2. COULD THE OWNER(S) OF THIS PLOT SELL IT IF THEY WANTED TO?	Q4.3. IF THE OWNER(S) OF THIS PLOT WANTED TO SELL IT, HOW MUCH WOULD THEY GET TODAY? (LOCAL CURRENCY)	Q4.4. HOW MUCH OF THIS WOULD BELONG TO MEMBERS OF THIS HOUSEHOLD? (LOCAL CURRENCY)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				

Q5. PLEASE TELL ME ABOUT THE AGRICULTURAL EQUIPMENT OWNED BY YOUR HOUSEHOLD.

	ITEM	Q5.1. DOES THIS HOUSEHOLD OWN ANY OF THE FOLLOWING ITEMS OF AGRICULTURAL EQUIPMENT? YES.....1 NO.....2 >NEXT ITEM	Q5.2. IF THE OWNER(S) OF THIS EQUIPMENT WANTED TO SELL IT, HOW MUCH WOULD THEY GET TODAY? (LOCAL CURRENCY)	Q5.3. HOW MUCH OF THIS WOULD BELONG TO MEMBERS OF THIS HOUSEHOLD? (LOCAL CURRENCY)
1	Large tractor (>=12 horse power)			
2	Small tractor (<12 horse power			
3	Machined pulled plough or harrower			
4	Animal pulled plow			
5	Mechanical water pump			
6	Sprinkler			
7	Motorized thresher			
8	Hand thresher			
9	Rice winnower			
10	Mill			
11	Machine to process livestock feed			
12	Motorized insecticide pump			
13	Ox cart			
14	Small cart pulled by person			
15	Fish pond			
16	Fishing boat			

Q6. PLEASE TELL ME ABOUT THE DIFFERENT TYPES OF LIVESTOCK OWNED BY YOUR HOUSEHOLD.

	LIVESTOCK TYPE	Q6.1. DOES THIS HOUSEHOLD CURRENTLY OWN ANY OF THE FOLLOWING TYPES OF LIVESTOCK? YES.....1 NO.....2 >NEXT TYPE	Q6.2.HOW MANY OF THIS TYPE OF LIVESTOCK ARE CURRENTLY OWNED BY MEMBERS OF THIS HOUSEHOLD? (NUMBER)	Q6.3. IF THE OWNER(S) OF THESE LIVESTOCK WANTED TO SELL ONE OF THEM TODAY, HOW MUCH WOULD THEY RECEIVE? (LOCAL CURRENCY)
1	Beef cattle			
2	Milk cows			
3	Breeding bulls			
4	Horses			
5	Donkeys/mules			
6	Pigs for breeding			
7	Sows			
8	Sheep			
9	Goats			
10	Chickens			
11	Ducks			
12	Other poultry			
13	Rabbits			
14	Bees			
15	Other (specify_____)			

Q7. PLEASE TELL ME ABOUT THE HAND TOOLS OWNED BY YOUR HOUSEHOLD.

	HAND TOOL	Q7.1. DOES THIS HOUSEHOLD OWN ANY OF THE FOLLOWING HAND TOOLS? YES.....1 NO.....2 >NEXT TOOL	Q7.2.IF THE OWNER(S) OF TOOLS WANTED TO SELL THEM, HOW MUCH WOULD THEY GET TODAY? (LOCAL CURRENCY)
1	Hoes		
2	Knives		
3	Axes		
4	Rakes		
5	Shovels		
6	Picks		
7	Sickles/Reaping hooks		
8	Fishing nets		

CREDIT

Q8. PLEASE TELL ME ABOUT ANY MONEY THAT IS CURRENTLY OWED BY NON-HOUSEHOLD MEMBERS TO MEMBERS OF YOUR HOUSEHOLD

	QUESTION	RESPONSE	
Q8_1	Does any non-household member owe money to one or more household members at this time?	Yes.....1 No.....2	>Q9
Q8_2	How many people currently owe money to members of this household?	— —	
Q8_3	How much in total is currently owed to members of this household?	_____ (*local currency*) Don't know.....999	

Q9. PLEASE TELL ME ABOUT ANY LOANS CURRENTLY OWED BY HOUSEHOLD MEMBERS TO NON-HOUSEHOLD MEMBERS OR OTHER EXTERNAL SOURCES.

	SOURCE	Q9.1. DO ANY HOUSEHOLD MEMBERS CURRENTLY HAVE OUTSTANDING LOANS BORROWED FROM THIS SOURCE? YES.....1 NO.....2 >NEXT SOURCE	Q9.2. HOW MUCH IS CURRENTLY OWED TO THIS SOURCE? (LOCAL CURRENCY) DON'T KNOW.....999
1	Family member, friend or other non-household member		
2	Employer or landlord		
3	Credit union, cooperative or NGO		
4	Bank or government agency		
5	Microfinance institution		
6	Money lenders/pawn shop		
7	Other source (specify_____)		

OTHER REAL ESTATE AND FINANCIAL ASSETS

Q10. PLEASE TELL ME ABOUT OTHER REAL ESTATE AND FINANCIAL ASSETS OWNED BY MEMBERS OF YOUR HOUSEHOLD.

	QUESTION	RESPONSE	
Q10_1	Does any member of your household own any land or buildings which you rent to others for residential or business purposes (do not include property used by the household for residential, farming or business purposes)?	Yes.....1 No.....2	>Q10_3
Q10_2	How much money would your household receive if it sold all this property today?	_____ (*local currency*) Don't know.....999	
Q10_3	Does any member of your household have investments in stocks, bonds or life insurance?	Yes.....1 No.....2	> Q10_5
Q10_4	What is the approximate current value of these assets?	_____ (*local currency*) Don't know.....999	
Q10_5	Does any member of your household have savings or other assets in a financial institution (including microfinance institutions)?	Yes.....1 No.....2	> Q10_7
Q10_6	What is the current value of these savings or other assets?	_____ (*local currency*) Don't know.....999	

Q10. PLEASE TELL ME ABOUT OTHER REAL ESTATE AND FINANCIAL ASSETS OWNED BY MEMBERS OF YOUR HOUSEHOLD.

	QUESTION	RESPONSE	
Q10_7	Is any member of your household currently participating in any informal savings association such as a [GIVE LOCAL NAME]?	Yes.....1 No.....2	>Q10_10
Q10_8	How much have household members contributed to this association since joining it?	_____ (*local currency*) Don't know.....999	
Q10_9	How much have household members received from the association since joining it?	_____ (*local currency*) Don't know.....999	
Q10_10	How much cash on hand does your household have currently (including the cash value of any gold, jewelry or other valuables)?	_____ (*local currency*) Don't know.....999	

CONSUMER DURABLES

Use the module on consumer durables in I.B.1 (questions Q1, Q2 and Q7 only)

C. BOTH URBAN AND RURAL WOMEN

1. INDIVIDUAL ASSETS (FINAL OUTCOME)

Indicator: Net value of woman's financial assets

Definition: The current value of the female respondent's share of household financial assets (Q3 below) less her share of household liabilities (Q8).

