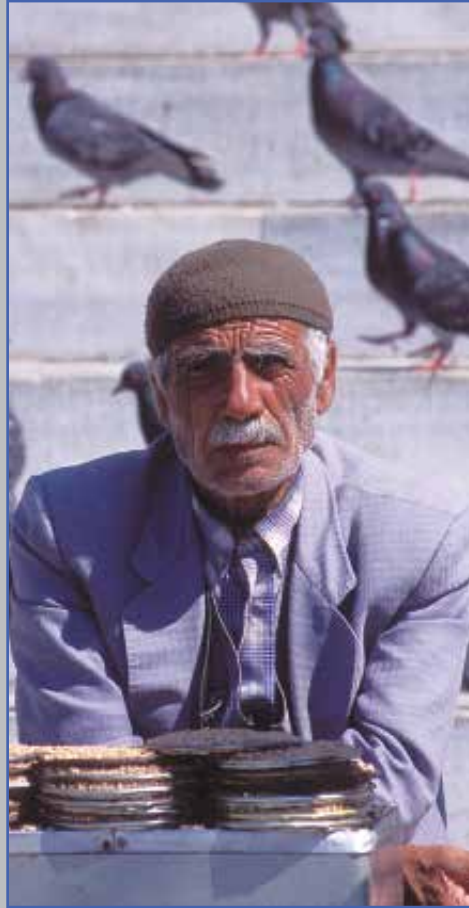


FINANCIAL
LITERACY
&
EDUCATION
RUSSIA TRUST FUND



A **TOOLKIT**
for the **evaluation** of
financial capability programs
in low- and middle-income
countries

Summary

A toolkit for the evaluation of
financial capability programs
in low- and middle-income
countries

Summary



© 2013 International Bank for Reconstruction and Development / The World Bank

1818 H Street, NW
Washington, DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

The findings, interpretations, and conclusions expressed here do not necessarily reflect the views of the Executive Directors of The World Bank or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Rights and Permissions

The material in this work is subject to copyright. Because the World Bank encourages dissemination of its knowledge, this work may be reproduced, in whole or in part, for noncommercial purposes as long as full attribution to this work is given.

Any queries on rights and licenses, including subsidiary rights, should be addressed to the Office of the Publisher, The World Bank, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2422; e-mail: pubrights@worldbank.org.

Cover photos: World Bank
Cover design/layout: Nita Congress



A large, bold, blue letter 'A' graphic that serves as a decorative element for the start of the text.

major initiative under the World Bank's Russia Financial Literacy and Education Trust Fund is the development Toolkit, which provides methodological guidance for the evaluation of financial capability programs. In low-income countries in particular, there are many obstacles to conducting rigorous evaluations. For example, there are problems with physical infrastructure, with securing financial and human resources, and with maintaining randomization. The Trust Fund set out to develop a toolkit specifically targeted to these circumstances.

The Toolkit is different from other monitoring and evaluation (M&E) handbooks in a number of ways:

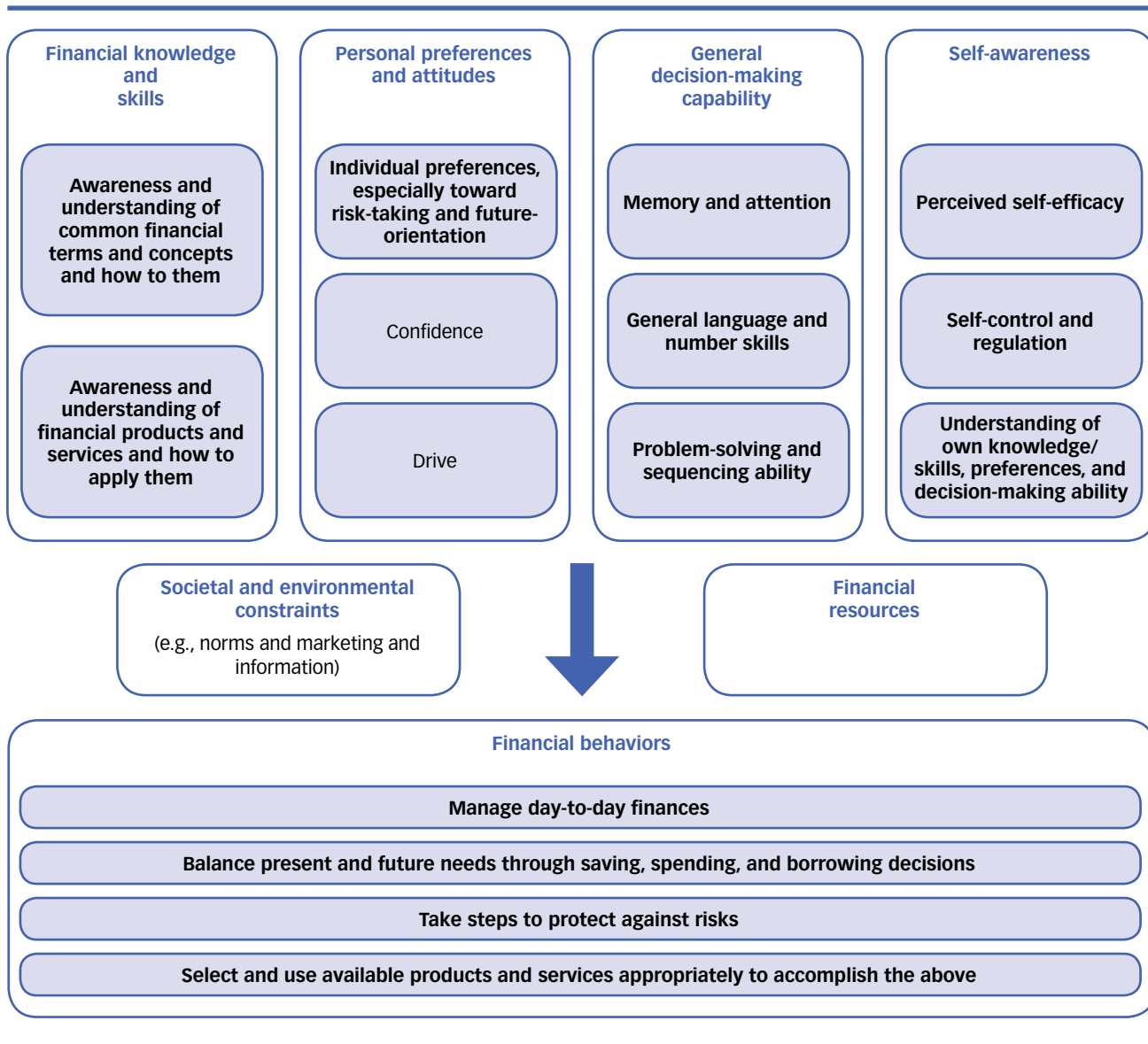
- It is designed to specifically address the issues and challenges of evaluating financial education and other financial capability enhancement programs, including developing outcome measures tailored to the unique characteristics of these kinds of programs.
- It provides an overview of all evaluation and research methods that are relevant to these programs, including impact and process evaluation, and quantitative and qualitative techniques.
- It is linked to a coordinated set of program evaluations supported by the Trust Fund to illustrate the use of the methods, describing the design of the programs and some of the results that were found.

The Toolkit is based on an emerging concept of financial capability that is more expansive and behavior-oriented than the earlier more knowledge-oriented concept of financial literacy. It defines financial capability as the internal capacity to act in one's own best financial interests, given the socioenvironmental conditions. It offers a framework and methods to evaluate programs in relation to behavioral outcomes, defining a financially capable person as one who is able to make and implement financial decisions that are considered appropriate for his or her peers and community, given the external circumstances. This initiative is closely related to a complementary program under the Trust Fund: to develop a survey instrument to measure financial capability and its distribution among the populations of low- and middle-income countries.

Figure 1 outlines this conceptual model of financial capability. The top four boxes are personal factors that comprise financial capability. The interaction of these factors with societal and environmental factors and financial resources determines financial behaviors.

The Toolkit discusses both theory and methods through real-world examples. It takes the reader through concrete steps from conceptual formulation of the evaluation, to choosing methods, to conducting data collection and analysis. It combines the latest evaluation technology with challenges and best practices realized among the diverse range of programmatic approaches, countries, and methods covered by the program evaluations supported by the Trust Fund. In this way, it aims to lower barriers to high-quality evaluation in low- and middle-income countries and advance evidence on the effectiveness of financial capability interventions.

FIGURE 1 A CONCEPTUAL MODEL OF FINANCIAL CAPABILITY



ORGANIZATION OF THE TOOLKIT

The Toolkit is presented as a series of short chapters. Although it is important to be familiar with all parts of the M&E process, it is not necessary to read the guide from beginning to end. Instead, each chapter is conceived as a stand-alone presentation that can be read independently of the others. Short summaries of the Toolkit’s main chapters are provided below.

1 FINANCIAL CAPABILITY: APPROACHES AND CONCEPTS

Leveraging the importance of improving financial capability

In the past decade or so, there has been an increasing awareness that **the ability of individuals to achieve personal financial well-being can have serious implications both for those individuals and for society as a whole**. Whether individuals achieve personal financial well-being depends on many things, including financial inclusion, consumer protection, over-indebtedness, and social protection (particularly in areas where consumers have to make their own provision for it). Financial capability programs in these policy areas can be useful tools in helping achieve this impact.

Interest in improving individuals' financial capabilities was initially concentrated in high-income countries, but now this interest has expanded to poorer parts of the world, including both low- and middle-income countries. Although people living on low incomes in a low-income country can have very sophisticated financial lives that may not necessarily require interaction with financial services, concerns can be subtly different in such settings. Research carried out as part of the Trust Fund confirms this conclusion, while reaffirming the existence of core competencies that are applicable across income levels and countries.

A changing focus on how to improve financial capability

There has been a shift in conceptualizing financial capability from a **cognitive/normative approach**—which assumes that a lack of financial knowledge is a key constraint and that appropriate knowledge and the ability to apply it determines appropriate behaviors (and outcomes)—to an **outcome-oriented/positive approach** that is more focused on behavior or outcomes and adopts an empirical approach to determining the behaviors that should be considered capable and those that should be considered less capable. The Trust Fund research initiative has adopted the outcome-oriented/positive approach.

Early programs tended to be education based and typically used workshops with adults and school-based teaching for children and young people. A shift in focus recognizes that education alone (in its narrowest sense) may not be the best (or even the most appropriate) way of influencing behaviors. This recognition has translated more recently into a wider range of interventions that seek to modify behaviors rather than just inculcate information. These interventions range from one-on-one guidance or counseling to social marketing, “edutainment,” and the use of specific design elements in general financial information and disclosure materials for consumers.

A very small evidence base on what actually works...

Unfortunately, despite this burgeoning interest and rapid growth in the number of financial capability interventions of all kinds, we actually know very little about what works in financial capability programs—not just in developing countries but also on a global scale. While there is a robust body of evidence that individual financial capability or literacy leads to better behavior and outcomes in both developed and developing countries, few studies are able to convincingly show the ability of financial capability programs to affect financial behavior, and even fewer convincingly document how. Also, the evidence on specific features of successful interventions—whether content, delivery, or timing—that work best is even more limited.

...But a growing awareness of the need for rigorous evaluation

In response to these limitations, there are an **increasing number of evaluations that explicitly aim to understand the causal relationship between interventions and financial knowledge, attitude, and behavioral outcomes**. These include Trust Fund pilot initiatives that deliver financial capability programs through social marketing or edutainment, based on large marketing campaigns in television, radio, and other media. They use drama, radio programming, television soap opera programs, specialized films, street theater performances, Internet-based media/ CD- or DVD-ROM modules, and comic books/animation. Such types of outreach may better capture individuals' attention and can target audiences that may be illiterate; thus, they may be particularly appropriate in lower-income countries.

