USAID
FROM THE AMERICAN PEOPLE

IMPACT EVALUATION: CRITICAL CHALLENGES / PROMISING SOLUTIONS

August 2019

This publication was produced for review by the United States Agency for International Development. It was prepared for the E3 Analytics and Evaluation Project by Molly Hageboeck, Jacob Patterson-Stein, and Irene Velez of Management Systems International, A Tetra Tech Company.

IMPACT EVALUATION: CRITICAL CHALLENGES / PROMISING SOLUTIONS

August 14, 2019 (Original version submitted November 2017)

Contracted under AID-OAA-M-13-00017

E3 Analytics and Evaluation Project

Cover Photo: Woman selling carrots in a farmer's market taken by #landmatters. *Credit: Fintrac.*

Annex A Cover Photo: A discussion about data driven decision-making for HIV/AIDS. *Credit: USAID.*

ABSTRACT

This paper examines critical challenges USAID staff may face when undertaking an impact evaluation to measure change in a development outcome that is attributable to a defined USAID intervention. After USAID's 2011 Evaluation Policy, impact evaluation garnered increased attention within the Agency. However, there has been little formal review of the challenges USAID has faced and the tools Agency staff developed and used to successfully implement impact evaluations. The report draws on the experiences of USAID staff in the Bureau for Economic Growth, Education, and Environment (E3) as well as staff in other bureaus and overseas missions who have managed impact evaluations. The report examines critical technical and managerial challenges that can emerge and presents promising solutions for addressing these challenges. Approaches and tools to which readers are directed include existing Agency guidance as well as innovative solutions USAID staff have developed and applied to impact evaluations in which they were involved. Overall, this paper is positive about the prospects for impact evaluation and the integration of impact evaluation findings into evidence-based decision-making in USAID. It also recognizes that the path to success in this regard is not always smooth.

PREFACE

This report is designed to help USAID staff who commission or manage impact evaluations understand the challenges they may face as they design and implement those evaluations. The report recognizes that these two audiences differ somewhat. Accordingly, the main volume of this report provides both those who commission impact evaluations and those who manage them with information that can help them identify and address challenges to ensure the evaluation is successful and useful. Annex A: Resources for USAID Impact Evaluation Managers provides additional practical suggestions for managers. While not as immediately relevant for those who commission impact evaluations, the resources in Annex A contain examples of how USAID staff have overcome challenges to successfully implement impact evaluations.

Structurally, the main section of this volume discusses two groups of challenges to impact evaluations: technical and managerial. The report discusses both challenges and possible solutions under each group, following the order shown in the table below.

Technical Challenges	Management Challenges
• Framing flaws • Design application issues	• Inadequate stakeholder engagement • Partner collaboration issues • Mechanism and management transfers • Evaluation process management

To address challenges faced by USAID staff who manage impact evaluations, this report presents issues, approaches, and tools for those managers, arranged chronologically across the planning, implementation, reporting, and utilization phases of an impact evaluation.

Annex A is structured around a checklist for impact evaluation managers that builds on an existing USAID checklist for managing any type of evaluation, adding to it elements that are specific to impact evaluations. For each phase of an impact evaluation, Annex A provides examples of approaches and tools for addressing technical and management challenges discussed in this report. Page references within the text tie Annex A to the main volume of the report, to facilitate movement between the two volumes.

CONTENTS

TABLES

FIGURES

TEXT BOXES

ACRONYMS

3ie	International Initiative for Impact Evaluation
ADS	Automated Directives System
AOR	Contracting Officer's Representative
CDCS	Country Development Cooperation Strategy
COR	Agreement Officer's Representative
DCHA	Bureau for Democracy, Conflict, and Humanitarian Assistance (USAID)
DDL	Development Data Library
DEC	Development Experience Clearinghouse
DIV	Development Innovation Ventures
DRG	Center for Excellence in Democracy, Human Rights, and Governance (USAID/DCHA)
E3	Bureau for Economic Growth, Education, and Environment (USAID)
IDB	InterAmerican Development Bank
LER	Office of Learning, Evaluation, and Research (USAID/PPL)
MAST	Mobile Application to Secure Tenure
M&E	Monitoring and Evaluation
MOU	Memorandum of Understanding
PLC	Office of Planning, Learning, and Coordination (USAID/E3)
PPL	Bureau for Policy, Planning, and Learning (USAID)
PPR	Performance Plan and Report
RFP	Request for Proposals
SOW	Statement of Work
USAID	United States Agency for International Development

INTRODUCTION

Impact evaluations, introduced to the United States Agency for International Development (USAID) through the Agency's 2011 Evaluation Policy, can be challenging to implement. This paper is one of the first attempts to formally reflect on the Agency's experience managing impact evaluations since 2011.

This volume of the report starts with a brief review of the emergence of impact evaluations. In separate sections, the paper then identifies technical and management challenges that have come to light as USAID and other donor institutions have expanded their use of rigorous evaluations to generate credible evidence for improving development policies and practices. An included timeline can help readers understand when various challenges are likely to emerge during an impact evaluation. The technical and management sections also discuss promising solutions to the identified challenges. These solutions draw on innovative approaches that staff in USAID's Bureau for Economic Growth, Education, and Environment (E3) and other Washington-based technical and geographic bureaus developed. The paper also highlights approaches, tools, and guidance for impact evaluations that USAID's Office of Learning, Evaluation, and Research in the Bureau for Policy, Planning, and Learning (PPL/LER) developed. Annex A: Resources for USAID Impact Evaluation Managers contains a checklist for evaluation managers that highlights actions to take at each stage to enhance impact evaluation quality and prospects for utilization.

Overall, this paper is positive about the prospects for impact evaluation and the integration of impact evaluation findings into evidence-based decision-making at USAID. It also recognizes that the path to success may contain bumps in the road.

THE EMERGENCE OF IMPACT EVALUATIONS – AN OVERVIEW

In 2006, the Center for Global Development released the findings of an Evaluation Gap Working Group study that, with the support of several private sector foundations, examined reasons for the limited knowledge about the types of programs and interventions that most effectively alleviate (if not eliminate) problems that developing countries face. Specifically, the working group asked why rigorous evaluations of development assistance interventions were so rare. This *When Will We Ever Learn?* report (2006) challenged international development institutions to more definitively answer questions about what works. To do this, the report said, institutions must routinely invest in rigorous impact evaluations that can determine not only the changes in outcomes of interest achieved when development assistance interventions were delivered, but also, using a constructed counterfactual, the status of the same outcomes over the same period when interventions were not provided. When paired with emerging evaluative research, primarily from the Abdul Latif Jameel Poverty Action Lab (J-PAL) at the Massachusetts Institute of Technology, this broad challenge led to evaluation policy refinements across the donor community and spurred an exponential increase in the number of impact evaluations conducted within a few years.

Reflecting on the rapid adoption and growth of impact evaluation across the development assistance community, Ruth Levine and William Savedoff in *The Future of Aid: Building Knowledge Collectively* (2015) cautioned that continuing efforts will be needed to ensure that evaluators produce high-quality impact evaluations and that development agency managers view such evaluations as relevant and useful for their work.