Source: Adapted from the Gender Asset Gap Project in Ecuador, Ghana and India (2009). The project team leaders are Hema Swaminathan, Indian Institute of Management Bangalore (IIMB); Abena D. Oduro, University of Ghana; Carmen Diana Deere, University of Florida; Cheryl Doss, Yale University; and Caren Grown, American University.

FINANCIAL ASSETS

Q5. PLEASE TELL ME ABOUT THE AGRICULTURAL EQUIPMENT OWNED BY YOUR HOUSEHOLD.

	FINANCIAL ASSETS CODE (READ OUT LIST)	Q1. DO YOU YOURSELF INDIVIDUALLY OR JOINTLY WITH SOMEONE ELSE HAVE ANY MONEY IN [FINANCIAL ASSET]? 1. YES, INDIVIDUALLY 2. YES, JOINTLY 3. NO >SKIP TO NEXT ASSET 98 DOES NOT KNOW 99. REFUSED TO RESPOND >SKIP TO NEXT ASSET (CIRCLE RESPONSE)	Q2. WHOSE NAME IS ON THE ACCOUNT? 1. YOURSELF ONLY 2. SPOUSE ONLY 3. YOURSELF & SPOUSE 4.YOURSELF & OTHER HOUSEHOLD MEMBER(S) 5. YOURSELF & OTHER NON-HOUSEHOLD MEMBER(S) 6. OTHER HOUSEHOLD MEMBERS ONLY 98 DOES NOT KNOW 99 REFUSED TO RESPOND (CIRCLE RESPONSE)	Q3. WHAT IS THE APPROXIMATE CURRENT VALUE OF YOUR SHARE IN THIS ACCOUNT OR SOURCE OF SAVINGS CURRENTLY? (LOCAL CURRENCY) 98 DOES NOT KNOW 99 REFUSED TO RESPOND
1	Savings Account in a bank	1 2 3 98 99	1 2 3 4 5 6 98 99	
2	Time deposits in a bank	1 2 3 98 99	1 2 3 4 5 6 98 99	
3	Current account in bank	1 2 3 98 99	1 2 3 4 5 6 98 99	
4	Cooperatives/NGO savings/ microfinance account	1 2 3 98 99	1 2 3 4 5 6 98 99	
5	Post office savings account	1 2 3 98 99	1 2 3 4 5 6 98 99	
6	Provident funds/pension account	1 2 3 98 99	1 2 3 4 5 6 98 99	
7	Other formal or informal savings account	1 2 3 98 99	1 2 3 4 5 6 98 99	
8	Stocks/shares/debentures	1 2 3 98 99	1 2 3 4 5 6 98 99	
9	Bonds/government certificates	1 2 3 98 99	1 2 3 4 5 6 98 99	
10	Life Insurance	1 2 3 98 99	1 2 3 4 5 6 98 99	
11	Deposit with another individual	1 2 3 98 99	1 2 3 4 5 6 98 99	
12	Mobile telephone account	1 2 3 98 99	1 2 3 4 5 6 98 99	
13	Money owed to you by other persons	1 2 3 98 99	1 2 3 4 5 6 98 99	
14	Cash at hand (including gold, jewelry and other valuables)	1 2 3 98 99	1 2 3 4 5 6 98 99	
15	Other financial asset, specify	1 2 3 98 99	1 2 3 4 5 6 98 99	

Q4. DO YOU ANY MEMBERS OF YOUR HOUSEHOLD CURRENTLY HAVE ANY OUTSTANDING LOANS?

1 YES > PLEASE PROVIDE THE FOLLOWING INFORMATION FOR EACH OUTSTANDING LOAN
2 NO > NEXT MODULE

	Q5. INDIVIDUAL OR INSTITUTION MONEY BORROWED FROM (USE CODE 1)	Q6. HOW MUCH OF THIS LOAN REMAINS TO BE PAID? (LOCAL CURRENCY) 98 DOES NOT KNOW 99 REFUSED TO RESPOND	Q7. ARE YOU YOURSELF INDIVIDUALLY OR JOINTLY OBLIGATED TO REPAY THIS MONEY? 1 YES, INDIVIDUALLY > NEXT LOAN 2 YES, JOINTLY 3 NO, SOMEONE ELSE IS OBLIGATED TO REPAY THE LOAN > NEXT LOAN (CIRCLE RESPONSE)	Q8. IF YOU WERE TO REPAY THIS LOAN FULLY TODAY, HOW MUCH WOULD YOU PERSONALLY HAVE TO PAY? (LOCAL CURRENCY) 98 DOES NOT KNOW 99 REFUSED TO RESPOND
1			1 2 3	
2			1 2 3	
3			1 2 3	
4			1 2 3	
5			1 2 3	
6			1 2 3	

CODE 1	TYPE OF GROUP	CODE 1	TYPE OF GROUP
1	Relative/family member	8	NGO
2	Friend/individual	9	Money lender
3	Employer	10	Business/shop
4	Private bank	11	Self-help group
5	Government bank	12	Credit card
6	Credit cooperative	96	Other, specify
7	Microfinance organization		

Indicator: Value of woman's bank/financial accounts
Definition: The current value of all bank and other financial accounts held in a woman's name.

Source: Adapted from the Gender Asset Gap Project in Ecuador, Ghana and India (2009). The project team leaders are Hema Swaminathan, Indian Institute of Management Bangalore (IIMB); Abena D. Oduro, University of Ghana; Carmen Diana Deere, University of Florida; Cheryl Doss, Yale University; and Caren Grown, American University.

	FINANCIAL ASSETS CODE (READ OUT LIST)	Q1. DO YOU HAVE MONEY IN ANY OF THE FOLLOWING TYPES OF ACCOUNTS IN A BANK OR OTHER FINANCIAL INSTITUTION THAT ARE UNDER YOUR NAME? 1. YES 2. NO >SKIP TO NEXT ASSET 98 DOES NOT KNOW 99. REFUSED TO RESPOND >SKIP TO NEXT ASSET (CIRCLE RESPONSE)	Q2. WHAT IS THE APPROXIMATE VALUE OF THIS ACCOUNT CURRENTLY? (LOCAL CURRENCY) 98 DOES NOT KNOW 99 REFUSED TO RESPOND
1	Savings Account in a bank	1 2 3 98 99	
2	Time deposits in a bank	1 2 3 98 99	
3	Current account in bank	1 2 3 98 99	
4	Cooperatives/NGO savings/ microfinance account	1 2 3 98 99	
5	Post office savings account	1 2 3 98 99	
6	Provident funds/pension account	1 2 3 98 99	
7	Other formal or informal savings account	1 2 3 98 99	

Indicator: Value of women's physical assets

Definition: Current value of female respondent's share of the household's physical assets, based on responses to Q4 below.

Source: Adapted from Uganda WEAI Questionnaire on "Women's Empowerment in Agriculture Index—New Questions" Pilot Version (August 2015).