Building capacity in financial capability evaluation

Drawing on both the literature in this area and the Trust Fund programs in particular as examples, the Toolkit provides nuts-and-bolts guidance to help build capacity in conducting financial capability programs for **researchers** who are interested in conducting an evaluation of a financial capability program and for **policy makers** and **practitioners** interested in commissioning an evaluation. It will also be useful to **evaluation researchers** who want to brush up on a research technique they are less familiar with or who are new to the area of financial capability and financial education, particularly in a low- or middle-income country.

2 MONITORING AND EVALUATION

The use and value of monitoring and evaluation

The ultimate goal of financial capability programs is to improve participants' financial capability, but how do we know if such a program accomplishes what it set out to do? For that matter, how do we know if it is delivering what is called for at the right time and in the right place, if it is being implemented as program designers envisioned, and if the impacts achieved are worth the cost of doing it?

Such questions are typically answered through **monitoring and evaluation**, which is not a single process but rather two distinctly different processes, although they may be linked to one another.

What is monitoring?

Monitoring involves regularly collecting, analyzing, and reviewing information about how a financial capability program is operating. It is primarily descriptive and can provide rapid, real-time insights into a program as it develops and on the progress it is making toward meeting workplans and milestones. It can also be used to track the use of a program's budget and to identify any emerging threats to the program's success. Monitoring information may be collected and brought together from a range of sources, including both the organization planning the program and those charged with delivering it. It may also involve third parties, particularly if monitoring changes in participants' behavior is desired.

What is evaluation?

Evaluation assesses how well a financial capability program is performing and how much success it is having in achieving its intended outcomes. If an evaluation is conducted during the start or in the midst of a program, it is typically referred to as a **formative evaluation**; its primary goal is to provide feedback on program delivery to inform the evolution of the program itself. An evaluation conducted after a program has been established or concluded is called a **summative evaluation**. It may focus on how well the program met its objectives with respect to deliverables, but also on the outcomes and impact of the services delivered.

Process evaluation

A process evaluation addresses a broad range of questions about the nature of a financial capability program's development and implementation—including its relevance, efficiency, and effectiveness—with an emphasis on lessons learned: what worked and did not work as the program and its services were delivered.

Impact evaluation

An impact evaluation is focused on outcomes—the causal effect of the program and its services on observed outcomes. In the case of a financial capability program, impact will refer particularly to changes in behavior among the beneficiaries, but also to changes in knowledge, skills, and attitudes related to financial decision making that underlie behavioral changes.

Table 1 compares process and impact evaluations.

TABLE 1 PROCESS EVALUATION AND IMPACT EVALUATION

GOAL	EXAMPLE QUESTIONS	EXAMPLE ACTIVITIES	TIMING/ FREQUENCY
Process evaluation			
Document and assess implementation, operation, and outcomes against the goals and objectives of the program	<ul style="list-style-type: none"> ▪ Were the goals and objectives appropriate? ▪ Was the program implemented as intended? ▪ Were the goals and objectives met? ▪ Was the quality of the implementation adequate? ▪ Did the program reach the intended beneficiaries? ▪ Were the intended outcomes achieved? If not, why not? And what lessons can be learned? ▪ Did the program have unplanned or unintended effects? 	<ul style="list-style-type: none"> ▪ Document administrative process ▪ Review all monitoring data for trends and patterns ▪ Interview people involved in the planning and delivery of the program ▪ Interview beneficiaries in a survey or in-depth interviews ▪ Directly observe the program 	Ad hoc, at inception at an interim point, and when service has been established
Impact evaluation			
Measure the causal effect of program	<ul style="list-style-type: none"> ▪ What impact has the program had? ▪ Does the impact justify the cost? 	<ul style="list-style-type: none"> ▪ Conduct baseline and follow-up survey on treatment and control participants 	Ad hoc, when service has been established

Cost analyses

Funders often face many competing needs for resources. Those who are financially supporting a program need to understand whether the program’s delivery and performance provide more value for their money relative to other alternative programs or whether the resources spent would be put to better use elsewhere. Thus, evaluators and other stakeholders (such as funders) and policy makers often also need to **assess the delivery and achievements of a program against the costs incurred in providing those, relative to other alternatives**. There are three different types of cost analyses: cost-benefit, cost-effectiveness, and cost-consequences; these are described in table 2.

TABLE 2 PROCESS EVALUATION AND IMPACT EVALUATION

Cost-benefit analysis	Weighs the total costs incurred by a financial capability program against the total benefits (outcomes), measured in monetary terms, to come up with the net benefit of undertaking the program
Cost-effectiveness analysis	Measures the cost per output or outcome in a financial capability program and usually compares this cost with other similar programs or other ways of achieving the same outcome
Cost-consequences analysis	Enumerates and characterizes all the relevant costs and benefits for a financial capability program, using quantitative (monetary) costs and benefits where possible and qualitative descriptions otherwise

When and how should M&E be undertaken?

Ideally, one would decide at the outset of planning a program that an evaluation will be carried out; this means the evaluation would be **prospective**. If evaluators are involved at the very beginning of a program or as part of a pilot test, developing the M&E framework and the program planning can go hand in hand. Of course, not everything follows the ideal. Often, planning turns out to be **retrospective**; this means that program officials or external stakeholders decide to evaluate an ongoing or even a completed program after the fact. While retrospective evaluation is not ideal, it does not mean it is not doable.

While we often think of research as being conducted by trained evaluators independent of those being evaluated, that is not always the case. Many research activities are participatory in nature, where evaluators work directly with stakeholders to seek change and thus involve them in evaluation processes or inquiries. This is obviously the case when evaluation is prospective, but it is also true for retrospective evaluation.

When should an evaluation be done?

A strong strategic case for evaluation can be made if any of the following are true of the program:

- Is the program in question innovative? Has the impact of similar programs never been tested?
- Is it a pilot program that will later be under consideration for expansion?
- Will the program results be useful in helping design and deliver the program in the future? Is it replicable and will others want to learn from it?
- Does the program involve significant resource allocation?
- Can strong and sustained stakeholder commitment be expected if the results are positive?

If the answers are no, then careful thought should be given to the scale and scope of evaluation. It does not mean that an evaluation is not needed, but rather that a more modest approach may be called for.

3 SETTING THE STAGE FOR MONITORING AND EVALUATION

The importance of having a map in place

Regardless of whether evaluators carry out monitoring, a process evaluation, an impact evaluation, or all of the above, all M&E efforts must start by formally laying out and providing a context for the financial capability program goals and objectives: setting the stage for what is coming. If we do not know what the financial capability program is trying to achieve (its goals) or how we plan to concretely achieve those goals (its objectives), then we cannot expect to be able to determine if it is being carried out effectively (through a process analysis) or if it is having the desired impact (through an impact analysis).

Thus, while it is tempting to hit the ground running, having a clear map of each program component and the required indicators is fundamental in generating an overall M&E framework that systematically captures program components and the types of information that evaluators would in theory need to be able to assess each component.

Formalizing program goals and objectives

A critical first activity is to step back and get a better understanding of the problem the financial capability program is trying to address and the potential ways to solve it. This process is often referred to as a **resources and needs assessment** or a **general situation analysis**. Such analyses—which can include, for example, a literature review of existing research related to the substantive area of the planned program and/or its target population—narrow down the specific problem to be addressed, the target populations to focus on, and what makes the most sense in terms of desired outcomes to be measured.

With the situation analysis done, the next step is to **clearly define the program goals and objectives**. While they may sound like the same thing—and indeed are aligned and mutually reinforcing—they are actually very different. **Goals** are broad program aims that are not necessarily expressed in concrete, quantifiable terms; while **objectives** are best defined as operationalized goals—what needs to be done in practice to actually achieve the broader program goals.

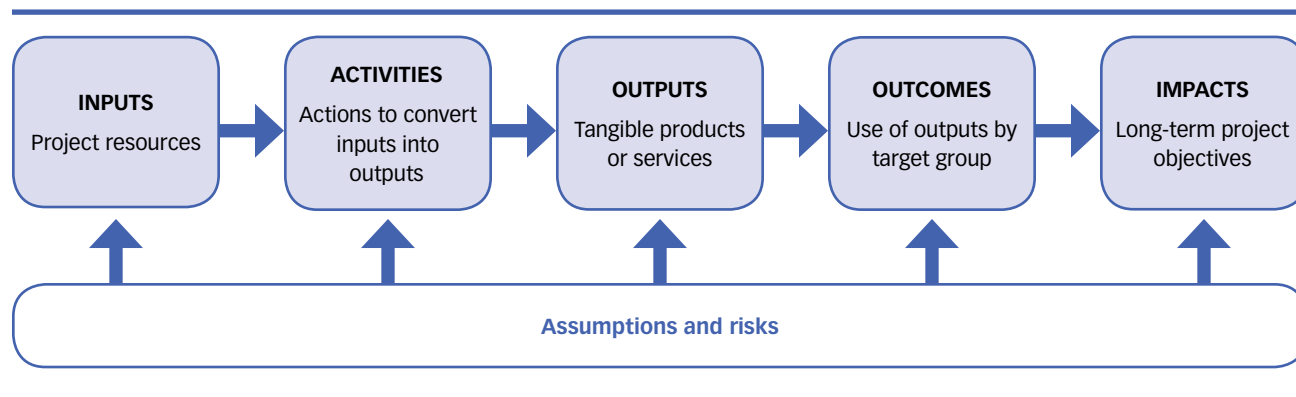
Assembling the M&E map

The link between a program’s design and its goals can be understood through the program’s **theory of change**—a logical, sequential argument for how and why an intervention will deliver the desired results, together with any assumptions necessary at each step of the way. Every financial capability program should be based on an underlying theory of change.

How does one depict a program’s theory of change in a way that is useful for program evaluation? A **results framework model** is one of the simplest and most useful ways to do so (figure 2). Such a model explicitly maps the planned program logic in terms of the program’s expected inputs, activities, outputs, outcomes, and impacts, as well as key assumptions, preconditions, and risks that could affect the project outcome.

In designing a program from scratch, it makes most sense for evaluators to start with objectives and then work backward to the input requirements (iterating the process

FIGURE 2 AN EXAMPLE OF A RESULTS FRAMEWORK MODEL



as needed). However, for an existing program where evaluators are evaluating retrospectively, it is often easiest to begin by making a list of all program activities and progressing from right to left, linking the inputs to the objectives.

Developing effective indicators

Going down deeper into the development of a robust M&E system, it is crucial to have a set of high-quality **indicators**—quantitative or qualitative variables that measure achievement or progress toward a specific goal or objective. Once the program components in the results framework are mapped—the inputs, activities, outputs, and outcomes—indicators should be determined for each component, from inputs to outcomes.

Standard indicators are often a good starting place for developing more customized ones, as long as evaluators avoid using standardized indicators when their content does not match that of the program. For example, there are two important and highly relevant new resources from the World Bank that evaluators can draw upon for their own work.

The financial capability measurement work funded by the Trust Fund and managed by the World Bank developed new methods for assessing levels of financial capability in a way that is relevant for low- and middle-income environments and that can be used consistently across countries to conduct international comparisons. The **measurement instrument developed by the Trust Fund** provides a diagnostic tool that evaluators can use to identify the key areas of financial capability (behavior, skills, attitudes, and knowledge) and which of these areas need improvement. The survey instrument can also be used to inform the design of targeted interventions in the area of financial education and financial capability enhancement by helping identify especially vulnerable groups in terms of financial capability.