> *"Despite the dynamism of the field of impact evaluation, and both the realized and the potential benefits, future progress is threatened by several forces. First and foremost is what Lant Pritchett and others have referred to as the 'hype cycle' (Pritchett 2013). In the face of real-world constraints and delays, overenthusiasm about what impact evaluation can achieve, and/or unrealistic expectations about how quickly policymakers will take up the findings from evaluations, may lead to disillusionment."*

Levin and Savedoff suggested that unrealistic expectations about the impact of rigorous evaluations on development programming could dampen enthusiasm for this type of investment, just as experience and skill with these techniques are beginning to expand.

Focusing on impact evaluation quality, *Failing in the Field: What We Can Learn When Field Research Goes Wrong* (2017) highlighted additional impediments to high-quality impact evaluations and their use, including issues in research settings, technical design flaws, challenges stemming from partner organization relationships and expectations, measurement problems, and low participation rates or other factors that limit the scale of the underlying experiments.

With expanded recent attention to impact evaluations as a tool to learn what works in development assitance, evaluators and evaluation managers at USAID and throughout the development community have come to a more mature understanding of what they need to plan and excute rigorous impact evaluations. They are also increasingly aware of the ways that various management challenges can threaten an impact evaluation's success. Reviews of completed impact evaluations echo these cautions. In the World Bank's 2012 review of impact evaluation utilization, *World Bank Impact Evaluations: Relevance and Effectiveness*, its independent evaluation group reported that the use of impact evaluations to make decisions about whether to "continue, expand, scale down, or cancel" projects was relatively low, "ranging from 22 to 33 percent, depending on the source of information." That report cited reasons for this modest level of utilization as including "poor timing, underdeveloped operational linkages, failure to engage project teams and decision-makers, or lack of dissemination." Similar findings emerged from a 2017 InterAmerican Development Bank (IDB) review, *IDB Impact Evaluation: Production, Use and Influence*. In addition, these findings are remarkably like those presented in *Evaluation Utilization at USAID* (2016), the Agency's study on its utilization of all types of evaluations. In both the IDB and USAID, impact evaluations can be difficult to find. The IDB reported that some of its impact evaluations are posted in its document collection. In USAID, where most evaluations are posted on its Development Experience Clearinghouse (DEC), one can narrow a search of its more than 140,000 evaluations using filters for sector, region, year, or language. No filter is available for evaluation type to locate potentially relevant impact evaluations and the term "impact" has been applied without regard to the definition in USAID's 2011 Evaluation Policy.

ORIGIN, METHODS, AND STRUCTURE OF THIS PAPER

This paper draws on discussions with USAID monitoring and evaluation (M&E) officers in Washington and overseas, and the authors' hands-on experience in the design, implementation, and dissemination of impact evaluations under the E3 Bureau's Analytics and Evaluation Project, as well as for Agency missions overseas through other mechanisms. The paper highlights innovative approaches developed by staff in E3 offices and other USAID bureaus, as well as useful guidance assembled by PPL/LER. To support the paper's development, the authors held discussions with M&E points of contact from various E3 offices and with PPL/LER staff. The authors also conducted telephone interviews with M&E staff from USAID's Center for Excellence in Democracy, Human Rights, and Governance in the Bureau for Democracy, Conflict, and Humanitarian Assistance (DCHA/DRG); the Bureau for Food Security; the Bureau for Latin America and the Caribbean; and the U.S. Global Development Lab. The paper describes two clusters of critical challenges – technical and management – that it further divides as shown in Table 1.

TABLE 1: CRITICAL CHALLENGES FOR IMPACT EVALUATIONS

Technical Challenges	Management Challenges
• Framing flaws • Design application issues	• Inadequate stakeholder engagement • Partner collaboration issues • Mechanism and management transfers • Evaluation process management

While conceptually distinct, these challenges often interact in practice. Some impact evaluations experience none of these problems, while others encounter more than one. These challenges can arise at various points during the planning and implementation of an impact evaluation, or after the study is completed. Figure 1 shows an illustrative impact evaluation timeline, showing phases of the evaluation in relation to phases in the activity on which the evaluation focused. Table 2 highlights when the types of challenges cited in Table 1 most often occur.

This report's discussion of specific technical and management challenges also presents promising solutions. Although technical challenges come first, this does not imply they are more important than management challenges. Rather, the order will help ensure readers understand the technical issues than can arise before examining how to use management approaches to avert or address them.

Following a discussion of impact evaluation process management at the end of this volume, Annex A provides tools and approaches for USAID evaluation managers at each phase of an impact evaluation as outlined in Figure 1. This sequentially structured annex includes approaches and tools by phase, as suggested in Table 2, as well as page references to discussions in the body of the report. The report's intent is to make promising solutions readily accessible to readers, whether they are looking for approaches to specific challenge types or for guidance on problems that may arise at specific stages of an impact evaluation.

For readers who are relatively new to impact evaluations, Box 12 in Annex A provides some readings.