NOW I WOULD LIKE TO ASK YOU ABOUT YOUR HOUSEHOLD'S ACCESS TO AND OWNERSHIP OF A NUMBER OF ITEMS THAT COULD BE USED TO GENERATE INCOME.					
	DESCRIPTION	Q1. DOES ANYONE IN YOUR HOUSEHOLD CURRENTLY HAVE ANY [ITEM]? YES.....1 NO.....2 >NEXT ITEM (CIRCLE RESPONSE)	Q2. DO YOU PERSONALLY OWN ANY OF THE ITEM? 1. YES, SOLELY 2. YES, JOINTLY 3. NO > NEXT ITEM (CIRCLE RESPONSE)	Q3. WHO WOULD YOU SAY CAN DECIDE WHETHER TO SELL, GIVE AWAY, MORTGAGE OR RENT [ITEM] MOST OF THE TIME? 1. SELF 2. PARTNER/SPOUSE 3. OTHER HH MEMBER 4. OTHER NON-HH MEMBER 97 DON'T KNOW 98 NOT APPLICABLE 99 REFUSED TO RESPOND (CIRCLE RESPONSE)	Q3. IF THIS ITEM WERE SOLD, HOW MUCH MONEY DO YOU THINK YOU WOULD PERSONALLY RECEIVE? 97 DON'T KNOW 98 NOT APPLICABLE 99 REFUSED TO RESPOND (LOCAL CURRENCY)
1	Agricultural land				
2	Large livestock (oxen, cattle)				
3	Small livestock (goats, pigs, sheep)				
4	Fowl (chickens, ducks, turkeys, pigeons)				
5	Fish pond or fishing equipment				
6	Farm equipment (non-mechanized: hand tools, animal-drawn plough)				
7	Farm equipment (mechanized: tractor-plough, power tiller, treadle pump)				
8	Nonfarm business equipment (sewing machine, computer)				
9	House or other structure				
10	Large consumer durables (refrigerator, TV, sofa)				

NOW I WOULD LIKE TO ASK YOU ABOUT YOUR HOUSEHOLD'S ACCESS TO AND OWNERSHIP OF A NUMBER OF ITEMS THAT COULD BE USED TO GENERATE INCOME.

	DESCRIPTION	Q1. DOES ANYONE IN YOUR HOUSEHOLD CURRENTLY HAVE ANY [ITEM]? YES.....1 NO.....2 >NEXT ITEM (CIRCLE RESPONSE)	Q2. DO YOU PERSONALLY OWN ANY OF THE ITEM? 1. YES, SOLELY 2. YES, JOINTLY 3. NO > NEXT ITEM (CIRCLE RESPONSE)	Q3. WHO WOULD YOU SAY CAN DECIDE WHETHER TO SELL, GIVE AWAY, MORTGAGE OR RENT [ITEM] MOST OF THE TIME? 1. SELF 2. PARTNER/SPOUSE 3. OTHER HH MEMBER 4. OTHER NON-HH MEMBER 97 DON'T KNOW 98 NOT APPLICABLE 99 REFUSED TO RESPOND (CIRCLE RESPONSE)	Q3. IF THIS ITEM WERE SOLD, HOW MUCH MONEY DO YOU THINK YOU WOULD PERSONALLY RECEIVE? 97 DON'T KNOW 98 NOT APPLICABLE 99 REFUSED TO RESPOND (LOCAL CURRENCY)
11	Small consumer durables (radio, cookware)				
12	Cell phone				
13	Other land not used for agricultural purposes				
14	Motorbike/motorcycle				
15	Car or truck				
16	Other item (specify) _____				

Indicator: Value of woman's motor vehicles
Definition: Current value of motor vehicles owned by woman (zero if none owned)

Indicator: Value of woman's mobile phones
Definition: Current value of mobile phone(s) owned by woman (zero if none owned)

	QUESTION	RESPONSE
Q1	Do you currently own one or more motorbikes, cars or trucks that are registered in your name?	Yes.....1 No.....2 > Q3
Q2	How many of each do you own? a. motorbikes b. cars/trucks	 (number owned) (number owned)
Q3	If you were to sell all of these motor vehicles today, how much money do you think you would receive after paying off all loans that you took out to purchase them?	(local currency)
Q4	Do you currently own more than one mobile telephone?	Yes.....1 No.....2 > Next module
Q5	If you were to sell all of your mobile phones today, how much money do you think you would receive after paying off all loans that you took out to purchase them?	(local currency)

2. SATISFACTION WITH LIFE (FINAL OUTCOME)

Indicator: Woman's overall satisfaction with life
Definition: Sum of the responses across all of the questions (excluding responses of zero), with higher values indicating greater satisfaction with life.

Source: Adapted from "Life Satisfaction Module" in Questionnaire for Individual Women, Multiple Indicator Cluster Survey (MICS), UNICEF (October 2013) (http://www.childinfo.org/files/MICS_Questionnaire_for_Individual_Women_20131022.docx)

	QUESTION	RESPONSE
LS2	Taking all things together, would you say you are very happy, somewhat happy, neither happy nor unhappy, somewhat unhappy or very unhappy? You can also look at these pictures to help you with your response. *Show side 1 of response card and explain what each symbol represents. Circle the response code selected by the respondent.*	Very happy.....5 Somewhat happy.....4 Neither happy nor unhappy.....3 Somewhat unhappy.....2 Very unhappy.....1
LS7	Now I will ask you questions about your level of satisfaction in different areas. In each case, we have five possible responses: please tell me, for each question, whether you are very satisfied, somewhat satisfied, neither satisfied nor unsatisfied, somewhat unsatisfied or very unsatisfied. Again, you can look at these pictures to help you with your response. *Show side 2 of response card and explain what each symbol represents. Circle the response code selected by the respondent, for questions ls7 to ls13.* How satisfied are you with your current work/job? If the respondent says that she does not have a job, circle "0" and continue with the next question. Do not probe to find out how she feels about not having a job, unless she tells you herself.	Very satisfied.....5 Somewhat satisfied.....4 Neither satisfied nor unsatisfied.....3 Somewhat unsatisfied.....2 Very unsatisfied.....1
LS10	How satisfied are you with how people around you generally treat you?	Very satisfied.....5 Somewhat satisfied.....4 Neither satisfied nor unsatisfied.....3 Somewhat unsatisfied.....2 Very unsatisfied.....1
LS13	How satisfied are you with your current income? *If the respondent says that she does not have any income, circle "0" and continue with the next question. Do not probe to find out how she feels about not having any income, unless she tells you herself.*	Does not have any income 0 Very satisfied.....5 Somewhat satisfied.....4 Neither satisfied nor unsatisfied.....3 Somewhat unsatisfied.....2 Very unsatisfied.....1
LS14	Compared to this time last year, would you say that your life has improved, stayed more or less the same, or worsened, overall?	Improved.....3 More or less the same.....2 Worsened.....1

	QUESTION	RESPONSE
LS15	And in one year from now, do you expect that your life will be better, will be more or less the same, or will be worse, overall?	Better.....3 More or less the same.....2 Worse.....1

RESPONSE CARD
SIDE 1

VERY HAPPY	SOMEWHAT HAPPY	NEITHER HAPPY, NOR UNHAPPY	SOMEWHAT UNHAPPY	VERY UNHAPPY
				

RESPONSE CARD
SIDE 2

VERY SATISFIED	SOMEWHAT SATISFIED	NEITHER SATISFIED, NOR UNSATISFIED	SOMEWHAT UNSATISFIED	VERY UNSATISFIED
				

Indicator: Woman's stress level

Definition: Sum of the coded responses to the questions below, with lower values of the indicator signifying lower stress levels.