An important category of outcome indicators is financial inclusion. While systematic indicators of the use of different financial services had been lacking for most economies, particularly in low- and middle-income environments, the **Global Financial Inclusion (Global Findex) database** released in April 2012 provides such indicators. This short list of key indicators covers a range of financial behaviors on saving, borrowing, making payments, and managing risk.

4 MONITORING

The importance of monitoring

Making sure that a financial capability program stays on course and adjusts quickly and in real time requires continuous and systematic tracking of performance over the program cycle. This is a key goal of monitoring. Monitoring a financial capability program from its inception helps ensure that those who manage it—along with policy makers and other stakeholders—can see that workplans and milestones are proceeding as planned in terms of how that program is delivered. If plans and milestones are not going as planned, monitoring can help ensure that plans are in place to rectify any problems. As such, **monitoring can serve as an early-warning system for identifying possible obstacles and can serve as tool to communicate with stakeholders and program staff.**

Beyond the value it offers in keeping programs on track, monitoring can also contribute to evaluations. Specifically, activities for monitoring and process evaluation often share information, infrastructure, and staff, which means that a robust and

explicit monitoring plan can not only help manage day-to-day program operations, but can also support other evaluation activities—and vice versa.

Building a monitoring plan based on the results framework

An effective monitoring system requires a plan that describes how the financial capability program will be implemented as it is rolled out over time. A results framework is a useful way to get organized for monitoring. For broad categories that are part of a results framework—inputs, activities, outputs, outcomes, and final outcomes—the monitoring plan should explicitly state the objectives, indicators, frequency of reporting, source of information, assumptions and risks, and responsible staff; table 3 presents a template for such a plan.

TABLE 3 EXAMPLE MONITORING IMPLEMENTATION PLAN

CATEGORY	INDICATORS	FREQUENCY	INFORMATION SOURCE	ASSUMPTIONS & RISKS	STAFF
Inputs					
Activities					
Outputs					
Outcomes					
Final outcomes					

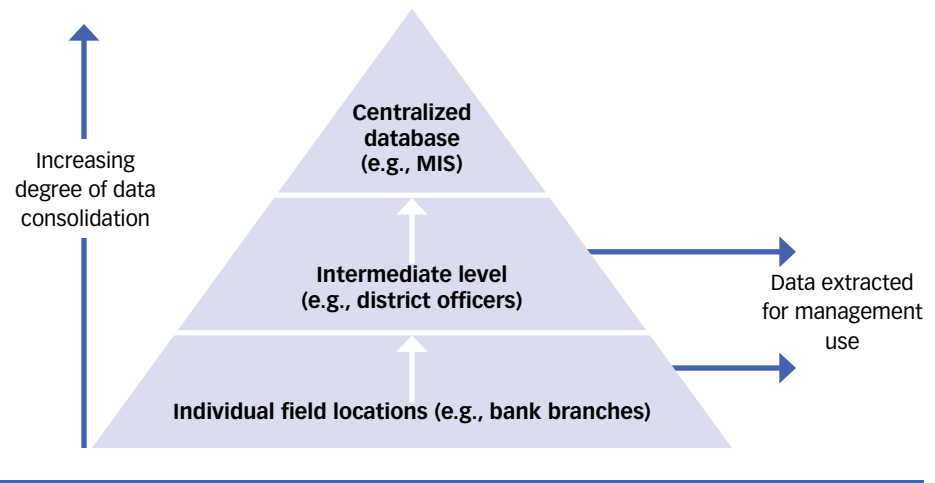
To ensure that the monitoring plan can work in practice and that the responsibilities are fairly and reasonably distributed, the plan must be reviewed together with the designated staff for each step. Do staff members understand their roles and activities? Will they have the necessary skills, time, and resources to carry out the monitoring tasks? Can the required work be made easier without compromising the quality of information (e.g., by reducing the frequency of data collection)?

Ensuring a clear and appropriate flow of information

Not surprisingly, the output of the monitoring plan is a large amount of data and information. The monitoring system should provide each level of the program with the information necessary to manage the relevant operations, but it also needs to avoid information overload. Different staff need different amounts of detail. Thus,

the system must aggregate and consolidate reports for higher-level managers but provide detailed feedback to field staff. Striking this balance is sometimes difficult and is made all the more difficult because there is a risk that useful details are lost. But this decision can be facilitated with protocols for data entry and reporting, with quality checks and validations along the way. Although this can be done through a computer-based management information system (MIS), a good paper-and-pencil approach can be quite sufficient if it is well thought through (figure 3).

FIGURE 3 PYRAMID OF MONITORING INFORMATION FLOWS



Getting the most out of the monitoring system

A good monitoring system is not a static system that is built and then left alone; it is constantly reviewed and adapted to guide project operations, track financial resources, provide managerial decision support, and facilitate communication with stakeholders.

The system can be used to report to stakeholders and provide feedback and guidance to program staff at all levels. In this latter role, the feedback process can take the form of automatic trigger points when the data and information collected are reviewed; such trigger points can be warnings when implementers have missed their targets or incentives when implementers have met them.

5 PROCESS EVALUATION

Assessing program relevance, efficiency, and effectiveness

In any setting, it is useful for program managers and stakeholders to understand how well implementation proceeds relative to the program's goals and objectives, how participants feel about the program and the services being delivered, and how any barriers to effective implementation can be addressed quickly. This is the domain of **process evaluations**, which evaluate the program's development and implementation, including its relevance, efficiency, and effectiveness.

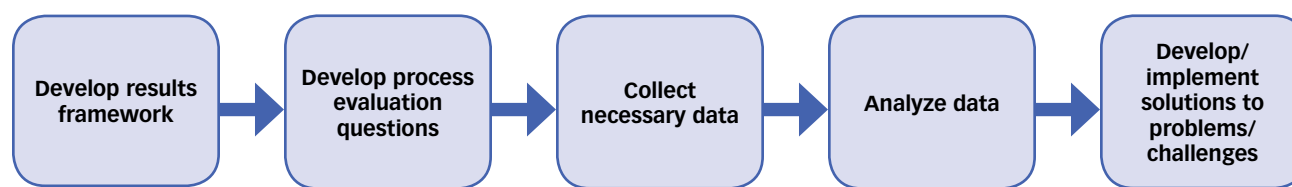
Clearly, there are overlaps between what monitoring does and what process evaluations do, but there also differences. While monitoring continuously assesses whether inputs, activities, and services are being delivered as anticipated and in a timely manner, process evaluations focus more on the effectiveness of those items relative to the program's goals and objectives and on whether it is being implemented as intended by the program's designers.

Process evaluations are particularly useful when a program is being piloted or in the early implementation stages; that way, if barriers are impeding effectiveness, improvements can be made before scaling up to a full implementation or can be altered along the way in full implementations. In terms of financial capability programs in particular, process evaluations are valuable in helping practitioners identify emerging challenges and areas for improvement when such programs involve new or complex interventions, such as multiple stakeholders or services; deal with sensitive issues; or take place in conflict areas.

Following a logical flow

The challenge in designing process evaluations is to ensure that they are valuable to the program but not too burdensome to implement. So, when it comes to actually conducting a process evaluation, it makes sense to follow a logical flow (figure 4) that begins with developing the results framework to conceptualize the intervention

FIGURE 4 STEPS IN CONDUCTING A PROCESS EVALUATION

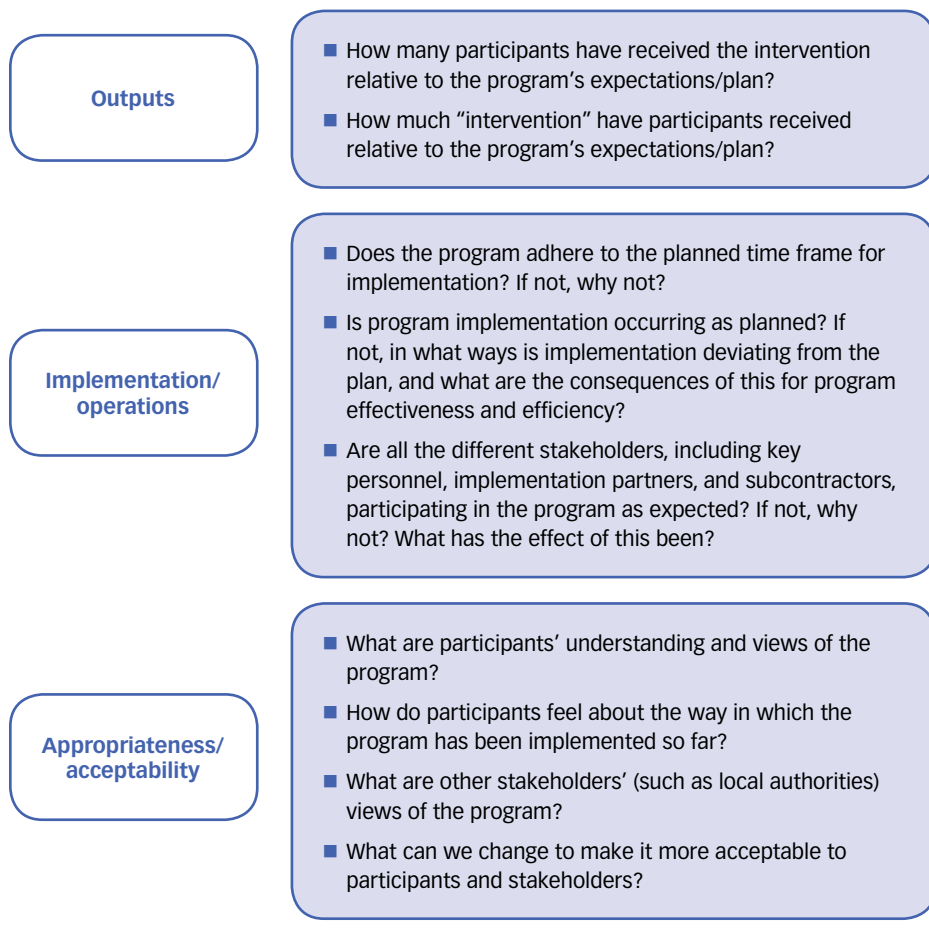


and reflect the program’s components (inputs, activities, outputs, and outcomes), what it is expected to achieve, and how it is expected to achieve it. The results framework can be helpful in identifying the right questions and the best information to answer them—which in turn leads to data collection, data analysis, and the development of solutions to any problems or challenges identified.

Asking the right questions to drive data collection

One of the advantages of driving a process evaluation by a results framework is that it helps ensure that evaluators ask the right kinds of questions. Generally, the questions asked in a process evaluation fall into three main groups: **outputs**, **implementation/operations**, and **appropriateness/acceptability** toward participants and other stakeholders (figure 5).

FIGURE 5 SAMPLE QUESTIONS FOR PROCESS EVALUATION



Asking the right questions, in turn, drives the data collection effort. Monitoring data can often be a useful input to the data that need to be collected for process evaluations. Such monitoring data are particularly useful for questions around outputs, processes, and even finances. Depending on the different questions, the other relevant data for process evaluations will either be qualitative or quantitative:

- **Qualitative data** are the main type of data collected for process evaluations and are used to answer questions about the participants' or implementers' experiences with the program and the acceptability of the program operations. This information often comes from interviews and focus groups with participants or implementers, from the review of program-related documents and materials, and from audit studies.
- Additional **quantitative data** can provide information beyond what is routinely collected by the monitoring system. These could include short surveys of participants.

Identifying and acting on potential problems

Process evaluations will logically reveal what is working well and what is not in implementing a financial capability program. The strength of a process evaluation is in its ability to provide implementers, practitioners, or other stakeholders with a comprehensive picture of how well the program is progressing relative to expectations.