FIGURE 1: ILLUSTRATIVE IMPACT EVALUATION TIMELINE

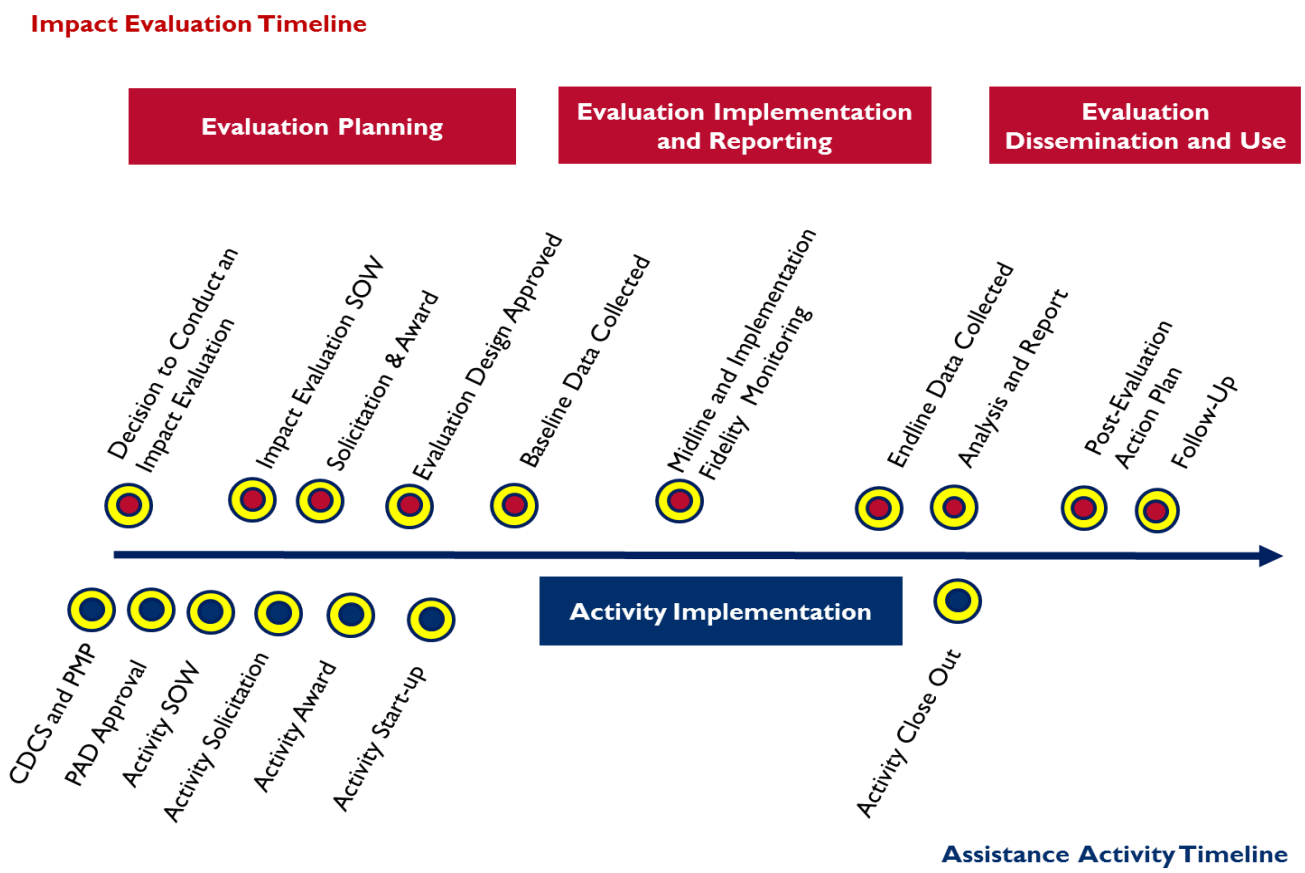


TABLE 2: TIMING OF CRITICAL CHALLENGES TO IMPACT EVALUATIONS

Impact Evaluation Timeline							
Critical Challenges to Impact Evaluation Success	**Planning**			**Implementation**		**Dissemination/ Utilization**	
	Decide Whether to Proceed with an Impact Evaluation	Prepare Scope of Work and Issue RFP	Approve Team, Create and Review Design	Apply Design and Collect Data	Analyze and Report	Share Results	Trace Post-Evaluation Action and Effects
Technical Challenges							
Framing Flaws	●	●	●				
Design Application Issues		●	●	●			●
Management Challenges							
Inadequate Stakeholder Engagement	●	●	●				●
Partner Collaboration Issues			●	●			
Mechanism and Management Transitions				●			
Evaluation Process Management		●	●	●	●	●	●

TECHNICAL CHALLENGES TO IMPACT EVALUATIONS

Technical challenges to impact evaluation affect USAID's ability to execute impact evaluations in a manner that produces credible inferences about causal linkages between USAID interventions and their intended outcomes. Technical challenges of two types are described below. The first type, *framing flaws*, includes challenges that make it difficult or impossible to select an appropriate impact evaluation design. The second type, *design application issues*, includes challenges that compromise the implementation of the impact evaluation design or the credibility of inferences based on the data produced.

FRAMING FLAWS

Framing flaws include problems that arise from the way a development assistance activity is structured and how plans for the corresponding impact evaluation are articulated. Framing flaws tend to become evident early in the impact evaluation design process. This section discusses four types of framing flaws:

1. The activity designated for an impact evaluation lacks a defined intervention.
2. The outcomes of interest are not well specified and metrics for assessing progress toward them have not been designated.
3. Questions to be examined do not align with stakeholder interests and needs for evidence to support decision-making.
4. An impact evaluation is launched before knowing whether the answers it would produce already exist.

These framing flaws are discussed briefly below, followed by promising solutions.

1. **The activity designated for an impact evaluation lacks a "defined intervention."** USAID's definition of an impact evaluation, shown in Box 1 assumes that a defined intervention exists that can be clearly specified and implemented in a consistent and stable manner for a long enough period to measure its effects on the development outcome of interest.

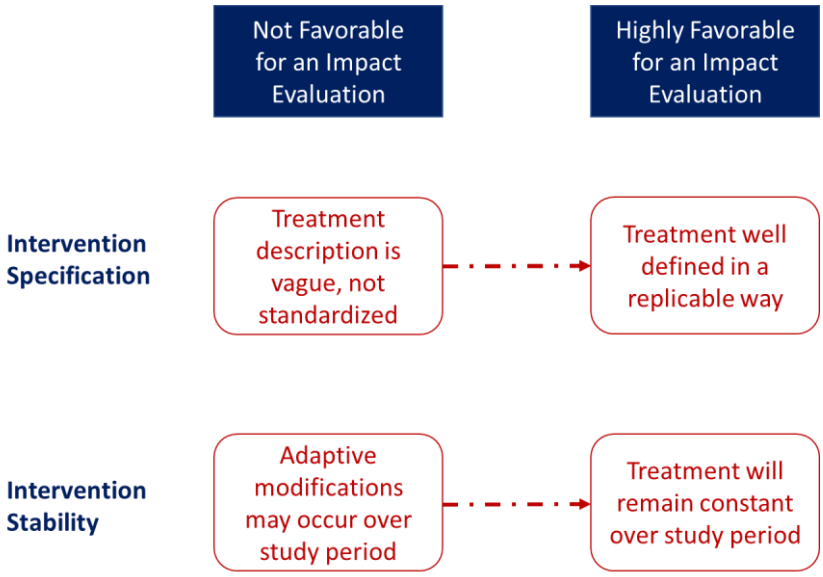
BOX 1: USAID'S IMPACT EVALUATION REQUIREMENT

"Impact evaluations measure the change in a development outcome that is attributable to a defined intervention. Each mission and Washington [operating unit] must conduct an impact evaluation, if feasible, of any new, untested approach that is anticipated to be expanded in scale or scope through U.S. Government foreign assistance or other funding sources (i.e., a pilot intervention). Pilot interventions should be identified during project or activity design, and the impact evaluation should be integrated into the design of the project or activity."

- USAID Automated Directives System (ADS) 201

Figure 2 illustrates how intervention specification and stability contribute to a judgment about whether a "defined intervention" exists, the effect of which could be determined by an impact evaluation.

FIGURE 2: FAVORABLE CHARACTERISTICS OF A "DEFINED INTERVENTION"



In practice, not all pilot or otherwise innovative interventions are well defined, and not all USAID staff and implementing partners are willing to refrain from altering an intervention for the period required to measure its effects. Some USAID activities are structured to provide grants that focus on a field or result, but even though it is not initially known what interventions the grants will deliver the Agency still wishes to rigorously evaluate the impact of the most innovative approaches. In other cases, USAID may ask an implementing partner to try several approaches to achieve a certain result and to modify its approaches as the activity evolves. While it is sometimes possible to build an impact evaluation around a somewhat fluid implementation design, it may be more appropriate in such situations to conduct operations research rather than undertake an impact evaluation.

2. **The outcomes of interest are not well specified and metrics for assessing progress toward them have not been designated.** Impact evaluations are more likely to fail if the evaluation's outcomes of interest are not well articulated, are impossible to measure, or no reliable and quantifiable proxies exist. Constructs that are not clearly defined, such as economic satisfaction or empowerment, require attention at the start of an impact evaluation to operationally define such terms and identify or create verifiable performance indicators and measurement tools that will reliably measure change over the life of the evaluation. For some evaluations, it may be necessary to operationalize stages of a process outcome to measure the results that interest USAID. For example, in an impact evaluation focused on capacity development, the evaluators may need to develop ways to measure changes in both the skills of beneficiaries and how beneficiaries apply new skills over time or to a specific problem.