Source: adapted from U.S. National Center for Health Statistics, NHANES Study (http://www.lmra.org/content/Facility/2/downloads/10-09-SelfTest-Stress.pdf)

	QUESTION	RESPONSE	CODE
Q1	How have you been feeling in general?	In an excellent frame of mind In a very good mood In a good mood mostly My mood has been up and down In a poor frame of mind mostly In a very poor frame of mind	1 2 3 4 5 6
Q2	Have you been bothered by nervousness?	Not at all A little Some—enough to bother me Yes—quite a bit Yes—very much so Extremely so—to the point where I could not work or take care of things	1 2 3 4 5 6
Q3	Have you been in firm control of your behavior, thoughts, emotions, and feelings?	Yes, definitely so Yes, for the most part Generally so Not too well No, and I am somewhat troubled by that No, and I am very troubled by that	1 2 3 4 5 6
Q4	Have you been feeling so sad, discouraged, or hopeless, or had so many problems that you wondered if anything was worthwhile?	Not at all A little Some—enough to bother me Yes—quite a bit Yes—very much so Extremely so—to the point that I have just about given up	1 2 3 4 5 6
Q5	Have you been feeling that you were under any strain, stress, or pressure?	Not at all A little About the same amount as usual Yes—more than usual Yes—quite a bit of pressure Yes—almost more than I could bear	1 2 3 4 5 6
Q6	How happy or satisfied have you been with your personal life?	Extremely happy Very happy Fairly happy Somewhat satisfied Somewhat dissatisfied Very dissatisfied	1 2 3 4 5 6
Q7	Have you had any reason to wonder if you were losing your mind or memory, or losing control over the way you act, talk, think, or feel?	Not at all Only a little Some—but not enough to be concerned Some, and I have been a little concerned Some, and I am quite concerned Yes, a lot, and I am very concerned	1 2 3 4 5 6

	QUESTION	RESPONSE	CODE
Q8	Have you been anxious, worried, or upset?	Not at all A little Some—enough to bother me Yes—quite a bit Yes—very much so Extremely so—to the point of being sick or almost sick	1 2 3 4 5 6
Q9	How often have you awakened refreshed and rested?	Every day Almost every day Fairly often Less than half the time Rarely None of the time	1 2 3 4 5 6
Q10	Have you been bothered by an illness, bodily disorder, pain, or fear about your health?	Not at all A little Some of the time Yes—a good bit of the time Yes—most of the time Yes—all of the time	1 2 3 4 5 6
Q11	Has your daily life been full of things that were interesting to you?	Yes—all of the time Yes—most of the time Yes—a good bit of the time Some of the time A little Not at all	1 2 3 4 5 6
Q12	Have you been feeling down-hearted and blue?	Not at all A little Some of the time Yes—a good bit of the time Yes—most of the time Yes—all of the time	1 2 3 4 5 6
Q13	Have you been feeling emotionally stable and sure of yourself?	Yes—all of the time Yes—most of the time Yes—a good bit of the time Some of the time A little Not at all	1 2 3 4 5 6
Q14	Have you been feeling tired, worn out, used-up, or exhausted?	Not at all A little Some of the time Yes—a good bit of the time Yes—most of the time Yes—all of the time	1 2 3 4 5 6
Q15	How concerned or worried have you been about your health?	Please respond with a number from 1-10 where 1 indicates “Not concerned at all” and 10 indicates “Very concerned”	# 1-10

	QUESTION	RESPONSE	CODE
Q16	How relaxed or tense have you been?	Please respond with a number from 1-10 where 1 indicates "Very relaxed" and 10 indicates "Very tense"	# 1-10
Q17	How much energy, pep, or vitality have you had?	Please respond with a number from 1-10 where 1 indicates "Very energetic, dynamic" and 10 indicates "No energy at all, listless"	# 1-10
Q18	How depressed or cheerful have you been?	Please respond with a number from 1-10 where 1 indicates "Very cheerful" and 10 indicates "Very depressed"	# 1-10

3. GENDER ROLES/NORMS (FINAL OUTCOME)

Indicator: Woman's roles in household decision-making

Definition: Sum of responses to the questions below, with the responses other than 1 coded zero. The exceptions are question WS24 (recode responses of 3 to 1 and others to zero) and question WS31 (recode responses of 2 to 1 in WS31a, WS31e and WS31f and other responses to zero). Higher values of this indicator signify a greater role for the woman in household decision-making.

Source: Adapted from "Women's Status Module" DHS Program (http://dhsprogram.com/pubs/pdf/DHSQMP/womens_status_module.pdf.pdf)

PLEASE RESPOND TO THE FOLLOWING QUESTIONS.

	QUESTION	
WS06	Who in your family usually has the final say on whether or not you should work to earn money?	Respondent =1 Husband/partner =2 Respondent & husband/partner jointly=3 Someone else =4 Respondent & someone else jointly =5 Decision not made /not applicable=6 Work.....1 2 3 4 5 6
WS08	Who in your family usually has the final say on the following decisions about your child(ren): Any decisions about children's schooling? What to do if a child falls sick? How children should be disciplined? Whether to have another child?	Respondent =1 Husband/partner =2 Respondent & husband/partner jointly=3 Someone else =4 Respondent & someone else jointly =5 Decision not made /not applicable=6 Schooling.....1 2 3 4 5 6 Medical.....1 2 3 4 5 6 Discipline.....1 2 3 4 5 6 Another child.....1 2 3 4 5 6

PLEASE RESPOND TO THE FOLLOWING QUESTIONS.

QUESTION

	Do you and your husband/partner talk about the following with each other often, sometimes, or never?		Often	Seldom	Never
WS10	Things that happen at this work/on the farm?	Events at work	1	2	3
	Things that happen at home?	Events at home	1	2	3
	What to spend money on?	Money matters	1	2	3
	Things that happen in the community?	Community matters	1	2	3
	Do you yourself control the money needed to buy the following things?		Yes	No	Does not buy
WS20	Vegetables or fruits?	Vegetables/fruit	1	2	3
	Clothes for yourself?	Clothes	1	2	3
	Any kind of medicine for yourself?	Medicine	1	2	3
	Toiletries for your like (*give local examples*)?	Toiletries	1	2	3

					WS25: If you ever need to, can you sell (ASSET) without anyone else's permission?		
	Now I would like to ask you some questions about financial matters. I ask these questions only to understand more about the financial position of women.						
	Please tell me if you alone, or jointly with your husband or someone else own....				Yes	No	Dk
WS24	Land?	1 ▼	2 ▼	3 ▶	1	2	3
	The house/dwelling you live in?	1 ▼	2 ▼	3 ▶	1	2	3
	Any other house, apartment, or dwelling?	1 ▼	2 ▼	3 ▶	1	2	3
	Jewelry or gems?	1 ▼	2 ▼	3 ▶	1	2	3
	Livestock such as (*give local examples*)?	1	2	3 ▶	1	2	3

PLEASE RESPOND TO THE FOLLOWING QUESTIONS.