Because process evaluations tend to be formative rather summative—that is, occurring as the program is being rolled out as opposed to after it is completed—it is important to remember that simply identifying problems or issues is not the ultimate goal. There is, in short, a natural feedback loop in such evaluations. Identifying problems early on is the first step toward addressing them and thereby improving the program's operations—and ultimately, the program's success. The insights from process evaluation can help tweak processes and resources and engage with stakeholders to make targeted improvements.

6 IMPACT EVALUATION

Impact evaluation: teasing out a financial capability program's effect

While a process evaluation answers questions about a financial capability program's development and implementation, an impact evaluation is a powerful tool designed to answer the question "What changes in outcomes are **directly attributable** to the intervention alone?" The key concern is being able to tease out whether the financial capability program directly causes the changes in the outcomes observed or whether those changes are the result of other factors in the environment. The aim of quantitative impact evaluation methods is to remove the effect of factors that confound the association between the financial capability program and financial behaviors as much as possible to establish a causal link.

To accomplish this goal, evaluators compare the outcomes of two groups of individuals: the **treatment group** (defined as participants who receive the financial capability program) and the **comparison** or **control group** (defined as similar participants who do not receive it). There are a number of ways to form the two groups of interest, which can be separated into two categories: (1) **experimental methods**, in which program participants are randomly assigned to the treatment and comparison groups; and (2) **quasi-experimental methods**, in which statistical methods are used to mimic random assignment.

While experimental designs are preferable when they are feasible to implement, quasi-experimental designs can also provide a reliable estimate of the causal impact of a financial capability program (table 4).

TABLE 4 OVERVIEW OF IMPACT EVALUATION METHODS

EVALUATION METHODOLOGY	IMPLEMENTATION TIMELINE	CHANGE CAN BE ATTRIBUTED TO INTERVENTION	STRENGTH OF METHODOLOGY
Experimental designs			
Randomized control trial	Prospective	Yes	****
Encouragement design	Prospective or retrospective	Yes	***
Quasi-experimental designs			
Regression discontinuity	Retrospective	Yes	***
Propensity score matching	Retrospective	Yes	**
Difference-in-difference	Retrospective	Maybe	*
Pre-/postcomparison	Retrospective	No	

Experimental evaluation methods

Experimental evaluation designs are those in which program participants are randomly assigned to the treatment and comparison groups. There are two types of experimental methods:

- **Randomized control trials.** Evaluators form the treatment and comparison (or control) groups by randomly assigning the financial capability intervention to a fraction of the eligible participants (considered the “gold standard” in impact evaluation design).
- **Encouragement designs.** Evaluators randomly assign who is encouraged to the financial capability intervention.

When implemented correctly, experimental methods are more credible than other methods in controlling for confounding factors. Also, the results from these methods are easily generalizable (conditional on the sample selection process), and they provide a transparent way to allocate programs with limited capacity. However, these methods must be implemented prior to the start of a financial capability program and can be costly to implement and maintain throughout the course of the program. While randomized control trials may not be feasible in some settings because it is not possible to exclude members from participating in an intervention, encouragement design provides an alternative experimental method, as long as the encouragement has the intended effect of increasing enrollment in the program.

Quasi-experimental evaluation methods

Quasi-experimental methods use statistical adjustments to mimic random assignment. Research has shown that quasi-experimental techniques can provide reliable results when the selection process is known and measured and/or when the treatment and comparison groups are matched on at least pre-intervention measures of outcome. There are three types of methods used:

- **Regression discontinuity design.** Evaluators assign participants to the financial capability program using the threshold of an eligibility index and take advantage of random error that places participants just above or just below the threshold.
- **Matching methods.** Evaluators use statistical methods to match the treatment group to a comparison group that is observably similar.
- **Difference-in-difference.** Evaluators estimate the change in the outcome of the treatment group over time, relative to the change in the outcome of the comparison group.

In general, quasi-experimental methods may be easier to implement because they do not change the implementation of the program and can be implemented retrospectively—that is, after a program has been implemented. But matching and difference-in-difference methods make strong assumptions about how the treatment group is formed, and these assumptions may invalidate the resulting impacts found.

Not all methods are created equal

While either the experimental or quasi-experimental designs discussed above can work in impact evaluations, some common research designs are relatively easy to put into practice but cannot provide good estimates of the causal intervention effect, because the counterfactual is not sound:

- Comparisons of outcomes **before** and **after** the program for those who participated
- Comparisons of outcomes between voluntary **participants** and **nonparticipants**

In the first case, there is no way to tell whether changes in outcomes resulted from the intervention or any other event that occurred at the same time. The second method suffers from **selection bias**, which occurs when some characteristics of the population cause them to self-select into a group (such as the treatment group), making it difficult to determine whether differences in outcomes are the result of these underlying characteristics or of the program itself. While these two methods are not useful for estimating the causal impact of a program, information collected from program participants can and should be used for monitoring and process evaluation, as long as the limitations of this information are understood.

Practical and logistical challenges to impact evaluation

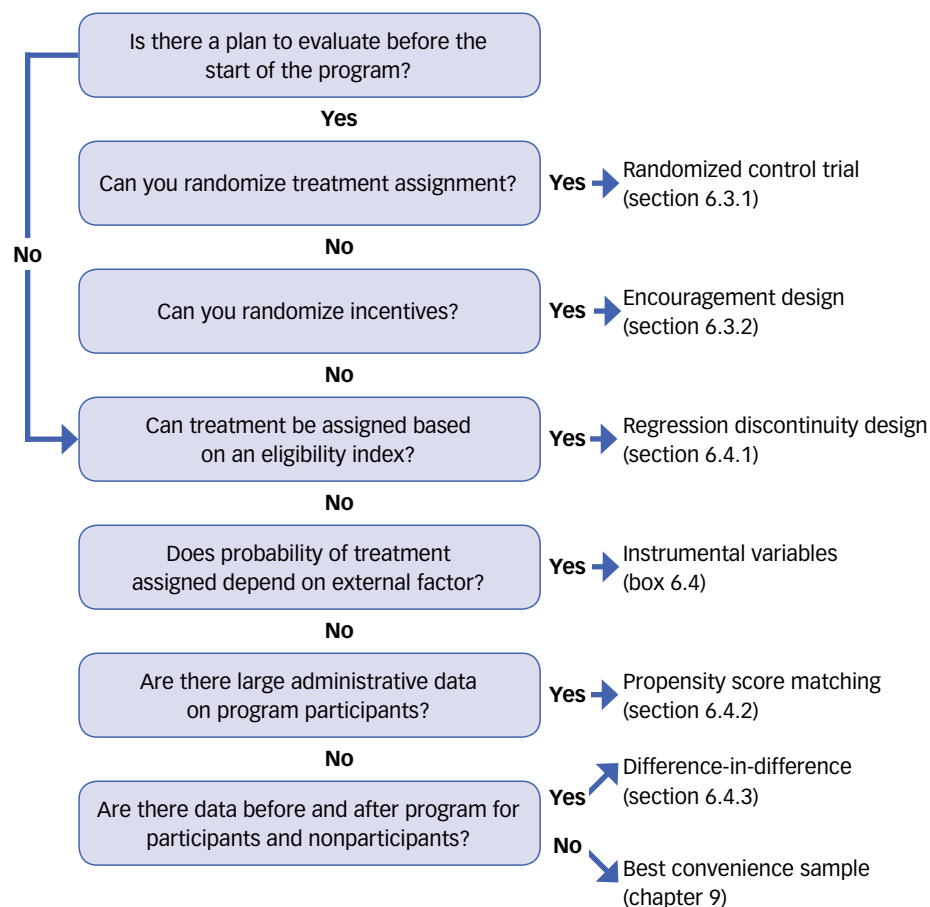
Often, even if a design is theoretically possible, practical challenges to implementation remain. While the types of challenges faced by evaluators depend on the evaluation methodology, they include **attrition** (loss of participants from the study), **spillovers** (participants in the control group are affected by the treatment), **partial compliance** (only a fraction of the participants assigned to the treatment group actually receive the intervention), **competition effects** (comparison group compensates for not receiving the intervention), and **Hawthorne effects** (participants modify behavior in response to being evaluated).

In summary

When it comes to doing an impact evaluation, there is a logical decision process policy makers can follow to choose the appropriate impact evaluation design for the program

setting (figure 6). And although all of the methods have some limitations, combining methods can help offset the limitations of any one method and create a more robust counterfactual. In practice, most of the methods described are used in combination.

FIGURE 6 DECISION TREE FOR CHOOSING IMPACT EVALUATION DESIGN



7 PUTTING IT ALL TOGETHER—COMPREHENSIVE EVALUATIONS

The importance of getting a bigger picture

While monitoring and process and impact evaluations are discrete events, in the real world, monitoring and the various forms of evaluation typically take place together as part of a **comprehensive evaluation** of a financial capability program.

By themselves, monitoring and evaluation are not going to provide the bigger picture needed to understand the implementation of a financial capability program. While an impact evaluation is necessary, it is not sufficient to achieve a full understanding of how a program performed. A process evaluation and monitoring are important parts of a comprehensive approach to evaluation, helping to answer the questions about how and why that ultimately make the overall evaluation’s results useful for program decision making.

A comprehensive evaluation lets us look inside the “black box” of the overall program effect to answer more basic questions about its theory of change and to help interpret the findings. This bigger picture is critical to understanding both program successes and shortcomings.

Answers to the kinds of questions shown in table 5 can help program implementers and stakeholders better understand what effects they are seeing and can make an evaluation valuable even if the program itself is not considered a success.

TABLE 5 QUESTIONS IN COMPREHENSIVE EVALUATION

<p>What led to a financial capability program’s successes?</p>	<ul style="list-style-type: none"> ▪ What were the underlying mechanisms that led to behavior change? ▪ What can this experience tell us about the most relevant barriers and constraints to financial capability to address in the target population, or other similar groups? ▪ What have we been able to learn about designing better programs in the future?
<p>What led to a financial capability program’s shortcomings?</p>	<ul style="list-style-type: none"> ▪ Why did it fail or fall short? ▪ Was there a fundamental conceptual problem (faulty program logic)? ▪ Were the participants inappropriate (faulty targeting)? ▪ Was it simply poorly executed (faulty implementation)?

Such comprehensive evaluations rely on a **“mixed methods”** approach that naturally arises in an evaluation strategy that combines impact evaluation, process evaluation, and analysis based on monitoring data. Combining different sources of information and evaluation methods allows us to assess both the “what” and the

“why” of program evaluation: Impact evaluation methods help estimate the magnitude of impacts, and process evaluation methods help shed light on the underlying causal mechanisms and processes.

When a comprehensive evaluation is most critical

Conducting a comprehensive evaluation using mixed methods that integrate both impact and process evaluations and the data they yield makes sense. But doing so can obviously go beyond the resources allocated for a program evaluation. Thus, if resources are limited, it is important to consider situations where it makes the most sense to strive to do one (table 6).