3. **Questions to be examined do not align with stakeholder interests and needs for evidence to support decision-making**. Evaluation differs from research in that its purpose is to provide managers with the empirical findings and conclusions about program effectiveness that they need to transparently report on results achieved and make evidence-based policy and future programming decisions. Evaluation's purpose, including for impact evaluations, thus goes well beyond simply fulfilling an Agency evaluation requirement, testing hypotheses, or even expanding the general body of knowledge in a field. An impact evaluation that is not structured to answer questions viewed as important and relevant to Agency leadership may also lack internal champions and the buy-in needed to facilitate its start-up and sustain that commitment over an extended period. In such instances, its results may never be used. If the intended users do not view an impact evaluation as relevant, it is poorly framed.

4. **An impact evaluation is launched before knowing whether the answers it would produce already exist.** Sound impact evaluation designs generally build on a solid understanding of what previous studies (e.g., needs assessments, government time series data, other evaluations) have shown. This helps clarify the knowledge gaps that an impact evaluation might be able to fill, as well as avoid unintentionally duplicating previous examinations of the same hypotheses, possibly in a neighboring country. Skipping this step in the impact evaluation design process may also yield treatment and control/comparison group sizes that are too small to detect changes emanating from the USAID interventions, or are larger than they need to be.

To avoid or address the *framing flaws* outline above, some USAID staff and the evaluators they work with are using promising approaches and tools described below. The authors utilize this structure – first presenting problems that can emerge when conducting impact evaluations and next detailing promising solutions for addressing those problems – throughout this report to acquaint readers with approaches and tools they may find useful.

Promising Solutions

To ensure that USAID is investing wisely in specific impact evaluations:

1. Review previous USAID evaluations and the published literature to determine whether questions about causality with respect to a specific intervention or type of intervention an impact evaluation would address have already been rigorously addressed by others.
2. Consider an evaluability assessment to determine whether, beyond meeting an Agency requirement, there are valid reasons for undertaking an impact evaluation of a specific activity/intervention, and no overriding reasons for not doing so.
3. Determine whether important pre-conditions for a successful impact evaluation exist, including:
    a. Stakeholder interest in and commitment to obtaining/using the evidence an impact evaluation would provide;
    b. Whether the cause-and-effect logic for the activity/intervention is adequately articulated to support the design and implementation of a rigorous evaluation; and
    c. Whether agreed metrics for measuring change in the outcomes of interest, and data on their status, already exist.
4. Support an external peer review of the evaluation design to ensure that plans for its execution are free from serious framing flaws and other technical impediments to its success.

These analytic steps, discussed further below, go beyond determining whether an intervention in an activity, project, or country strategy fits into a category for which USAID warrants an impact evaluation under the guidelines in Box 1. Each one challenges, in a constructive way, whether USAID should proceed with a specific impact evaluation. As a blog posted by USAID's Chief Economist, which reviews the 2018 article *Ten Reasons Not to Measure Impact and What to Do Instead,* points out, there are some situations in which proceeding with an impact evaluation is not the best use of USAID resources. These include:

- When the answer has already been established through previous rigorous studies;
- When the scale of the intervention and number of units available to study is too small (i.e., underpowered) to support drawing inferences about causality;
- When:
    - The intervention to be examined is not stable;
    - Implementing partners intend to continue to modify their approach during implementation;
    - Identifying and gathering data for a valid counterfactual is unlikely to be feasible; or

- Nobody cares about the intervention or, because of its lack of relevance, expects to be able to use the evidence an impact evaluation might produce.

While some of these issues may be obvious early on, other warning signals about potential impact evaluations may not emerge until experienced USAID staff or internal evaluators undertake a scoping trip to determine which impact evaluation approaches might be feasible and sufficiently powered to merit funding. Technical feasibility is not the only arena in which warning signals are important. A lack of serious interest in what the Agency could learn from an impact evaluation, among otherwise likely evaluation users, is also a red flag; it suggests that such investments, while possibly successful from a technical perspective, may fail to produce value.

❖ REVIEW COMPLETED EVALUATIONS AND RESEARCH TO VERIFY INFORMATION GAPS

A review of prior evaluations and other research situates a proposed impact evaluation within existing knowledge in its sector or field, revealing whether previous studies have already answered questions about causality for the type of intervention and outcomes USAID plans to evaluate. A review of prior studies could result in a decision to not proceed with an impact evaluation, or to proceed in another way. Scanning the literature helps ensure that impact evaluations will potentially help USAID decision-makers arrive at sound, evidence-based decisions.

Literature reviews can also make USAID and its impact evaluators aware of factors that are pertinent for determining the scale and duration of an impact evaluation design, including the sample size of individuals, farms, or other units from which data will need to be collected to be able to detect expected differences in outcomes measures. A scan of prior studies does not necessarily take a long time or extensive effort to complete. Most impact evaluations with an international development focus are housed in a few locations (e.g., the International Initiative for Impact Evaluation [3ie] repository; the J-PAL website; the World Bank's Development Impact Evaluation collection, USAID's DEC). For education, a broader array of completed domestic and international impact evaluations is available through the U.S. Department of Education's What Works Clearinghouse.

Box 27 on page 58 provides an example of a literature review for an impact evaluation design report.

❖ CONSIDER AN EVALUABILITY ASSESSMENT OR SIMILAR ANALYSIS

Evaluability assessment has a long history in U.S. domestic agencies but is relatively new to USAID. It focuses on three important preconditions for an evaluation, including whether:

- Clear stakeholder interest in and support for an impact evaluation exists, including defined stakeholder needs for evidence to inform specific future policy and programming decisions;
- An adequate description (theory of change) exists describing, usually graphically, the causal pathways through which USAID hypothesizes an intervention, activity, or strategy will bring about changes in outcomes of interest; and
- Existing documents or other sources of evidence that would help the evaluation team complete its work, including performance monitoring data, administrative records, or studies by other donors.

As part of an evaluability assessment, a mission or Washington office that commissions an impact evaluation starts by undertaking a stakeholder analysis to determine who needs evidence about the validity of USAID's hypotheses about the effects of a specific intervention, for what purpose, and by when. Answers to these initial questions may include USAID mission and Washington staff as well as country partners.