	QUESTION				
WS31	Now I would like to get your opinion on some aspects of family life. Please tell me if you agree or disagree with each statement:		Agree	Disagree	Dk
	a. The important decisions in the family should be made only by the men of the family.	Family decisions by men	1	2	8
	b. If the wife is working outside the home, then the husband should help her with household chores.	Husband should help	1	2	8
	c. A married woman should be allowed to work outside the home if she wants to.	Women should work	1	2	8
	d. The wife has a right to express her opinion even if she disagrees with what her husband is saying.	Wife to express opinion	1	2	8
	e. A wife should tolerate being beaten by her husband in order to keep the family together.	Tolerate being beaten	1	2	8
	f. It is better to send a son to school than it is to send a daughter.	Better to school son	1	2	8
WS32	Are you usually permitted to go to the following places on your own, only if someone accompanies you, or not at all?		Alone	Not alone	Never
	To the local market to buy things?	Market	1	2	3
	To a local health center or doctor?	Health center	1	2	3
	To the community center or other nearby meeting place?	Community center	1	2	3
	To homes of friends in the neighborhood?	Friends	1	2	3
	To a nearby shrine/mosque/temple/church?	Religious place	1	2	3
	Just outside your house or compound?	Outside the home	1	2	3
WS33	Are you a member of any type of association, group or club which holds regular meetings?	YES.....1 NO.....2			
WS35	When there is a local or a national election of any kind do you vote always, sometimes, or never?	Always votes.....1 Sometimes votes.....2 Never votes.....3 Too young to vote.....4 Never an election.....5			

4. SELF CONFIDENCE (INTERMEDIATE AND FINAL OUTCOME)

Indicator: Woman's overall self-confidence

Definition: Sum of the responses to the statements below (excluding responses of "Don't know/No opinion/Does not apply"), with the responses to items 7, 8 and 14 reverse scored (i.e., "Disagree strongly"=5, "Disagree"=4, etc). Higher values of the indicator signify a higher level of the woman's overall self-confidence.

Sources: Adapted from Sri Lanka Female Enterprise Survey, Questionnaire for Female Business Owners Not Previously Surveyed (http://microdata.worldbank.org/index.php/catalog/1553/related_materials); and Kenya Female Enterprise Survey (2013), Baseline Questionnaire (version 10) (http://microdata.worldbank.org/index.php/catalog/1985)

PLEASE INDICATE HOW MUCH YOU AGREE/DISAGREE WITH EACH STATEMENT BELOW

CODES FOR RESPONSES

1 = DISAGREE STRONGLY
2 = DISAGREE
3 = NEUTRAL
4 = AGREE
5 = AGREE STRONGLY
6 = DON'T KNOW/NO OPINION/DOES NOT APPLY

	STATEMENT	RESPONSE					
1	I plan tasks carefully	1	2	3	4	5	6
2	I make up my mind quickly	1	2	3	4	5	6
3	In uncertain times I usually expect the best	1	2	3	4	5	6
4	I can think of many times when I persisted with work when others quit	1	2	3	4	5	6
5	I continue to work on hard projects even when others oppose me	1	2	3	4	5	6
6	I like to juggle several activities at the same time	1	2	3	4	5	6
7	If something can go wrong for me, it will	1	2	3	4	5	6
8	I never try anything that I am not sure of	1	2	3	4	5	6
9	I'm always optimistic about my future	1	2	3	4	5	6
10	A person can get rich by taking risks	1	2	3	4	5	6
11	It is important for me to do whatever I'm doing as well as I can even if it isn't popular with people around me	1	2	3	4	5	6

PLEASE INDICATE HOW MUCH YOU AGREE/DISAGREE WITH EACH STATEMENT BELOW

CODES FOR RESPONSES

1 = DISAGREE STRONGLY
2 = DISAGREE
3 = NEUTRAL
4 = AGREE
5 = AGREE STRONGLY
6 = DON'T KNOW/NO OPINION/DOES NOT APPLY

	STATEMENT	RESPONSE					
12	When a group I belong to plans an activity, I would rather direct it myself than just help out and have someone else organize it	1	2	3	4	5	6
13	It is important to me to perform better than others on a task	1	2	3	4	5	6
14	I rarely count on good things happening to me	1	2	3	4	5	6
15	Even when by business/farm is doing well I keep my eyes open in case I find a way to improve it	1	2	3	4	5	6

Indicator: Woman's willingness to assert herself
Definition: Sum of the responses across all six questions (omitting responses of "don't know/not applicable"). Higher values of the indicator signify higher levels of the woman's willingness to speak out. Note that Question #2 may not be relevant in some settings.

Source: Kenya Female Enterprise Survey (2013), Baseline Questionnaire (version 10) (http://microdata.worldbank.org/index.php/catalog/1985)

PLEASE INDICATE YOUR LEVEL OF COMFORT IN SPEAKING OUT IN DIFFERENT KINDS OF SITUATIONS.

CODES FOR RESPONSES

1 = NO, NOT AT ALL COMFORTABLE
2 = YES, BUT WITH A GREAT DEAL OF DIFFICULTY
3 = YES, BUT WITH A LITTLE DIFFICULTY
4 = YES, FAIRLY COMFORTABLE
5 = YES, VERY COMFORTABLE
6 = DON'T KNOW/NOT APPLICABLE

	QUESTION	RESPONSE					
1	Speaking out at a meeting of other women to talk about some common issue?	1	2	3	4	5	6
2	Speaking out at a meeting of men and women to talk about some common issue?	1	2	3	4	5	6
3	Talking to people who work for you about a disagreement?	1	2	3	4	5	6
4	Refusing someone who has asked to buy something for less than you feel is a fair price?	1	2	3	4	5	6
5	Bargaining with a supplier to get a lower price on something?	1	2	3	4	5	6
6	Do you feel comfortable speaking out about a household money issue with your spouse if you are not in agreement on what to do?	1	2	3	4	5	6

Indicator: Woman's willingness to take risk

Definition: An indicator or "willingness to take risk" can be constructed from the responses to the questions below as follows. The indicator has a value of 1 if the respondent prefers a certain payoff to the flip of a coin even with an expected payoff equal to twice the certain payoff (Option 1 in Question 2), a value of 2 if the respondent is willing to toss the coin for an expected payoff equal to twice the certain payoff (Option 2 in Question 2), a value of 3 if the respondent is willing to toss the coin if the expected payoff is only 50% higher than the certain payoff (Option 2 in Question 1), and a value of 4 if the respondent is willing to toss the coin even if the expected payoff is just equal to the certain payoff (Option 2 in Question 3).