TABLE 6 WHEN TO CONDUCT COMPREHENSIVE EVALUATION

Conducting the exploratory (or pilot) phase of an evaluation	<ul style="list-style-type: none"> Quantitative impact estimates are important when key resource decisions are being made Qualitative data collection can be used to develop hypotheses, research questions, and survey instruments; to define indicators for quantitative evaluation; and even to design interventions themselves
Following up after an impact evaluation and understanding differences in program effects	<ul style="list-style-type: none"> Quantitative analysis in impact evaluations can establish average program effects by comparing a large group of participants to another comparison group To get at subgroups, quantitative data from a survey can identify different or heterogeneous effects among different groups and provide breadth, while in-depth case studies (a qualitative method) can help describe and understand these groups in greater detail
Increasing the validity of evaluation findings	<ul style="list-style-type: none"> Triangulation can provide more weight to controversial findings by using interviews or focus groups (qualitative methods) to confirm the findings from a survey (quantitative method)
Conducting case studies of small or very sensitive populations	<ul style="list-style-type: none"> Triangulation can combine surveys (quantitative method) to look at big questions while using case studies (qualitative method) to gain an understanding of important programs or policies and subgroups that are too small to survey but important to study

Things to consider before doing a comprehensive evaluation

Properly conducting mixed-methods evaluation can be challenging. Evaluators should be careful not to embark on conducting such an evaluation without a clear rationale for why this approach is necessary. They should also think carefully about the mix of skills and expertise in the evaluation team. If the team consists only of researchers with quantitative analysis skills, it may be difficult to successfully incorporate insights from qualitative data into the evaluation and interpretation of results and vice-versa. It may also be important to ensure that experts with different methodological strengths communicate effectively and cooperate well.

8 DATA COLLECTION METHODS

The Importance of getting the data—the precursor to analysis

In conducting M&E efforts for financial capability programs, evaluators can employ both quantitative or qualitative approaches (or, more likely, a combination of both) to determine how well those programs are being implemented and whether they are achieving their intended outcomes.

Ultimately, such determinations will be made only when evaluators analyze the data collected during evaluations, which means that such data collection underpins the ability to do any type of evaluation. Data collection efforts can support qualitative approaches (such as focus groups with participants) or quantitative approaches (such as surveys of participants). What types make sense—and in what combination—will depend on the program objectives and activities. While qualitative data typically goes with process evaluations and quantitative data with impact evaluations, both types of data can and do get used in both, as they do in mixed-methods approaches in comprehensive evaluations.

The value added of qualitative data

The ultimate value of qualitative data is that they complement quantitative data by providing a depth of contextual understanding and a level of detail that cannot be obtained with quantitative data alone. They can add value by **helping develop the quantitative research materials**, such as survey questionnaires—both in terms of the substance of the material on such questionnaires and question wording and syntax. As part of a **formative process evaluation**, qualitative data can add value by providing a detailed understanding of the development of the financial capability program as it is being implemented—in particular, program implementation and operational issues. As part of **summative research**, including both process and impact evaluations, qualitative data can add value by shedding light on the causal mechanisms of the program—by allowing us to ascertain how any observed program effects came to pass (or failed to do so).

The uses of qualitative data

At the highest level, qualitative data collection methods focus on gathering information on people's perceptions and experiences with the process of implementing a financial capability program and on the meaning they ascribe to particular events in that implementation. A number of qualitative data collection methods are typically used, each with its uses and value (table 7).

TABLE 7 TYPES, USES, AND VALUE OF QUALITATIVE DATA

QUALITATIVE DATA TYPE	USES AND VALUE
In-depth interviews	<ul style="list-style-type: none"> ▪ Use to explore in detail the expectations, experiences, views, and feelings of participants (through topic guides) ▪ Are particularly useful for discussing personal details or sensitive issues that people may be hesitant to discuss in a group setting ▪ Use to explore in detail the expectations, experiences, views, and feelings of participants (through interview protocol that specify wording and sequence of questions and provide interviewer prompts)
Semi-structured and cognitive interviews	<ul style="list-style-type: none"> ▪ Are particularly useful and appropriate for program evaluation, because reliance on interview protocol increases the comparability of responses and enable clearer insights into particular issues of interest ▪ Are particularly useful as cognitive interviews in developing quantitative data collection materials, such as a survey instrument
Focus groups	<ul style="list-style-type: none"> ▪ Use to discuss topic(s) in small group of people ▪ Are useful in eliciting a person's own views in a context in which the person can consider the views and perceptions of other people and contribute to them ▪ Are particularly useful when informants are too intimidated or otherwise unwilling to have one-on-one interactions with researchers
Desk review of documents/ materials	<ul style="list-style-type: none"> ▪ Use to provide information on a program's stated aims and objectives, design, implementation, processes, stakeholders, inputs, outputs, and results ▪ Use to document changes in a program during its lifetime that may be relevant ▪ Use to provide direct input to the assessment of the quality of a financial capability program
Audit or mystery shopping studies	<ul style="list-style-type: none"> ▪ Use to provide useful information on how a program is being delivered and whether delivery varies significantly depending on perceived characteristics of the beneficiary ▪ Use to study the behaviors of service providers in particular industries

The uses of quantitative data

Quantitative data are data that are either collected as numbers or that can be easily translated into numbers. **The key advantage of quantitative data is that they can be used to statistically test hypotheses and relationships.** Where qualitative research can help probe relationships between a financial capability program and observed outcomes, quantitative research can be used to describe, predict, and establish those relationships in the aggregate.

There are two broad categories of quantitative data. Researchers collect **primary data** first hand, directly from the subjects under study. Surveys, direct observation, experiments, and field experiments are all examples of primary data collection

methods. **Secondary data** are existing data that are collected by someone other than the evaluation team but can be used to help answer relevant questions. Administrative and census data are examples of secondary data. In many cases, evaluators will link administrative data with survey data to provide a richer and more complete answer to questions about the effects of financial capability programs.

SURVEYS

Surveys are typically the best data collection method for determining information about large populations, with a known level of accuracy. Beyond enabling evaluators to assess outcome indicators of interest, it is important for survey instruments to also collect background information on respondents because this allows evaluators to better understand the study population, control for various sociodemographic characteristics, and look for heterogeneous effects.

In developing a survey, it makes sense to (1) cognitively test questions to make sure they correctly capture the researchers' underlying intent, (2) ensure the same questions are presented identically across all respondents so that biases are not inadvertently introduced, and (3) explicitly consider the order of questions to avoid the problems of saliency and inadvertent framing.

SITE VISITS AND OBSERVATIONS

Site visits and nonparticipant observation can be either qualitative or quantitative or both. They can help program evaluators get inside the "black box" of program activities and see for themselves what is actually going on. How are the program activities being delivered, and how are the participants receiving them? While site visits may not in themselves be sufficient sources of information about a program, triangulating these descriptions with other data (such as from surveys, interviews, focus groups, and document reviews) can provide rich and nuanced insights for an evaluation.

Regardless of the type of information collected during a site visit and observation, it is important to collect it systematically, using either a template or topic list that has been compiled in advance. If observation techniques are used, it is very important that the people being observed are put at ease and feel comfortable with the observation; otherwise, what they do will be influenced by their unease, and the observations will be invalid.

EXISTING ADMINISTRATIVE DATA

Data are being collected all the time, independent of the data being conducted to evaluate a financial capability program, and those data can be helpful in supplementing primary data collection activities. It is common to merge or link administrative data sets that already exist with survey data that are being collected as part of the evaluation process.

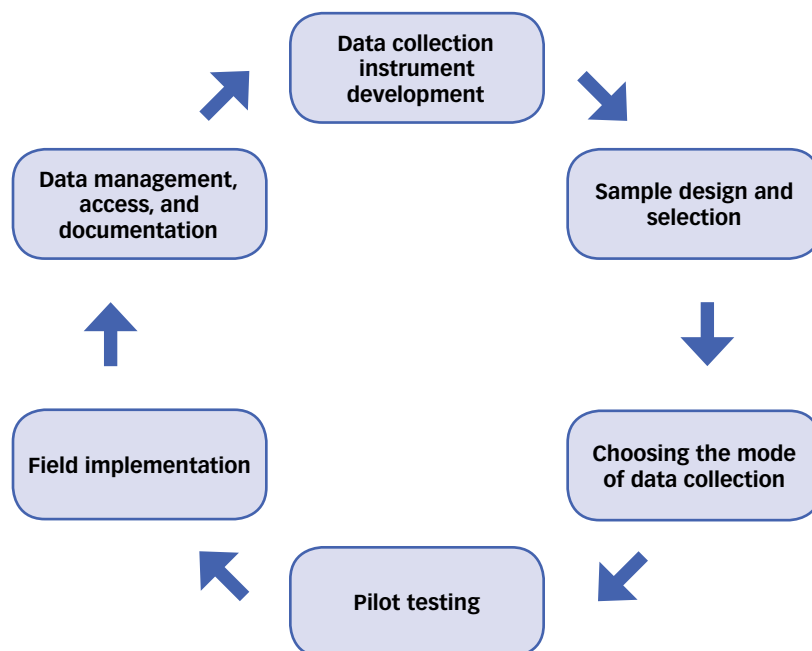
9 THE PROCESS OF COLLECTING DATA: PRACTICAL GUIDANCE

The importance of collecting data correctly

While collecting data is key to enabling evaluators to assess financial capability programs, the **processes** of actually collecting the data are also crucial. Those processes are central to getting the most out of the various data collection methods and ensuring the quality of any evaluation. How does one actually conduct a focus group? How does one select an appropriate sample population for a survey to ensure that the analysis of the resulting data will be both valid and useful?

If such data collection is not done correctly, it will undermine the study's validity or compromise the usefulness of its findings. Proper care must be exhibited at every stage of the data collection process, which includes designing survey instruments and interview protocols, selecting the right approach to sampling, preparing and training for data collection, implementation of the instruments in the field, and finally the storage and documentation of the data collected. All of these are critical to the process of collecting data (figure 7).

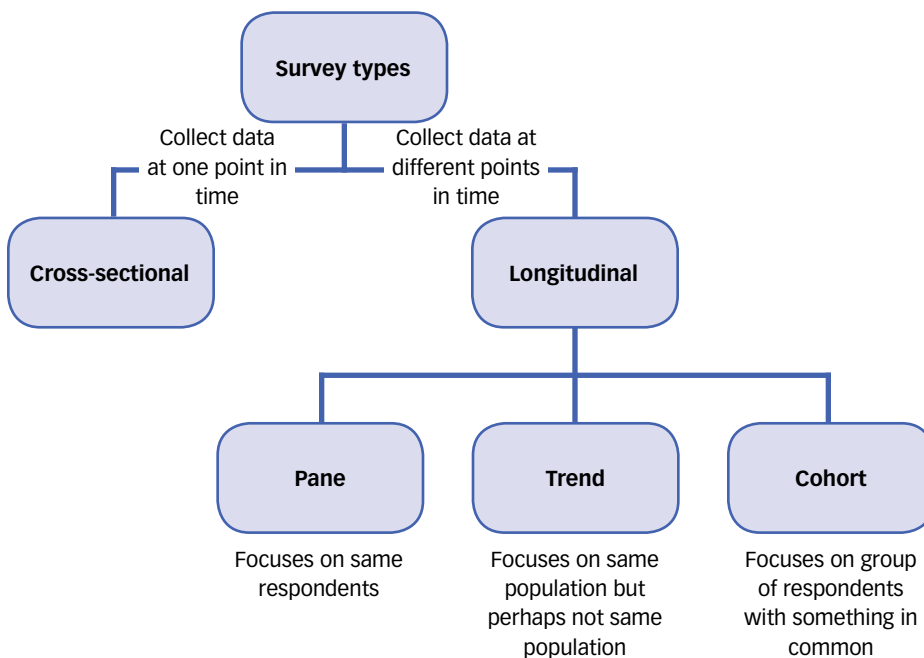
FIGURE 7 THE PROCESS OF DATA COLLECTION



Choosing a survey instrument design

The key in survey instrument design is the type of survey—cross-sectional or longitudinal—and, if longitudinal, the type of longitudinal survey, all of which depend on the survey’s intended purpose (figure 8).

FIGURE 8 SURVEY TYPES



The challenge of sampling design and selection

Drawing a good sample means that the chosen sample accurately represents the entire population of interest; this in turn means that the conclusions about outcomes that evaluators make from analyzing the survey data can be considered valid. Sampling design and selection applies to both quantitative data collection (e.g., surveys) and to qualitative data collection (e.g., in-depth interviews and focus groups), although sampling is a much more formal part of quantitative data collection.