BOX 2: INVOLVE STAKEHOLDERS AND OTHER INTENDED USERS

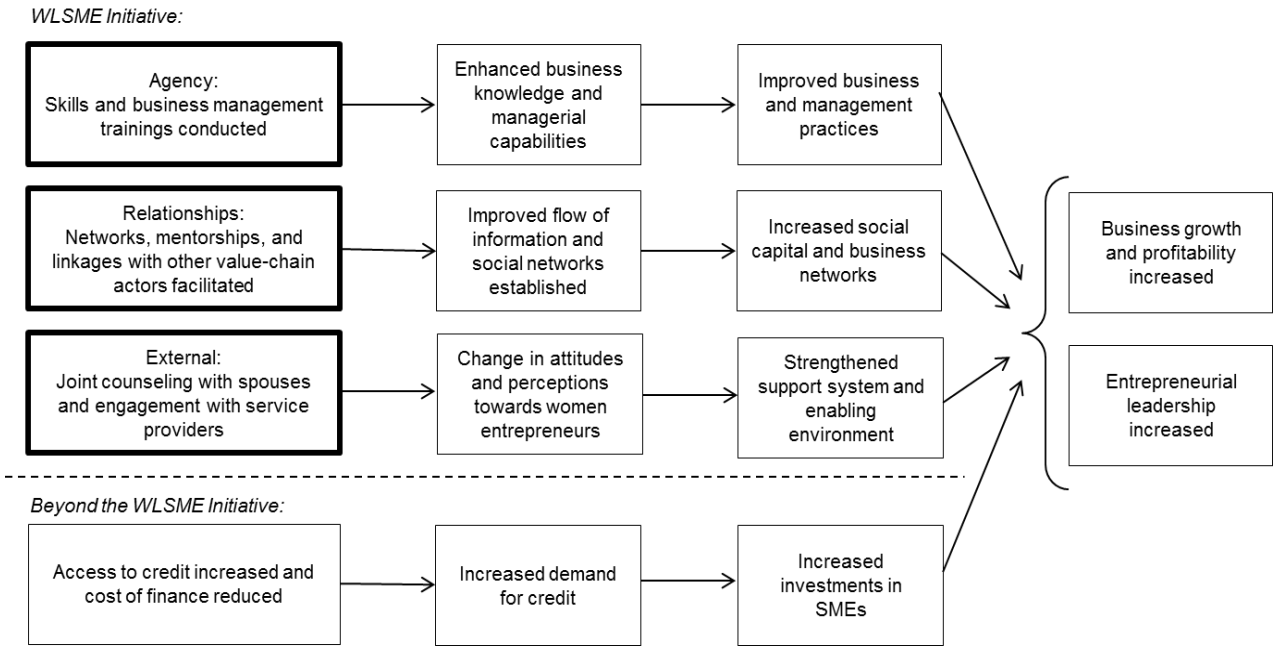
A wide variety of stakeholders often have a role in contributing to and using evaluation findings, including implementing partners, partner country governments, and other donors. Engaging stakeholders in the evaluability assessment helps to understand the possible demand for the evaluation and to generate buy-in for the evaluation. Use consultations with stakeholders to determine their needs and expectations. This includes establishing a common understanding of each stakeholder's role and ensuring a transparent process for decision-making – both for the evaluability assessment and the evaluation itself. Use the consultation process to understand what type of feedback each stakeholder group would find most helpful, and how they would like to receive this feedback.

- USAID Evaluation Resource: *Conducting an Evaluability Assessment for USAID Evaluations (2017)*

In parallel, or subsequently, USAID stakeholders who fund and manage an impact evaluation will work with other relevant actors to determine whether an adequate theory of change exists to guide the evaluation. To distinguish an evaluation-specific theory of change from a USAID results framework or logical framework that focuses on results beyond those directly affected by the intervention to be examined, USAID's E3 Analytics and Evaluation Project has often used a horizontal depiction of an evaluation theory of change. This displays both the intervention to be examined and the specific outputs and outcomes that will be measured to determine the intervention's effects. Figure 3 shows the evaluation theory of change used to examine the effects of a USAID activity that provided business training, market linkages activities, and gender-sensitive joint counseling to women entrepreneurs of small- and medium-sized enterprises in India.

Annex A provides an expanded description of the approaches and tools used in an evaluability assessment, along with links to reference materials on stakeholder analysis and theories of change. The next section, regarding management challenges, also focuses on stakeholders.

FIGURE 3: THEORY OF CHANGE EXAMPLE FOR A USAID ACTIVITY ON WOMEN ENTREPRENEURSHIP



❖ SUPPORT AN EXTERNAL PEER-REVIEW OF THE EVALUATION DESIGN

One way to increase the likelilhood that an impact evaluation is properly framed and methodologically sound is to submit the evaluation design to a panel of external peer reviewers. Under the E3 Analytics and Evaluation Project, which encourages the publication of impact evaluations in peer-reviewed journals, the E3 Bureau mandated an external peer review for impact evaluation designs. This additional peer review helps ensure that impact evaluations adhere to high professional standards and would likely be viewed favorably if submitted for publication.

To this end, the E3 Analytics and Evaluation Project has impact evaluation experts and/or sector specialists separate from the evaluation team conduct external peer reviews of draft evaluation designs. These experts are asked to identify and assess any methodological weaknesses or limitations of the draft design. The peer reviewers are given a questionnaire to guide their assessment of the impact evaluation design that includes the key impact evaluation design elements suggested by USAID's Technical Note on Impact Evaluations and reflects the research standards established by two well-regarded sources:

- Research acceptance standards of the U.S. Department of Education's What Works Clearinghouse; and
- Publication standards codified in the Publications Manual of the American Psychological Association, on which research publication standards in numerous journals in the social and behavioral sciences are based.

Page 65 in Annex A provides an illustrative protocol for carrying out an external peer review of an impact evaluation design. In addition, page 63 provides an illustrative outline of an impact evaluation design report.

DESIGN APPLICATION ISSUES

Design application issues arise when conditions that must exist for a successful impact evaluation either are not present or change substantially during the evaluation period. The following issues illustrate the range of design application problems observed in USAID-funded impact evaluations:

1. Constructing a comparison group is not practical.
2. The scale of the intervention may not be appropriate.
3. The context in which an impact evaluation would be undertaken is not suitable.
4. Timing issues emerge that may preclude the selection of some impact evaluation design options or negatively affect the implementation of the design USAID approves, including:
    - Site selection starts before the impact evaluation is structured.
    - Implementation starts before baseline data are collected.

Any of these four impediments can cause an impact evaluation to fail. They are discussed below, followed by promising solutions that USAID staff developed and used to address them.

1. **Constructing a comparison group is not practical.** In some situations, it may initially seem possible to identify a counterfactual, but in practice, the implementation context does not allow for it. The reasons for this are diverse. For example, similar villages might be too close to each other, raising concerns about crossovers, or they may be so far apart that they raise logistical and cost issues. For other evaluations, the physical safety of implementing partner staff may preclude selecting otherwise appropriate intervention sites, or evaluators may learn that the activity never collected some types of data, making it impossible to determine what members of the population received the intervention.

2. **The scale of the intervention may not be appropriate.** Detecting differences in outcomes between treatment and control/comparison groups requires appropriate minimum group sample sizes. Discussions with USAID staff and experienced evaluators suggest that intervention scale issues often explain why studies initially envisioned as impact evaluations ended up reverting to being performance evaluations. Some interventions reach too few areas (e.g., districts, schools), or take-up is too low for the evaluation to be able to conduct analysis with enough power. This happens with activities in which USAID's intervention was delivered to only a small number of units – or an intervention reached many units, but the evaluation budget is inadequate to collect data from enough of them. Poor choices about the number of units from which data will be needed to adequately power an impact evaluation also link back to what previous studies have learned about the effect size that similar interventions produced, and whether the evaluation team is aware of such facts. The appropriateness of the scale of data collection can also err in the opposite direction when evaluators calculate the size of treatment and control/comparison groups based on outcome measures further down the theory of change that require much larger samples sizes to detect expected changes, but USAID is unable to incur such high costs for data collection. It can also be an issue if the Agency is not willing or able to replicate a large data collection effort at endline that is comparable to the scale of the baseline, thus forcing a late-stage reduction in the variables examined.