Source: 2012 STEP Household Questionnaire, Lao PDR (World Bank)

	QUESTION	
1	Imagine that you have a choice between the following two options: **Option 1 – Receive $50 for sure.** OR **Option 2 – Flip a coin and receive 0 if it's tails or $150 if it's heads.** Which option would you take?	Option 1 (TAKE THE SURE MONEY)...1 Option 2 (FLIP THE COIN).....2 >>3
2	Now imagine that you have a choice between the following two options: **Option 1 – Receive $50 for sure.** OR **Option 2 – Flip a coin and receive 0 if it's tails or $200 if it's heads.** Which option would you take?	Option 1 (TAKE THE SURE MONEY)...1 >>Next module Option 2 (FLIP THE COIN).....2 >>Next module
3	Now imagine that you have a choice between the following two options: **Option 1 – Receive $50 for sure.** OR **Option 2 – Flip a coin and receive 0 if it's tails or $100 if it's heads.** Which option would you take?	Option 1 (TAKE THE SURE MONEY)...1 Option 2 (FLIP THE COIN).....2

5. SELF-ESTEEM

Indicator: Woman's self-esteem

Definition: This indicator is based on the Rosenberg Self-Esteem Scale, which is the most widely used measure of self-esteem. The scores are summed, except for items 2, 5, 6, 8 and 9, which are reverse-scored (i.e., "Strongly agree"=1, "Agree"=2, "Disagree"=3, and "Strongly disagree"=4). Higher scores signify a higher level of women's self-esteem.

Source: http://fetzer.org/sites/default/files/images/stories/pdf/selfmeasures/Self_Measures_for_Love_and_Compassion_Research_SELF-ESTEEM.pdf

PLEASE INDICATE HOW STRONGLY YOU AGREE OR DISAGREE WITH THE FOLLOWING STATEMENTS.

	QUESTION	RESPONSE			
1	On the whole, I am satisfied with myself.	1	2	3	4
2	At times I think I am no good at all.	1	2	3	4
3	I feel I have a number of good qualities.	1	2	3	4
4	I am able to do things as well as most other people.	1	2	3	4
5	I feel I do not have much to be proud of.	1	2	3	4
6	I certainly feel useless at times.	1	2	3	4
7	I feel that I'm a person of worth, at least on an equal plane with others.	1	2	3	4
8	I wish I could have more respect for myself.	1	2	3	4
9	All in all, I am inclined to feel that I am a failure.	1	2	3	4
10	I take a positive attitude toward myself.	1	2	3	4

II. INTERMEDIATE OUTCOMES

A. URBAN AND RURAL WOMEN ENTREPRENEURS

1. BUSINESS PRACTICES (INTERMEDIATE OUTCOME)

Indicator: Woman's adoption of recommended business practices

Definition: Two alternative modules are provided, one relatively short and the other more detailed. For both modules, the indicator of the adoption of recommended business practices is defined as the sum of the responses to the questions, with "Yes"=1 and "No" or "999"=0. Higher values of this indicator signify a higher level of the woman's adoption of recommended business practices.

Module #1

Source: Adapted from multiple sources.

	QUESTION	RESPONSE	YES.....1 NO.....2
Q1	Do you keep records of your business, including sales, expenses and inventory?		
Q2	Do you keep your business and household finances separated?		
Q3	Does your business have an updated business plan?		
Q4	Have you ever applied for a loan from a bank or other formal financial institution for your business?		
Q5	Do you visit your main customers at least once in three months?		
Q6	Do you advertise at least once in six months?		

Module #2

Source: Adapted from Sri Lanka Female Enterprise Survey, Questionnaire for Female Business Owners Not Previously Surveyed (http://microdata.worldbank.org/index.php/catalog/1553/related_materials)

PLEASE INDICATE WHICH OF THE FOLLOWING HAVE YOU DONE IN THE LAST THREE MONTHS?

	QUESTION	RESPONSE
MARKETING		
Q1	Visited one of your competitor's businesses to see what prices they are charging?	1. Yes 2. No 999. No Competitors N/A
Q2	Visited one of your competitor's businesses to see what products they have available for sale?	1. Yes 2. No 999. No Competitors N/A
Q3	Asked your existing customers whether there are any other products they would like you to sell or produce?	1. Yes 2. No
Q4	Talked with a former customer to find out why they have stopped buying from your business?	1. Yes 2. No 999. Don't have a former customer
Q5	Asked a supplier about which products are selling well in your industry?	1. Yes 2. No 999. Supplier has no knowledge of my industry (e.g., provides general inputs)
Q6	In the last three months have you used a special offer to attract customers?	1. Yes 2. No
Q7	In the last six months, have you done any form of advertising?	1. Yes 2. No
Q8	Did you do anything to measure the effectiveness of the advertising?	1. Yes 2. No
BUYING AND STOCK CONTROL		
Q9	In the last three months have you attempted to negotiate with a supplier for a lower price on raw materials?	1. Yes 2. No

PLEASE INDICATE WHICH OF THE FOLLOWING HAVE YOU DONE IN THE LAST THREE MONTHS?

	QUESTION	RESPONSE
Q10	In the last three months, have you compared the prices or quality offered by alternate suppliers or sources of raw materials to the supplier or source you have?	1. Yes 2. No
Q11	Do you have a record-keeping system which allows you to know how much stock of goods to sell or raw materials you have on hand?	1. Yes 2. No 999 No inventories

COSTING AND RECORD-KEEPING

Q12	Do you keep written business records?	1. Yes 2. No (skip to Q16)
Q13	Do you record every purchase and sale made by the business?	1. Yes 2. No
Q14	Are you able to use your records to easily see how much cash your business has on hand at any point in time?	1. Yes 2. No
Q15	Do you regularly use your records to know whether sales of a particular product are increasing or decreasing from one month to another?	1. Yes 2. No
Q16	Have you worked out the cost to you of each main product you sell?	1. Yes 2. No
Q17	Do you know which goods you make or sell are most profitable?	1. Yes 2. No
Q18	Do you have a written budget, which tells you how much you have to pay each month for rent, electricity, equipment maintenance, transport, advertising, and other indirect costs of the business?	1. Yes 2. No
Q19	Do you sell any goods on credit to customers?	1. Yes 2. No (skip to Q21)
Q20	Do you have a written record of how much each customer owes you?	1. Yes 2. No
Q21	If you wanted to apply for a bank loan, and were asked to provide records to show that you have enough money left each month after paying business expenses to repay a loan, would your records allow you to document this to the bank?	1. Yes 2. No

PLEASE INDICATE WHICH OF THE FOLLOWING HAVE YOU DONE IN THE LAST THREE MONTHS?