Options for quantitative analysis include a **random or probability sample**, **systematic sample**, **stratified sample**, or **cluster sample**, all of which have pros and cons. The question of sample size in any option is very formalized. Evaluators need an explicit **power calculation** to derive estimates of the sample size for each design needed to convincingly distinguish between a lack of actual impact and an inability to statistically detect meaningful differences. Also, because data collected for impact evaluations are intended to make inferences about the larger population of interest,

sampling weights are used so the sample can be rebalanced to look like the community of interest.

When it comes to doing qualitative analysis, sampling is still important, but there are no widely accepted rules about adequate sample sizes. Because the aim of qualitative research is to provide greater depth of contextual understanding, rather than generate statistics that can be generalized to the population, sample sizes and design can vary depending on the specific aims, circumstances, and resources of a particular study. **Samples in qualitative research are intended to maximize information**; the sample size should strive to reach the point at which the information obtained becomes redundant.

Choosing the mode of data collection

Choosing the mode of data collection is another decision that often embodies trade-offs and requires careful consideration of specific local conditions. Data collection modes can affect the quality of data, but can vary significantly in cost and feasibility. Common modes are self-reporting through mail-in paper forms, telephone interviews, face-to-face interviews, and online surveys. Access to respondents through mail, telephone, face-to-face interviews, and the Internet may vary significantly in different settings and lead to sample selection bias. Whatever mode is chosen, it is important to consider maintaining the mode of data collection across respondents and waves, unless there is a compelling logistical constraint.

The value of pilot testing

Before going into the field in force with a data collection effort, it makes sense to pilot test the data collection instruments and procedures first. It is easy to underestimate the importance of pilot testing instruments and procedures in the field before deployment. Time and budget constraints sometimes make testing sound like a waste of precious time. But in almost all cases, even minimal testing can help evaluators refine the instruments they will use in their research and prevent problems later on, such as respondents not understanding questions or protocols failing to draw the information evaluators are after. Once evaluation tools have been pilot tested, the necessary adjustments can be made before full implementation.

Implementing the program in the field

Evaluations typically require managing a large team, so an important part of ensuring smooth coordination involves preparing and finalizing an overall **workplan** ahead of time with all participants prior to launching any field operations.

In quantitative data collection, the process tends to be much more formalized, including a detailed training and field manual, clearly articulated quality standards, and in-place quality assurance procedures. Without such standards, the effort can be poorly executed, which in turn can affect rates of nonresponse (refusal to answer questions) and attrition (refusal to participate further in the study) and increase the level of errors in measurement.

In qualitative research, the nature of the interaction with the respondent is critical. Because in-depth interviews and focus groups call for a high level of skill, it is important that those facilitating them (1) have substantial interviewing experience; (2) do not influence what the respondent says and, above all, allow and encourage the respondent to speak at length on the topics to be covered; and (3) have well-developed listening skills and be familiar with techniques to probe replies and encourage the respondent to elaborate.

The criticality of data management, access, and documentation

Once data are collected, proper measures must be taken to store data effectively and ensure that they can be disseminated to key users. Part of the appropriate management of such individually identifiable data is having in place the protocols and procedures for storing, transferring, and accessing such data. All organizations that collect data on human subjects should ensure that secure storage facilities are available and used. When individually identifiable data are to be shared, data users must have proper authority to access those data, data managers must be responsible for verifying the credentials of users, and data transfer systems should be secure.

10 ANALYZING QUANTITATIVE AND QUALITATIVE DATA

Getting to findings: analyzing the data

Assuming that the data for evaluation have been collected using the appropriate quantitative and qualitative tools, the analysis of the data will inform evaluators and other stakeholders of whether a financial capability program is being appropriately and effectively fielded and whether participants are improving in terms of the outcomes of interest that the program was designed to address.

The credibility of those results depends on how well evaluators conduct the analysis of the data collected. Whether evaluators are analyzing quantitative data collected during experimental or quasi-experimental research designs or qualitative data from interviews, document reviews, or focus groups, they must approach these analyses

rigorously and systematically, following the procedures and rules that govern the types of analysis.

Analyzing quantitative data

Quantitative data analysis for impact evaluations involves estimating the effect of a program by comparing the outcomes of the treatment group (which received the financial capability intervention) with the outcomes of the comparison group (which did not).

GENERATING DESCRIPTIVE STATISTICS

Once any quantitative data are collected, the first thing to do in analyzing these raw data is to generate a set of simple descriptive statistics. Doing so helps both evaluators and their eventual audiences begin to understand what is happening in the raw data. Evaluators should first report the sample size for the overall sample of data and then produce summary statistics of key demographic variables, such as age, gender composition, rural/urban status, education, and income (if reliable measures are available), either through tabulations in tables or visually in charts and graphs. Descriptive statistics can be categorized into the following measures.

- **Central tendency**, such as the mean, mode, and median, because these statistics are measures of the value around which the data appear to be clustered
- **Variability**, such as the range, quartile, standard deviation, and skewness, because these statistics describe how the data are dispersed

EXPLORING RELATIONSHIPS BETWEEN TWO OR MORE VARIABLES

Because we are often dealing with more than one variable, we often want to analyze the relationship between them (known as bivariate and multivariate analyses). The most basic form of bivariate analysis is **cross-tabulation**, which provides a table that shows the values of one variable against the other.

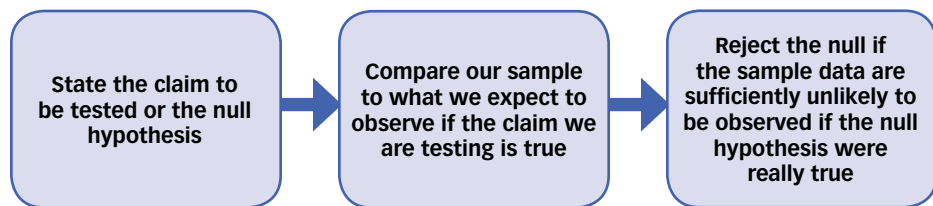
With two continuous variables, a visual representation of the relationship known as a **scatter plot** may be useful; it shows the relationship in terms of how the two variables “scatter” in the graph. The relationship between any two variables can also be captured by looking at the **correlation coefficient**, a standardized measure of the dependence between two variables which ranges from -1 to 1 . A zero value implies no relationship at all between the two variables, and a high absolute value indicates a strong relationship.

INFERENCEAL STATISTICS AND HYPOTHESIS TESTING

Descriptive statistics ensure that we truly understand the data, but evaluators also need **inferential statistics**, or the type of statistical analysis that allows them to draw conclusions about the causal impact of program participation on outcomes of interest.

One of the basic concepts in inferential statistics is **statistical significance**—defined as whether an event or occurrence is considered unlikely to have occurred simply by chance. Statisticians make this determination by setting up and formally testing the hypothesis or claim that the event did indeed occur by chance (figure 9). The statistical significance of an event is completely separate from the **practical significance** or size of the event. An event can be small in magnitude and statistically significant or large in magnitude but statistically insignificant (or statistically not different from zero).

FIGURE 9 HYPOTHESIS TESTING



REGRESSION ANALYSIS

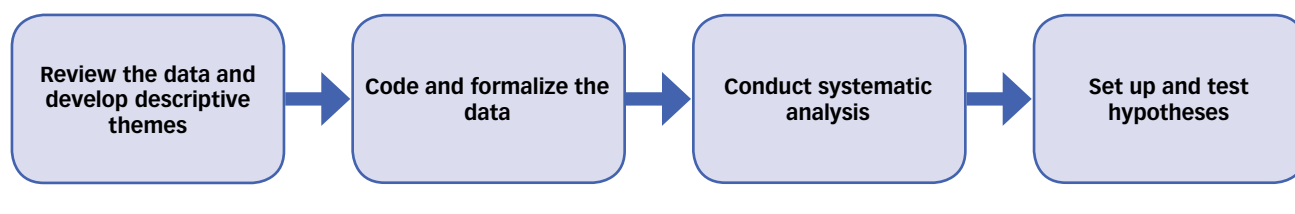
Regression analysis helps explain how the typical value of the dependent variable changes when any one of the independent variables is changed, while the other independent variables are held fixed. While a useful and commonly applied tool to understand the relationship between two variables, regression analysis has some challenges:

- It is demanding, because it requires quantitative data relating to a large number of individuals.
- To draw conclusions, there must be differences in the observed measures of the dependent and explanatory variables.
- Outliers (a few observations with values far away from all others) can compromise the accuracy of a linear regression as they can significantly affect mean values.

Analyzing qualitative data

While analyzing quantitative data clearly has a rigorous set of rules and procedures, analyzing qualitative data—while perhaps not as formally rigorous—must still be done systematically and in a way that others can assess and replicate (figure 10). In terms of process evaluations, this means thinking about the analysis as the fieldwork

FIGURE 10 A RIGOROUS APPROACH TO QUALITATIVE ANALYSIS



is under way and producing analytical notes as one goes along; being systematic and deriving a structure from and for the data collected that is interpreted flexibly throughout the analysis; and aiming to go beyond the purely descriptive to provide a depth of understanding or to generate hypotheses that can be tested quantitatively. In process evaluations, the analysis of qualitative data requires that evaluators have a very thorough understanding of the program under scrutiny to ensure that findings are interpreted appropriately.

Descriptive themes—typically, common patterns, ideas, and concerns that emerge throughout the data—are based on the initial research questions and aid in organizing findings. Data analysis does not need to wait until all data have been collected. It is important to look for themes and patterns in the data even at the fieldwork stage.

Coding and formalizing the data consists of identifying the key themes in the data and labeling the data according to these themes. Evaluators should be looking for patterns in the data: for similarities and consistencies across documents and differences and inconsistencies, with the goal of explaining why and how these occur. At the end of this process, the codes created can then be organized into a hierarchy of ideas (often known as a **code frame**).

The coding of data can be performed in a number of ways—such as annotated transcripts, computer programs, or thematic grids—but all should be systematic.

Annotated transcripts may involve creating folders, one for each issue or theme, where sections of text from transcripts are assigned to one or multiple folders. This approach is very flexible and allows for modifying, combining, or splitting categories as the analysis proceeds and new insights emerge. **Computer programs** help emulate this process, creating an initial, largely descriptive, electronic coding frame and then refining this into one that is both analytical and hierarchical. Evaluators may also use **thematic grids**, which summarize the transcripts, with the columns corresponding to the hierarchical code frame developed and transcripts or documents entered in the rows. Within the context of process evaluations, analysis of qualitative data will inform evaluators and stakeholders on how well a financial capability program has been implemented.

11 COST ANALYSIS

Laying out the costs and benefits of financial capability programs

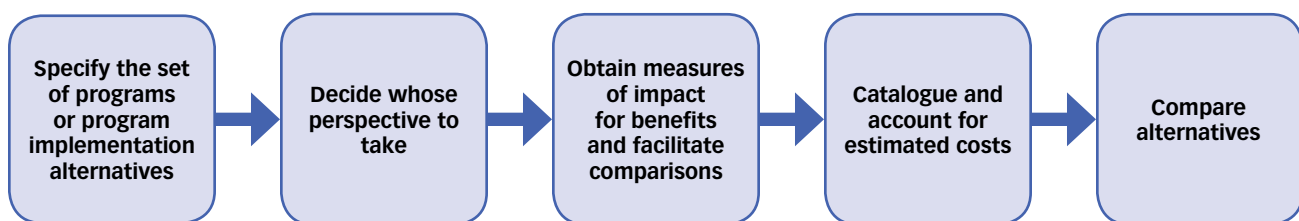
While evaluation often focuses on whether a financial capability program has achieved the program's objectives, there are costs to implementing any program, and if policy makers and stakeholders want to understand whether their investment in a particular financial capability program was worth it—especially when resources are scarce—then evaluators will also need to conduct a cost analysis. Determining whether a program was worth it typically revolves around two questions:

- Did the benefits of the program outweigh the costs of undertaking it?
- Did the program perform efficiently when compared to other options addressing the same program goals and objectives?