3. **The context in which an impact evaluation would be undertaken is not suitable.** Partner country circumstances can negatively affect the chances for a successful impact evaluation. For example, governments may say they are open to participating in an evaluation, but later indicate that their regulations prohibit sharing administrative data that the evaluation team needs. Political changes or natural disasters can limit the evaluation team's access to areas where it needs to collect data. Agency policy shifts can affect funding for a study or result in requests from USAID to collect additional data on a variable or hypothesis at midline or endline, when that type of information did not seem important at baseline. While impact evaluation teams seek assurances in this regard, it is not uncommon for the evaluation to fail or be reconstituted as a performance evaluation because the activity dropped an intervention or expanded its target area to include villages that the evaluation had designated as control/comparison sites. For an E3 Bureau impact evaluation in the planning stage, the only possible configuration for an experimental design could have endangered implementing partner staff in some possible treatment and control villages that the partner deemed unacceptable.

4. **Timing issues emerge that may preclude the selection of some impact evaluation design options or negatively affect the implementation of the design USAID approves.** While Agency Evaluation Policy encourages having parallel contracts for implementing development assistance activities and conducting impact evaluations of those efforts, E3 Bureau staff and the authors have observed multiple impact evaluations that were initiated after key activity tasks were already underway. The most common problems of this type involve (1) the selection of implementation sites before an evaluation team is engaged, and (2) the initiation or delivery of the intervention to be studied before the evaluation team has completed pre-intervention tasks including the construction of the control/comparison group and the collection of baseline data. Box 3 discusses these two issues.

BOX 3: NOTES FROM THE FIELD – TIMING ISSUES

Notes from the Field

- USAID canceled an impact evaluation of a water replenishment activity in Southeast Asia after the baseline was completed because the sites selected were not similar enough pre-intervention, so the resulting hydrological data would not show causal impact.
- For a proposed impact evaluation of a climate resiliency activity to improve food security in Central Asia, the activity had already started implementation while the evaluation was early in the design phase. The mismatch in timelines meant that the impact evaluation was being designed retroactively, after it was no longer possible to collect baseline data.

Promising Solutions

Several non-exclusive approaches can help ensure that design application issues are fully addressed early in the design phase of an impact evaluation.

- Hold a post-award workshop with the impact evaluation and activity implementation teams soon after both teams are in place, to assess the feasibility of conducting an impact evaluation.
- Conduct scoping visits to validate conclusions from the workshop in the environment where an impact evaluation would be conducted. Ideally, implementation and impact evaluation team members will jointly carry out these scoping visits.
- Require evaluators to fully discuss technical design options in their design report or in an inception brief produced for that purpose before a design report is drafted.
- Raise awareness about issues and actions that undermine impact evaluations.

❖ HOLD A POST-AWARD WORKSHOP SOON AFTER THE IMPACT EVALUATION AND ACTIVITY IMPLEMENTATION TEAMS ARE IN PLACE

A post-award workshop is an important first step towards building a shared understanding among USAID staff and its implementing and evaluation partners about a planned impact evaluation and how it can positively interface with the implementation of an intervention, activity, or strategy to generate evidence that supports future decision-making. USAID, like the World Bank, is increasingly encouraging its Washington offices and missions to hold such post-award impact evaluation workshops. These workshops are held before scoping and site selection visits are made, and before work on the implementing partner's monitoring, evaluation, and learning plan is completed. Experience indicates that if the contracting officer's representatives (CORs) or agreement officer's representatives (AORs) for both the implementing partner and impact evaluation teams are involved in these workshops, cooperation between the implementation and evaluation teams with respect to the design, data collection, and its timing will be enhanced.

The post-award workshop should be held as early as possible following award of both mechanisms. Several potential impact evaluations commissioned through the E3 Analytics and Evaluation Project faced critical challenges because kickoff workshops were delayed and missions or implementing partners lost interest or moved to select sites and make other activity roll-out decisions without the impact evaluation team's involvement. On the other hand, a couple of impact evaluations under this Project did undertake early kickoff workshops that increased collaboration between implementing partner and impact evaluation teams and resulted in evaluation designs that were aligned to the practicalities of the interventions and were adequate to infer causality. These experiences prompted the E3 Bureau's Office of Planning, Learning, and Coordination (PLC) and the Project team to more formally integrate a kickoff workshop into its process for all new impact evaluations. Mission staff working with the E3 Bureau on impact evaluations under the Project now receive information on a kickoff workshop soon after discussions begin. This includes a description of the Project's structured approach for these workshops and a sample workshop agenda in which several standard elements are embedded, including:

- An orientation to impact evaluation for the activity implementing partner, mission staff, and country stakeholders, if appropriate.
- A discussion of the purpose of the assistance activity; USAID's hypotheses about the intended activity outcomes; the implementing partner's strategy, plans, and schedule for delivering the activity; and questions about activity effects that the impact evaluation might be able to answer.
- The evaluation team's initial ideas about how to construct a counterfactual that could support reaching conclusions about the relationship between the activity's interventions and observed changes in outcomes that are of interest to USAID and other stakeholders. If a theory of change

diagram does not yet exist for the activity, this workshop is a good time to develop and reach agreement on one. One tool that can be helpful at this stage is an impact evaluation decision tree, as shown in Figure 10 on page 60. This type of diagram helps stakeholders understand the information sequence that impact evaluation teams use to determine whether a counterfactual is possible, and whether the number of people or areas that will receive the intervention will be large enough to support making inferences about causality.

Two key outcomes sought by the end of the post-award workshop are agreements about feasible evaluation design options and the roles and responsibilities of all evaluation stakeholders. A collective review by all parties of the implementation and evaluation timelines and next steps are also important for impact evaluation success.

Roles and responsibilities of impact evaluation teams and the implementing partner team responsible for an intervention that will be evaluated are discussed further in the management challenges section that follows. The section starting on page 53 in Annex A provides additional resources for organizing post-award workshops, including an illustrative schedule for a post-award impact evaluation workshop.

❖ MAKE SCOPING VISITS AN INTEGRAL PART OF THE EVALUATION DESIGN PROCESS

In-country scoping visits, particularly joint visits by implementing partner and impact evaluation teams, can illuminate many issues that, if not considered during design, could cause an impact evaluation to fail. The feasibility of specific designs, the realism of constructing a valid counterfactual, and the ease of data collection tasks can all be better assessed in place than from a distance. A joint scoping visit allows the implementing partner and evaluation team to explore ideas about treatment sites and populations, or how they should be selected, and learn from each other how various choices would affect the others' work. The interactive nature of these discussions illustrates why such an in-country visit is important before the implementing partner or evaluation team locks its choices into a final design and work plan.

Scoping visits have proven effective in highlighting aspects of activity and evaluation planning where ideas converged, as well as pointing out areas of disagreement, providing a basis for discussion about areas for compromise. These visits serve multiple purposes: they help teams avoid framing flaws, but also help strengthen partnership relationships if conducted with implementing partners.

Pages 61-62 in Annex A provides additional information on organizing scoping visits.

❖ REQUIRE EVALUATORS TO FULLY DISCUSS TECHNICAL DESIGN CONSIDERATIONS IN THEIR DESIGN REPORT OR IN AN OPTIONS BRIEF

A number of suggested steps for avoiding impact evaluation framing flaws and design application issues warrant attention before finalizing evaluation design decisions, including:

- The results of an evaluability assessment, if undertaken;
- The findings from the evaluation team's review of published literature and locally available information, which help address whether answers to a proposed impact evaluation's main questions about causality are already known, and what data may be available in-country to monitor changes in outcomes of interest;
- Agreement about roles and responsibilities, timelines, and other aspects of evaluation team and implementing partner cooperation reached during a kickoff workshop;
- Scoping visit results regarding the feasibility of establishing control or comparison groups;
- The scale of the intervention;

- Factors in the activity context that could make it impossible to achieve adequate power for drawing inferences from study data, or otherwise render it difficult to implement an impact evaluation; and
- The evaluation team's work plan, including a Gantt chart, for producing a final design report and implementing the evaluation.