	QUESTION	RESPONSE
FINANCIAL PLANNING		
Q22	Do you have a target set for sales over the next year?	1. Yes 2. No
Q23	Have you made a budget of what costs facing your business are likely to be over the next year?	1. Yes 2. No

2. VALUE OF BUSINESS TRAINING (INTERMEDIATE OUTCOME)

Indicator: Woman's willingness to pay for general business training

Definition: The maximum amount urban or rural female entrepreneurs are willing to pay to receive general business training (based on responses to Question 7.9 below), including zero values for responses of "No" to Question 7.8 for those unwilling to pay.

Indicator: Woman's willingness to pay for specialized technical business training

Definition: The maximum amount urban or rural female entrepreneurs are willing to pay to receive specialized technical business training (based on responses to Question 7.11 below), including zero values for responses of "No" to Question 7.10 for those unwilling to pay.

Source: Adapted from Sri Lanka Female Enterprise Survey (2009-2011), Questionnaire for Female Business Owners Not Previously Surveyed (http://microdata.worldbank.org/index.php/catalog/1553/related_materials)

	QUESTION	RESPONSE
7.8	If you were offered training related to general business skills at a reasonable price would you be interested?	Yes.....1 No.....2 Not sure/Don't know.....3
7.9	If you were offered such general business training for 40-45 hrs (i.e. 5-6 days), how much would you be willing to pay for such a training program? IF RESPONDENT IS HESITANT, PROBE WITH QUESTIONS OF THE FORM: "Would you be willing to pay [insert an amount above the cost of the training]?" IF THE RESPONDENT SAYS "YES" INITIALLY, KEEP INCREASING THE PRICE AND ASK AGAIN UNTIL THE RESPONDENT SAYS "NO". IF THE RESPONDENT SAYS "NO" INITIALLY, KEEP LOWERING THE PRICE AND ASK AGAIN UNTIL THE RESPONDENT SAYS "YES".	_____ (*local currency*)

	QUESTION	RESPONSE
7.10	If you were offered specialized technical training related to your business or intended business sector at a reasonable price would you be interested?	Yes.....1 No.....2 Not sure/Don't know.....3
7.11	If you were offered such specialized technical training for 40-45 hrs (i.e. 5-6 days), how much would you be willing to pay for such a training program? IF RESPONDENT IS HESITANT, PROBE WITH QUESTIONS OF THE FORM: "Would you be willing to pay [insert an amount above the cost of the training]?" IF THE RESPONDENT SAYS "YES" INITIALLY, KEEP INCREASING THE PRICE AND ASK AGAIN UNTIL THE RESPONDENT SAYS "NO". IF THE RESPONDENT SAYS "NO" INITIALLY, KEEP LOWERING THE PRICE AND ASK AGAIN UNTIL THE RESPONDENT SAYS "YES".	_____ (*local currency*)

3. GENDER ROLES/NORMS (INTERMEDIATE OUTCOME)

Indicator: Woman's decision-making role in own business

Definition: Sum of the responses for all tasks, with each response given a value of 2 if the response to Q1= "Yes" and the response to Q2="No", a value of 1 if the response to both Q1 and Q2="Yes", or a value of 0 if the response to Q1="No" (and with responses of "Does not apply" ignored). Higher values of this indicator signify greater decision-making power of the woman in her own business.

Source: Adapted from Kenya Female Enterprise Survey (2013), Baseline Questionnaire (version 10) (http://microdata.worldbank.org/index.php/catalog/1985)

I AM GOING TO LIST SOME TYPICAL TASKS YOU HAVE TO PERFORM IN YOUR BUSINESS. PLEASE TELL ME IF YOU PARTICIPATE IN THESE TASKS, EITHER BY YOURSELF OR WITH ANOTHER PERSON?

	TASK DESCRIPTION	Q1. ARE YOU INVOLVED IN THIS TASK? 1 = YES 2 = NO > NEXT TASK 97 = DOES NOT APPLY TO THIS BUSINESS	Q2. IS ANYONE ELSE INVOLVED IN THIS TASK? 1 = YES 2 = NO
1.	Deciding which products or services to make or sell		
2.	Procurement of inputs or goods		
3.	Deciding whether to invest in the business (a machine, large tool, adding more stock to sell)		
4.	Deciding whether to take out a loan to invest in the business		
5.	Negotiating with suppliers		
6.	Setting the prices of goods or negotiating with buyers/ middlemen		
7.	Selling goods to customers		
8.	Dealing with officials (banks, market, government)		

B. RURAL WOMEN ENTREPRENEURS AND FARMERS

1. AGRICULTURAL PRACTICES (INTERMEDIATE OUTCOME)

Indicator: Woman's adoption of recommended agricultural practices

Definition: The indicator can be calculated as the sum of the responses for the various agricultural practices, with a "Yes" response=1 and a "No" response=0. Higher values of this indicator signify greater use of improved agricultural practices by the woman.

Source: Adapted from USAID Sudan Food, Agribusiness, and Rural Markets (FARM) Project (http://www.usaid.gov/developer/SouthSudanBaseline); and Ethiopia Farmer Innovation Fund Impact Evaluation (2012), Midline Survey (Women's Module 2) (http://microdata.worldbank.org/index.php/catalog/2042)

DURING THE LAST YEAR (12 MONTHS) DID YOU APPLY/USE ANY OF THE FOLLOWING AGRICULTURAL [PRACTICES]?

(THE LIST OF PRACTICES IS ONLY INDICATIVE AND SHOULD BE MODIFIED AS APPROPRIATE FOR A GIVEN APPLICATION)

	PRACTICE DESCRIPTION	1=YES	2=NO	98=DOES NOT APPLY
1	New/improved seed varieties			
2	Applied additional fertilizer			
3	Different planting method (e.g., row planting, spacing)			
4	New weeding methods			
5	New pest control measures			
6	Deep plowing			
7	New crop rotation			
8	New crop storage method*			
9	New marketing method			
10	New natural resource management method (e.g., soil conservation, water management)			
11	New livestock types			
12	Improved livestock breeds			
13	Sell farm products in more distant markets			
14	Sell to a commodity buyer			

DURING THE LAST YEAR (12 MONTHS) DID YOU APPLY/USE ANY OF THE FOLLOWING AGRICULTURAL [PRACTICES]? *(THE LIST OF PRACTICES IS ONLY INDICATIVE AND SHOULD BE MODIFIED AS APPROPRIATE FOR A GIVEN APPLICATION)*				
	PRACTICE DESCRIPTION	1=YES	2=NO	98=DOES NOT APPLY
15	Hire labor			
16	Keep written records of farming activities			
17	Keep farm income in a bank			

** IMPROVED STORAGE INCLUDES IN-HOME, BAGGED AND STACKED ON PALLETS; BRICK STORE, BAGGED AND STACKED ON PALLETS, METAL CRIB OR SILO*

2. VALUE OF ACCESS TO NEW/IMPROVED AGRICULTURAL TECHNOLOGY

Indicator: Woman's willingness to pay for access to new/improved agricultural technology

Definition: The maximum amount that a woman is willing to pay to have access to a specified new/improved technology.