The basic steps of any cost analysis

Regardless of the type of cost analysis conducted, evaluators should start by **defining the alternatives under consideration** (figure 11). In many instances, it may be obvious what these alternatives are—a comparison of two programs, say—but in some cases, a program may have any number of dimensions and variations, and these variations will most likely change the program's impact and costs.

FIGURE 11 BASIC STEPS FOR COST ANALYSIS



The next step is to **decide whose perspective to take**, because a financial capability program's costs and benefits mean very different things to different stakeholders—participants, the program itself, the government, and society as a whole. The evaluator must define the stakeholder(s) from whose perspective the analysis will be conducted before being able to identify the impacts of relevance and catalogue the relevant costs. If multiple stakeholders are of interest, it may be best to do a separate cost analysis for each.

Like any other form of evaluation, the value of a cost analysis relies on what goes into it—in this case, the accuracy of **measuring the estimated benefits and costs**. Evaluators must consider only those impacts directly attributable to a program to accurately reflect the trade-off between costs and benefits. It may seem that estimating benefits is far more complicated than estimating costs, especially if programs report their total budgets.

But getting at the true **economic costs** of a program can be more complex than it first appears. Some costs may be straightforward, such as dedicated program spending. But others may not, including the value of inputs funded by partners and those donated or volunteered (e.g., building space to conduct the program or computers). There will also be **opportunity costs**—the value of the best possible alternative use of all program resources. In other words, if the program were not in place, resources used would be deployed elsewhere. The key here is to think about costs in a comprehensive way, such as in terms of a checklist of important categories of program inputs.

While the first steps are generic and relevant to all types of cost analyses, the final step involves **understanding and comparing the different kinds of cost analyses** and what they entail (table 8).

TABLE 8 TYPES OF COST ANALYSES

	COST-BENEFIT ANALYSIS	COST-EFFECTIVENESS ANALYSIS	COST-CONSEQUENCES ANALYSIS
What are the alternatives for comparison?	One or more programs to a threshold or benchmark for decision making	Two or more programs	One or more programs to one another
How many outcomes?	Multiple	One	Multiple
How do we make comparisons?	Convert all outcomes to a dollar value; then compare costs per dollar benefits	Select a common outcome measure for each intervention and compare costs per unit outcome across all programs	List and characterize all costs and benefits across all programs (qualitatively and/or quantitatively)
Result	Ratio of costs to benefits, “return on investment,” or “net present value”	Ratio of cost to common outcome measure	Costs and benefits displayed in tabular form

- **Cost-benefit analyses** monetize all program impacts, including intangibles, something evaluators do by estimating the different returns to financial market participation, such as access to credit, different savings mechanisms, and financial security.
- **Cost-effectiveness analyses** compare multiple programs in terms of their impact on a single, common outcome, which may be expressed in monetary terms or other units.

- **Cost-consequences analyses** simply enumerate and characterize all relevant costs and benefits for the alternative programs, side by side, numerically where possible and qualitatively where not.

The three types of cost analysis differ in terms of the alternatives for comparison, how many outcomes they evaluate, how the comparisons are made, and the result. In doing any type of cost analysis, be sure to estimate all the costs and benefits accurately, both the straightforward and tangible ones as well as the less straightforward and intangible ones. Not considering all the costs and benefits when doing a cost analysis undermines the usefulness of the results.

12 IMPLEMENTING THE EVALUATION

Translating the evaluation into action on the ground

Evaluations are the key drivers in ensuring that we are able to understand whether a financial capability program achieved the impacts it set out to accomplish. Having a sound evaluation plan that carefully considers all the things that are critical to an evaluation and all the things that could go wrong (along with mitigation plans to deal with them) will ensure that the evaluation ends up being effective.

Evaluation logistics and timing

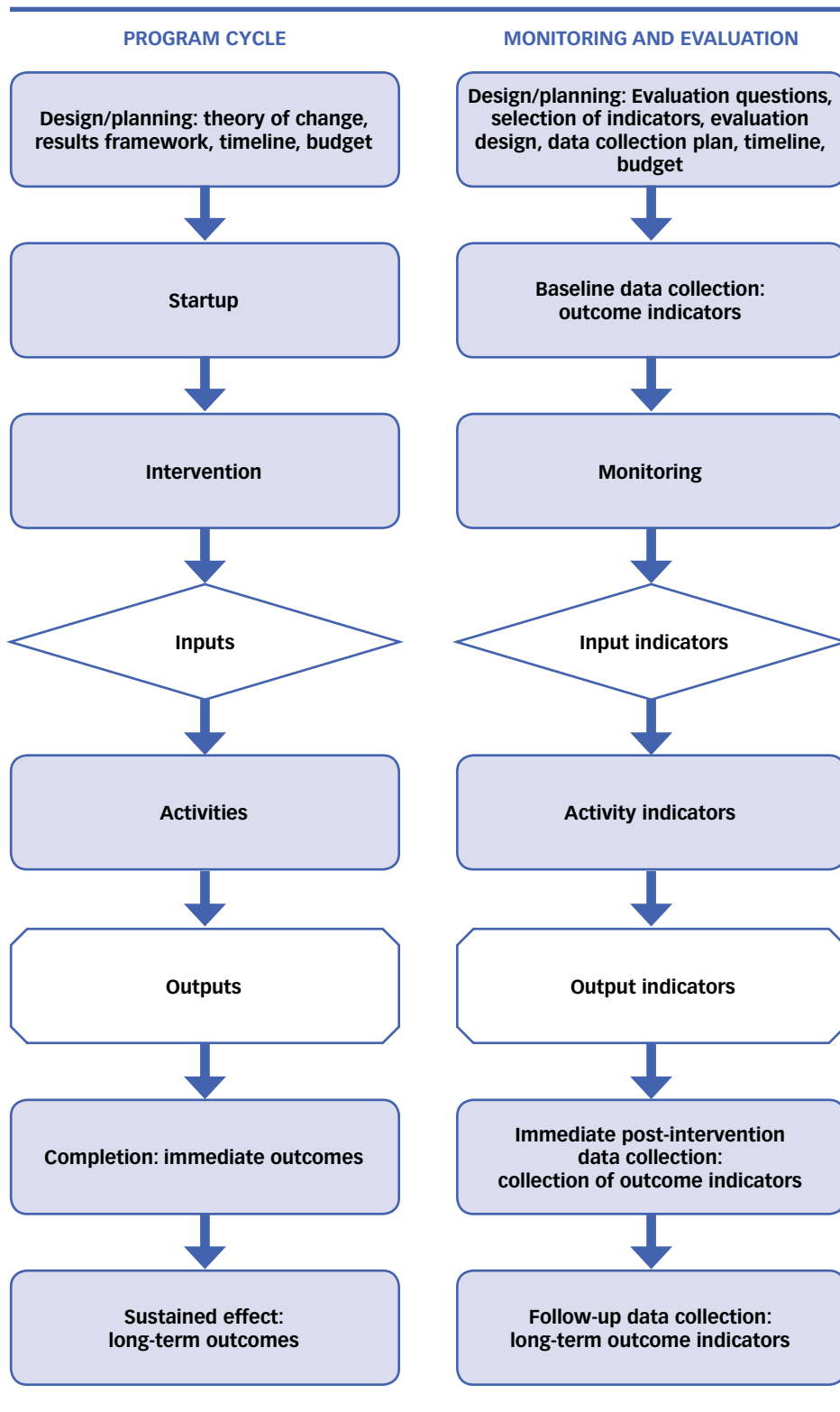
Four key factors that should be considered for the logistics and timing of the evaluation are the program cycle, the reasonably expected time required for results to manifest themselves, the logistical constraints of fieldwork, and internal or external decision points that affect the program. Just as the evaluation design should be fitted to the program design, evaluation timing also needs to be fitted to the program cycle (figure 12).

Forming an evaluation team

Once the general scope of the evaluation is determined, it is possible to consider the range of skills and roles needed to design and carry it out. The process of staffing the team; ensuring a clear and mutually agreed-upon set of roles and responsibilities; and setting up a system for discussion, ongoing communication, and feedback can be critical to evaluation success.

Individual team members may play multiple roles based on capacity and resources. In such cases, it is helpful to combine roles that have overlapping tasks but ensure that each person can receive support and feedback from other team members.

FIGURE 12 MEASURING PERFORMANCE AND EFFECTIVENESS:
EXAMPLE OF LINKING TO THE PROGRAM CYCLE



Important decisions must also be made about the allocation of internal and external responsibility. Using external evaluators can help achieve the right mix of technical skills and establish the objectivity and credibility of evaluation, but there are important reasons to keep internal stakeholders engaged.

Estimating total evaluation costs

Not surprisingly, evaluations are highly resource intensive, which is an important reason why many programs are not evaluated. Because evaluation costs can vary so widely, it is best to develop a preliminary budget based on rough but relevant information rather than to proceed on assumptions that may be very far from reality. Important cost categories to consider in estimating total evaluation costs are staff time, travel and subsistence costs, and data collection costs.

Planning for contingencies

An important part of evaluation planning is to carefully consider the risks and challenges likely to occur. While many of these are likely to be out of the evaluators' control, articulating these risks and instituting a plan for their management is an important precautionary step.

Sound contingency planning should allow for flexibility—some changes in operations without derailing the overall evaluation, particularly in large, multifaceted programs. A simple rule of contingency planning is to build in a sufficient cushion of budget and time whenever feasible. If the evaluation plan is not sufficiently flexible, problems that arise could lead to the end of the evaluation. With sufficient flexibility, such problems can be dealt with efficiently.

Preparing and reviewing a formalized evaluation plan

A written document that specifies the evaluation design and protocols is a helpful planning, communications, and commitment tool. Moreover, it is often required by external stakeholders to gain acceptance and permissions, set mutual expectations, and almost certainly as a prerequisite to secure funding. Evaluation plans should—ideally—be prepared several months in advance and be frequently reviewed. Regardless of whether ethical review is formally required, the plan should be reviewed with respect to ethical challenges relevant to financial capability program evaluation.¹ Apart from ethical review requirements, external review by a selected group of individuals is desirable and should include program or organizational leadership, program staff, advisory board members, and participants and community members.

¹ Ethical considerations are covered in chapter 13 of the Toolkit and are not included in this summary.

13 ETHICAL CONSIDERATIONS

The importance of protecting participants

When quantitative and qualitative research is done with human subjects, a number of ethical issues may arise with regard to designing the evaluation and collecting and analyzing data. These issues stem primarily from a concern that participants who take part in certain types of research may be negatively affected by it—either in terms of their health or social and economic well-being. For these reasons, ethical considerations should be fundamental to doing evaluations.

Evaluations of financial capability programs have some special ethical concerns, in particular in low- and middle-income countries. First, where education levels and exposure to formal finance are both low, many people actually may not be able to make informed judgments about their own finances. Second, financial capability programs may introduce participants to new risks, which can affect their well-being and may have long-term effects on financial decision making and other aspects of household behavior.

Evaluators should seek to mitigate the effects of such ethical concerns by building safeguards into their research and evaluation design, including strict procedures for recruitment, informed consent, and confidentiality procedures, among others.

Informed consent: a basic safeguard

Potential evaluation participants—whether in the financial capability program itself or the control group—should understand what the program is and be provided with an opportunity to give their consent to participate, based on fully available and easily accessible information (table 9).