In cases where there may be a significant budget constraint or if timing is critical, a short brief or options memo summarizing the above points, but without as much detail as a full report, might be considered. The E3 Analytics and Evaluation Project typically identifies impact evaluation design options prior to preparing a full evaluation design report. A concept paper or even a simple options memorandum can serve this purpose. Sharing options early in the design process is particularly important when options vary as to the level of effort, data collection costs, or evaluation duration. In one instance where several options were being considered, the Project team created a table that showed how many household surveys would be needed for each option. This helped USAID assess and identify its preferred option for the evaluation.

❖ RAISE AWARENESS ABOUT ISSUES AND ACTIONS THAT UNDERMINE IMPACT EVALUATIONS

Timing issues often stem from human errors that result from limited awareness of the work that is underway on an impact evaluation. Premature commitments about where to locate a USAID activity or when to start delivering services can introduce bias or preclude use of some of the most rigorous impact evaluation designs, particularly experimental designs which involve the randomized assignment of individuals, schools, or districts to receive or not receive the intervention that will be examined. For these reasons, USAID missions and Washington offices need to ensure that all Agency (and embassy) staff, including at the executive level, are aware of USAID's impact intentions and the impact that decisions about timing, location, and participation can have on an impact evaluation. To this end, mission directors could highlight in senior staff meetings when an impact evaluation is being planned and that decisions that could undermine its rigor must have front office authorization. In parallel, program officers could use the announcement that an impact evaluation is being planned to explain to new and long-term staff how hand-picking assistance targets can introduce bias and why delaying an intervention until after baseline data have been collected is of value to the Agency, even when a technical office or implementing partner would like to start implementation sooner. To this end, it is useful for those involved with impact evaluations to brief mission or Washington-based leadership early in the evaluation process on the type of designs being considered and what kinds of actions could negatively affect those choices or the evaluation's overall rigor.

On a more systemic basis, USAID mission orders can establish rules of engagement for all staff when impact evaluations are being planned and implemented. Program offices, which normally fund external evaluations, play an important role in developing and updating mission orders on monitoring and evaluation. Program office staff are also well positioned to implement guidelines that a mission considers appropriate for ensuring all staff avoid taking actions that might negatively affect evaluations. Mission order guidance could also set consultation requirements on issues such as selecting field sites prior to USAID's decision on whether to randomize site selection to avoid bias in expensive and important impact evaluations.

The preceding section focused on technical issues that arise during an impact evaluation and previewed several potential solutions developed by USAID staff and evaluation partners for addressing those challenges. The following section shifts to management challenges that can impede efforts to conduct high-quality impact evaluations that USAID and its country partners utilize to inform policy and programming decisions. Both types of challenges are important for Agency missions and offices to consider when they undertake this type of evaluation.

MANAGEMENT CHALLENGES TO IMPACT EVALUATIONS

Management challenges to impact evaluation are human and systems challenges that affect USAID's ability to proceed smoothly through the phases of an impact evaluation, even when no serious technical challenges to the work exist. The first challenge, *inadequate stakeholder engagement*, can also be one of the most pervasive, appearing more than once and negatively affecting an impact evaluation's design, implementation and use. *Partner collaboration issues*, both internally and with external partners, emerge during the impact evaluation design phase as well as during implementation. A third management challenge, involving *mechanism and management transfers*, is most likely to be disruptive in an impact evaluation's implementation phase and can result in different firms carrying out an evaluation's baseline and endline studies.

Finally, this section addresses *impact evaluation process management and quality control gaps*. USAID staff managing impact evaluations face this challenge when trying to implement best practices. Relevant guidance is scattered, and effective approaches for ensuring quality at every stage of the impact evaluation process may not be documented or easily accessed. In addition, principal investigators for impact evaluations may be working with USAID for the first time and not have much experience crafting reports that align with every relevant USAID technical and style guidance. Similarly, Agency managers who may have previously managed performance evaluations can be surprised by the additional steps and technical issues their first impact evaluation involves. For these reasons, this section's approach to *impact evaluation process management and quality control* is more of a guide to promising practices and solutions than a critique.

INADEQUATE STAKEHOLDER ENGAGEMENT

"Stakeholder engagement," when linked to an impact evaluation, refers to all aspects of the involvement of a variety of actors from the planning stage through post-evaluation dissemination and utilization of findings and recommendations when making forward-looking decisions, as depicted in Figure 1. Buy-in on the evaluation questions and agreed design as well as clear intent to utilize the evaluation results – irrespective of how long it takes to obtain them – and a commitment to fund all phases of the evaluation are central stakeholder engagement concerns. This focus aligns with Agency program cycle principles in ADS 201, particularly the commitment to apply analytic rigor to support evidence-based decision-making. Inadequate stakeholder buy-in can impede completion of an impact evaluation and limit its value.

Signals that USAID managers view an impact evaluation as unlikely to yield information they would consider important, relevant, or timely are relatively easy to spot, even when a program office is enthusiastic. Such signals include:

- Meetings become difficult to schedule or are postponed indefinitely;
- An evaluation kickoff workshop is delayed until the implementing partner has completed its work plan and M&E plan; or
- Start-up is delayed until after the implementing partner has settled into the field sites where it intends to work.

These and other indications that the value to Agency management of undertaking an impact evaluation had not been established have surfaced under the E3 Analytics and Evaluation Project. They have also emerged in interactions between this report's authors and staff from USAID missions and Washington offices under other evaluation mechanisms, and in conversations with other external evaluators.

No matter how technically strong an impact evaluation is, it will have failed if, in the end, no one in USAID or among country partners and the evaluation's broader stakeholder community considered its findings when deciding whether to continue, expand, or stop funding the intervention that the impact evaluation examined. Reviews conducted by impact evaluation funders and publications by impact evaluation specialists, such as Dean Karlan and Michael Quinn Patton, point to weak stakeholder buy-in as an important explanatory factor.

"Weak buy-in" suggests a lack of agreement with impact evaluation findings. However, the literature surrounding "weak buy-in" suggests that what undermines impact evaluation completion and use is not disagreement about results, but rather a lack of stakeholder engagement in the impact evaluation process. Data from studies cited in Box 4 show that groups identified in evaluations as "intended users" – left largely out of the impact evaluation process – often end up not even learning about the findings of impact evaluations, let alone seeing the questions that the evaluation asked as relevant to their needs. The experiences of this report's authors also suggest that inadequate stakeholder involvement early in the evaluation process at least partly explained why several E3 offices ended up converting impact evaluations they had planned to undertake to performance evaluations instead.

BOX 4: EVIDENCE OF IMPACT EVALUATION INFLUENCE ON PROGRAMMING DECISIONS

In 2012, a World Bank impact evaluation review reported that, with a few well-publicized exceptions, a direct link from impact evaluations it had completed to the kinds of programming decisions listed above could be found in around 22 to 33 percent of cases. A similar review by the IDB in 2017 also reported limited evidence impact evaluation results affecting programming and policy decisions, as well as a few notable exceptions where impact evaluation evidence had a strong and well-publicized effect. USAID's 2016 examination of evaluation utilization, which did not focus on impact evaluations, found extensive use of evaluation results on programming decisions about ongoing and follow-on projects in the same location, but little use of that evaluation evidence in other countries or on Agency strategies or policies, or by country partners and other external stakeholders.