Source: adapted from Malawi Technology Adoption Risk Initiative, Household Baseline Survey (2006) (http://microdata.worldbank.org/index.php/catalog/1541)

	QUESTION	RESPONSE (LOCAL CURRENCY)
	Suppose you could purchase [insert brief description new/improved technology, e.g., improved seeds for a given crop]. How much would you be willing to pay for [insert quantity, e.g., 30 kilogram of improved seeds]?	
M5	IF RESPONDENT IS HESITANT, PROBE WITH QUESTIONS OF THE FORM: "Would you be willing to pay [insert an amount above cost]?" IF THE RESPONDENT SAYS "YES" INITIALLY, KEEP INCREASING THE PRICE AND ASK AGAIN UNTIL THE RESPONDENT SAYS "NO". IF THE RESPONDENT SAYS "NO" INITIALLY, KEEP LOWERING THE PRICE AND ASK AGAIN UNTIL THE RESPONDENT SAYS "YES".	

3. GENDER ROLES/NORMS (INTERMEDIATE OUTCOME)

Indicator: Woman's decision-making role in her own farm

Definition: This indicator is defined as the sum of the responses for all tasks, with each response given a value of 2 if the response to Q1= "Yes" and the response to Q2="No", a value of 1 if the response to both Q1 and Q2="Yes", and with a value of 0 if the response to Q1="No" (and with responses of "Does not apply" ignored).

Higher values of this indicator signify greater decision-making power of the woman in her farm.

Source: Adapted from USAID Sudan Food, Agribusiness, and Rural Markets (FARM) Project http://www.usaid.gov/developer/SouthSudanBaseline

I AM GOING TO LIST SOME TYPICAL TASKS YOU HAVE TO PERFORM IN YOUR FARM. PLEASE TELL ME IF YOU PARTICIPATE IN THESE TASKS, EITHER BY YOURSELF OR WITH ANOTHER PERSON?

	TASK DESCRIPTION	Q1. ARE YOU INVOLVED IN THIS TASK? 1 = YES 2 = NO > NEXT TASK 97 = DOES NOT APPLY TO THIS FARM	Q2. IS ANYONE ELSE INVOLVED IN THIS TASK? 1 = YES 2 = NO
1	Choice of crops to plant		
2	Method of planting (rows, broadcast, number of seeds per hole)		
3	Type of seeds to use		
4	Timing of planting and harvesting		
5	Whether to use fertilizer and how much to apply		
6	How to store crops after harvest		
7	Where to sell crops		
8	Buying farming inputs (seeds, fertilizer, etc.)		
9	Keep written records of farming activities		
10	Which types of livestock/poultry to raise		
11	Whether to purchase additional livestock/poultry		
12	Whether to sell livestock/poultry		

C. BOTH URBAN AND RURAL WOMEN

1. TECHNOLOGY ADOPTION AND EFFECTIVE USE (INTERMEDIATE OUTCOME)

Source: Adapted from Booz&Co, ExxonMobil and Cherie Blair Foundation for Women. 2012. Mobile Value Added Services: A Business Growth Opportunity for Women Entrepreneurs. Appendix E: Primary Research: Survey of Women Entrepreneurs.

Indicator: Woman's intensity of mobile phone use for business purposes
Definition: The value of this indicator is the response to Q2 in the questionnaire module below. If the response to Q1=2 "No"), the value of the indicator equals zero.

Questionnaire module

	QUESTION	RESPONSE
Q1	Do you personally own or have regular access to a mobile phone?	Yes.....1 No.....2 > next module
Q2	How often do you use your mobile phone for business purposes?	Never.....0 Seldom (e.g., once per month).....1 Sometimes (e.g., once per week).....2 Often (e.g., every day).....3 Very often (e.g., several times per day).....4

2. WOMAN'S SELF-CONFIDENCE (INTERMEDIATE OUTCOME)

Use the indicators and questionnaire modules for women's self-confidence in I.C.4 above

3. GENDER NORMS/ROLES (INTERMEDIATE OUTCOME)

Indicator. Sharing of house work between spouses/partners

Definition: Sum of the responses across all tasks (excluding responses of "does not apply"). Higher values of this indicator signify increased sharing of house work between spouses/partners.

Source: Adapted from ICRW International Men and Gender Equality Survey (IMAGES) Questionnaire (http://www.icrw.org/sites/default/files/publications/International-Men-and-Gender-Equality-Survey-IMAGES.pdf)

	QUESTION	RESPONSE
Q1	Do you have a spouse or regular partner?	Yes.....1 No.....2 > next module
Q2	Does your spouse/partner live with you?	Yes.....1 No.....2 > next module
Q3	If you disregard the help you receive from other household members, how do you and your spouse/partner divide the following tasks	

	TASK	I DO EVERYTHING	USUALLY ME	SHARED EQUALLY OR DONE TOGETHER	USUALLY PARTNER	PARTNER DOES EVERYTHING	DOES NOT APPLY
1	Washing clothes						
2	Repairing the house						
3	Buying food						
4	Cleaning the house						
5	Cleaning the bathroom/toilet						
6	Preparing food						
7	Paying bills						
8	Caring for small children						
9	Playing with children						
10	Helping children with their school work						
11	Taking children to or from school						
12	Caring for elderly household members						
13	Caring for sick household members						

4. ENGAGEMENT/PARTICIPATION IN COMMUNITY, BUSINESS, OR FARMER GROUPS (INTERMEDIATE OUTCOME)

Indicator: Woman's participation in groups

Definition: The number of hours per month women spend participating in groups (based on responses to Q7 below)

Indicator: Woman's participation in mainly women's groups

Definition: This indicator is defined as the number of hours per month women spend participating in groups (based on responses to Q7 below) a majority of whose members are female (based on responses to Q4 and Q5 below).

Source: Adapted from Ethiopia Farmer Innovation Fund Impact Evaluation (2012), Midline Survey (Women's Module 6) (http://microdata.worldbank.org/index.php/catalog/2042)

Q1. DO YOU PARTICIPATE IN ANY GROUPS, FOR EXAMPLE, BUSINESS ASSOCIATIONS, FARMER'S GROUPS, CREDIT GROUPS?

1 YES > PLEASE PROVIDE THE FOLLOWING INFORMATION FOR EACH GROUP
2 NO > NEXT MODULE

	Q2. LIST THE NAME OF EACH GROUP THAT THE RESPONDENT BELONGS TO ON A SEPARATE LINE	Q3. WHAT TYPE OF GROUP IS IT? (USE CODE 1)	Q4. HOW MANY MEMBERS ARE IN THE GROUP?	Q5. HOW MANY MEMBERS ARE FEMALE?	Q6. HOW MANY HOURS PER MONTH DO YOU SPEND ON ACTIVITIES OR MEETINGS WITH THIS GROUP?
1					
2					
3					
4					
5					
6					
7					

CODE 1	TYPE OF GROUP	CODE 1	TYPE OF GROUP
1	Trade or business association	7	Youth group
2	Farmers' group	8	Civic/community organization
3	Agricultural cooperative	9	(other group)
4	Credit association / microfinance	10	(other group)
5	Informal savings group	11	(other group)
6	Religious group	12	Other (specify_____)