When seeking informed consent, evaluators should provide the necessary information to participants in a format that is complete yet also understandable and meaningful to them and within time frames that suit both the participants and the study. Considerations include being aware of the following:

- Even basic financial terms may not be understood and should be defined in the consent process.
- Obtaining informed consent may be difficult or have adverse consequences.
- Lengthy informed consent procedures that are too burdensome may deter participants.

TABLE 9 WHAT PARTICIPANTS NEED TO KNOW AT A MINIMUM

Ensuring confidentiality through data safeguarding

Ensuring confidentiality is key to developing trust and to getting the most useful information from participants. Thus, having a data safeguarding plan that specifies storage, use, and destruction of the data upon completion of the evaluation is critical (table 10).

TABLE 10 KEY ELEMENTS OF A DATA-SAFEGUARDING PLAN

AREA	ELEMENTS
Data sensitivity	<ul style="list-style-type: none"> ▪ Are data de-identified (i.e., no personal identifiers that allow users to determine the individual's identity)?
Responsibility for data safeguarding	<ul style="list-style-type: none"> ▪ Who has overall responsibility for data safeguarding? ▪ Who else will have access to the data? ▪ Will all who have access to the data be trained in appropriate safeguarding procedures?
Data safeguarding procedures	<ul style="list-style-type: none"> ▪ Who is responsible for recruiting/enrolling participants? ▪ Are unique identifiers assigned? ▪ What (if any) personal details will be recorded? ▪ Where are data recorded? ▪ Will there be copies of the data (hard copies, soft copies, web-based storage, etc.)?
Data transmittal	<ul style="list-style-type: none"> ▪ How will data be transmitted, e.g., from program sites to evaluation "headquarters"?
Data disposal	<ul style="list-style-type: none"> ▪ How and when will data be disposed of when the evaluation is complete?

Risk assessment and mitigation planning

Risk assessments should consider all the ways an evaluation may harm participants and others (the researchers themselves, subjects' families, etc.) and, if the evaluation involves a new and untested intervention, the potential risks from the intervention

itself. Important issues to consider should include financial and economic harm as well as other types of harm:

- Could the research **compromise the physical safety or lead to psychological stress** for researchers or subjects?
- Can participation in a program evaluation **negatively affect the subjects' continued participation** in the program?
- Can participation in a program evaluation **have social repercussions** for subjects?
- Can participation in the evaluation **negatively affect the participants' well-being**?

Clearly communicating any risks that participation may entail is an important part of obtaining true informed consent.

Ethics in evaluations with randomized interventions

Randomization raises the key concern that it is “unethical” or “unfair” to randomly assign certain people to a purportedly helpful program while excluding other equally needy candidates—or, conversely, to assign the latter group to something unknown and possibly harmful. While this is intuitively understandable, randomization can sometimes be more “ethical” than it may first appear.

Still, it is critical to put in place mechanisms and safeguards to prevent conflict (e.g., within a community) that may arise from randomization. These mechanisms and safeguards must be appropriate for both the community and the evaluation’s requirements. Public announcements and commitment to a schedule of implementation, with the support of local community leaders, is an important step. Such an announcement can be good for the evaluation design because it can help prevent contamination by helping to ensure that control groups are not exposed to the treatment condition or aspects of it.

Other ethical obligations

Broader ethical obligations—including objectivity, transparency, integrity, fairness, and professional competence—are also important and can ensure that evaluation findings are reliable, trustworthy, and empowering to stakeholders, donors, and the broader public. Setting out basic **rules of engagement** at the outset of an evaluation, or working with an external evaluator, can prevent problems down the line.

14 DOCUMENTING AND COMMUNICATING RESULTS

The importance of documentation

Effectively conducting M&E is not the end of the process. Equally important is making sure that the program effort is properly documented and that the results are communicated to those who need or want to see them. Regardless of how compelling those results are, they will not be useful if no one sees them. Moreover, even if they are seen, they will not be as useful as they could be if they are not appropriately communicated to the specific audiences that need them.

Documentation is critical to ensuring that interested audiences—including program participants and managers, organization members, funders, and organizations that operate similar programs—can learn from what was found in an evaluation and apply the results. Such learning is important because the base of evidence on what makes for an effective financial capability program is currently very weak. Despite the large numbers of financial capability programs and interventions, few have been evaluated at all and, where they have, the evaluation designs have tended to limit what others can learn from them.

Obviously, failure to document and communicate results impedes the ability of others to learn from an evaluation. But such a failure can also contribute to the problem of “publication bias”—the tendency of researchers and evaluators to only publish those results that show large intervention effects. The best way to ensure that key audiences learn from an evaluation is to document and publish the results, no matter what they say.

Thinking about the relevant audiences

The process of documentation is driven by a need to think about the key audiences that one wants to inform and how to communicate with them effectively. One way to identify audiences is to determine who will be interested in the results, what information they will look for, and how they will use this information; this assessment depends on whether the audiences are internal or external (table 11).

Communicating with key audiences

The most common way of reaching all these audiences is a formal report that includes a two- to four-page summary, which can potentially stand alone as a separate document. Such reports are useful for accountability purposes and because both upper management in the evaluator’s own organization and program sponsors will expect it. Depending on which audiences one is most interested in reaching,

TABLE 11 KEY AUDIENCES AND THEIR INTERESTS

INTERNAL	Program staff and participants	<ul style="list-style-type: none"> ▪ Interested in a more detailed discussion of the program ▪ Seeing how well the program is performing and what program effects may be ▪ Seeing that their input was captured accurately ▪ Seeing any identified problems that they may need to fix ▪ Seeing that the discussion of any problems was handled constructively and with tact and sensitivity ▪ Understanding lessons learned and deciding whether and how to modify the program to incorporate those lessons
	Upper management	<ul style="list-style-type: none"> ▪ Interested in what they are being asked to do—action items identified and how these emerge from evaluation results ▪ Interested in the “short story” about program outcomes and key takeaway messages ▪ Less interested in process issues surrounding program or methodological details about how the evaluation reached its results
EXTERNAL	Funders	<ul style="list-style-type: none"> ▪ Interested in whether the program is performing effectively—whether it should be maintained, scaled-up, modified, or terminated ▪ Interested in high-level summary that goes light on methodology, sets the context for the evaluation, emphasizes results, and points to recommendations
	Policy audiences	<ul style="list-style-type: none"> ▪ Interested in how results are relevant to their interests and broader implications of results for similar programs that address similar issues ▪ Less interested in program-specific actions or recommendations that emerge from the evaluation
	People involved with similar programs	<ul style="list-style-type: none"> ▪ Interested in the details of the program and the purpose of evaluation, key questions asked, approach to data gathering and analysis, and implications of results for modifying or expanding the program
	Researchers and other evaluators	<ul style="list-style-type: none"> ▪ Interested in what the evaluation adds to the literature or evidence base ▪ Interested in the methodology and validity of the evaluation’s results, strength of the evidence that inheres in the results, and how the results add to or modify what is known about the type of program under evaluation

reports can be supplemented by other forms and venues of communication that can enhance the effectiveness of the communication effort.

Given this expectation, it is important to think about how to use this expected—and often mandatory—deliverable to effectively address the audience or audiences the evaluation results are to reach. Typically, such reports have special sections that do specific things (table 12).

TABLE 12 GENERIC TEMPLATE FOR A FORMAL REPORT

SECTION	CONTENTS
Summary	<ul style="list-style-type: none"> ▪ Brief discussion of context, approach, key findings, and recommendations for action
Introduction	<ul style="list-style-type: none"> ▪ Brief overview of the program ▪ Purpose of the program ▪ Brief overview of the approach ▪ Evaluation stakeholders ▪ Evaluators and their relationship to stakeholders ▪ Overview of the contents of the report
Program Description	<ul style="list-style-type: none"> ▪ Program history, background, and development ▪ Program goals and objectives ▪ Program participants and activities
Evaluation Design and Methods	<ul style="list-style-type: none"> ▪ Evaluation questions ▪ Data collection methods used to address each question ▪ Analysis methods for each type of data collected
Findings and Results	<ul style="list-style-type: none"> ▪ Description of how the findings are organized (e.g., by evaluation questions, themes/issues, etc.) ▪ Results of analyses of quantitative or qualitative data collected (usually represented in tables, graphs, or other visual illustrations, and text)
Conclusions and Recommendations	<ul style="list-style-type: none"> ▪ Conclusions drawn about the evaluation results ▪ Recommendations for action based on these conclusions ▪ Suggestions for further study, if applicable

Dissemination: getting the word out

With a well-crafted report and any other products tailored to key audiences in hand, there is one more step: getting the findings in the hands of those audiences. Without this final step, the evaluation report alone will not have the desired effect. Having that effect will require different modes of engagement—referred to as “channels”—for in-person communications, print distribution, electronic dissemination, and other media (such as television, radio, and blogs). Such channels vary and can be combined to create a dissemination plan (table 13).

TABLE 13 SAMPLE DISSEMINATION PLAN

TYPE OF AUDIENCE	PRODUCTS AND ACTIVITIES	DISSEMINATION CHANNELS
Program participants	<ul style="list-style-type: none"> ▪ Informal briefing 	<ul style="list-style-type: none"> ▪ Meetings or discussion groups ▪ Regular program communication channels
Program Staff/ Organization Upper Management	<ul style="list-style-type: none"> ▪ Report ▪ Informal briefing 	<ul style="list-style-type: none"> ▪ Website (intranet) ▪ Meetings
Client	<ul style="list-style-type: none"> ▪ Report ▪ Formal briefing ▪ Multimedia presentation 	<ul style="list-style-type: none"> ▪ Website ▪ Meetings
Practitioners (operators of or participants in similar programs)	<ul style="list-style-type: none"> ▪ Report ▪ Journal article ▪ Electronic newsletters ▪ Briefings and podcasts ▪ Press release ▪ Webinars and webcasts 	<ul style="list-style-type: none"> ▪ Website (external) ▪ Annual conferences ▪ Email distribution list ▪ Smaller professional conferences, seminars, and workshops
Federal Policy makers	<ul style="list-style-type: none"> ▪ Report ▪ Short policy briefs ▪ Electronic newsletters ▪ Briefings and podcasts ▪ Webinars and webcasts 	<ul style="list-style-type: none"> ▪ Website (external) ▪ Email distribution list ▪ Individual meetings, briefings, and testimonies before Congress and agency officials
General Public	<ul style="list-style-type: none"> ▪ Report ▪ Short policy briefs ▪ Press release ▪ Electronic newsletters 	<ul style="list-style-type: none"> ▪ Website (external) ▪ Traditional media campaign ▪ Social media campaign ▪ Innovative channels developed in consultation with clients
Researchers	<ul style="list-style-type: none"> ▪ Report ▪ Working paper and journal article 	<ul style="list-style-type: none"> ▪ Website (external) ▪ Presentations at academic conferences

The Russia Financial Literacy and Education Trust Fund was established in 2008 at the World Bank with funding provided by the Ministry of Finance of the Russian Federation. The work supported by the Trust Fund is jointly managed by the World Bank and the Organisation for Economic Co-operation and Development (OECD) and is directed toward improving public policies and programs to enhance financial knowledge and capabilities in low- and middle-income countries. This effort has focused on the review of national strategies for financial education, the development of methods for the measurement of financial knowledge and capabilities, methods for evaluating the impact and outcome of programs, and research applying these methods to programs in developing countries. The products of this program of work can be found at the Trust Fund website at:

www.finlitedu.org



THE WORLD BANK



MINISTRY OF FINANCE OF
THE RUSSIAN FEDERATION



OECD