Learning what Agency leadership, program managers in missions and Washington offices, partner countries, and other key stakeholders want or feel they need to know from evaluations requires direct interaction with those audiences, which are functionally an evaluation's "customers." In lieu of interactive stakeholder consultations, USAID M&E staff and evaluation managers often do their best to anticipate the questions that Agency leadership, partner countries, and other audiences want evaluations to address. This can be helpful but representing stakeholder interests without directly engaging stakeholder representatives falls short of what best practices guidance (including the ADS) recommends.

Promising Solutions

A few non-exclusive approaches for ensuring strong stakeholder engagement over the course of an impact evaluation are discussed below:

- Identify all key stakeholders for an impact evaluation.
- Foster long-term buy-in and use through effective initial consultations.
- Develop a stakeholder engagement plan spanning all evaluation phases.

❖ IDENTIFY ALL KEY STAKEHOLDERS FOR AN IMPACT EVALUATION

Identification of the stakeholders for an impact evaluation and initial consultations may take place as part of an evaluability assessment (described on pages 9 and 35). A comprehensive review of evaluation stakeholders and their interests is equally important when an impact evaluation is initiated without conducting an evaluability assessment. When identifying stakeholders, it is often useful to pinpoint the

Agency operating unit that represents a stakeholder perspective, since within an organization as diverse as USAID views on the value of an impact evaluation and the questions it should address may vary between field offices and headquarters, and between functions such as those that technical and program offices perform. The same is true for partner government stakeholders, which may include a technical ministry as well as a government-wide M&E unit, or a ministry's representative at the district level.

For any given impact evaluation there may be a wide variety of stakeholders. Table 3 illustrates this range. However, not every type of stakeholder listed will be relevant for all impact evaluations USAID undertakes. This list can serve as a starting point for developing a more precise list of organizations, specific units within those organizations, and individuals in those units who constitute the clients, other intended users, and broader audience for the evaluation.

Page 36 in Annex A provides approaches and tools for identifying an impact evaluation's stakeholders, and page 38 shares resources on stakeholder analysis and consultations.

TABLE 3: ILLUSTRATIVE TYPES OF IMPACT EVALUATION STAKEHOLDERS

• USAID technical office staff	• Other involved U.S. government entities
• USAID senior managers	• Bilateral or multilateral donors
• Implementing partner team	• Private donors (e.g., Gates Foundation)
• Country partner – ministry level	• Assistance beneficiaries - institutions
• Country partner – local government	• Assistance beneficiaries - individuals
• Country partner(s) – civil society	• Partner country citizens
• Country partner(s) – private sector	• U.S. Congress/American people

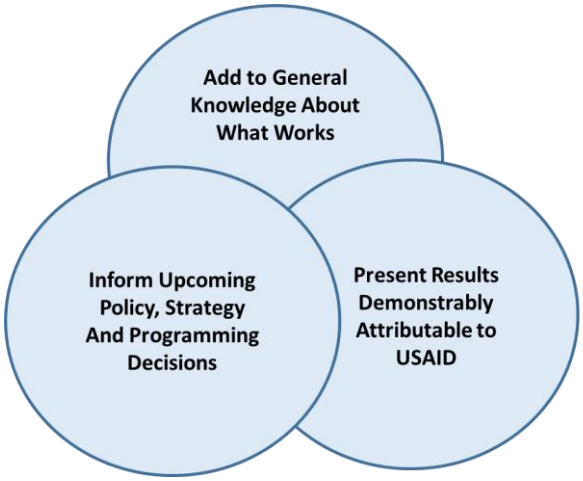
❖ FOSTER LONG-TERM BUY-IN AND USE THROUGH EFFECTIVE INITIAL CONSULTATIONS

In addition to explaining USAID's plans for an impact evaluation of a specific activity, several important discussions and consultation outputs merit attention in an initial session and require setting aside enough time to achieve them. These include:

- A clear articulation from each stakeholder group about how they expect to use the results of the evaluation. This should clarify where stakeholders want the evaluation to fall along a continuum, from knowledge-building to evidence directly relevant for specific programming or policy decisions, as illustrated in Figure 4. It should also clarify what actions stakeholders may be able to take, given differences in their authority, resources, etc.
- Clarity about each stakeholder group's interest in the evaluation process, and the points at which they most want to be engaged.
- The communication methods that stakeholders and users prefer.
- Confidence on the part of stakeholders concerning USAID's intention to continue to engage them over the course of the evaluation timeline, and on a post-evaluation basis.

Some of the strongest arguments for determining in a first consultation how key stakeholder groups anticipate using the results of an impact evaluation are captured in the 3ie Working Paper *Evaluations with Impact*. This paper argues that initial consultations with stakeholders are critical for determining where along a continuum between knowledge-focused and decision-focused a given impact evaluation falls. It also contends that decisions about an evaluation's questions, duration, and methods should follow from – rather than lead or be made independent of – an understanding of stakeholder needs for evidence and the study's decision-making context. The paper's authors posit that starting with an understanding of the decisions that stakeholders need to make will reduce a mismatch between stakeholder and evaluator perceptions and choices they frequently see.

FIGURE 4: INTERRELATED USES OF IMPACT EVALUATION FINDINGS



❖ DEVELOP A STAKEHOLDER ENGAGEMENT PLAN SPANNING ALL EVALUATION PHASES

In addition to initial consultations with stakeholder groups, it is important to keep stakeholders engaged throughout the evaluation. Periodic follow-up with stakeholders helps build their commitment to the process and using the evidence the evaluation will yield. To this end, evaluation managers may find it useful to develop a stakeholder engagement plan, building off the experience of USAID activities and partners in building stakeholder engagement to help achieve development results.

For impact evaluations, a stakeholder engagement plan outlines the ways in which an evaluation team will engage stakeholders in each phase of the evaluation process that Figure 1 outlines. Moreover, USAID staff are developing new and more effective ways of maintaining stakeholder interest over time, such as including country partner representatives in the review of an impact evaluation design report; sharing the results of the evaluation's baseline and midline studies; and engaging at least some stakeholders, including country partners, in discussions about the implications of impact evaluation findings prior to finalizing the study's recommendations.

Page 41 in Annex A provides additional information about approaches and tools for developing and implementing a stakeholder engagement plan for an impact evaluation.

PARTNER COLLABORATION ISSUES

USAID prefers that evaluations be conducted by independent teams that were not involved in the design and implementation of the examined activities and interventions. To this end, it generally relies on different partners for activity implementation and conducting impact evaluations. In most instances, the USAID COR/AOR overseeing the activity will differ from the one who oversees the evaluation. In missions, technical office staff oversee implementing partners, while a program office manages evaluations. When a Washington office supports an impact evaluation in the field, its internal partners may include a mission's technical office, program office, or both. Less frequently, one individual has served as the COR/AOR for both activity implementation and the impact evaluation.

Figure 5 illustrates how internal USAID partnerships and Agency relationships with external partners are structured for most impact evaluations, and how communications between implementing partners and evaluation teams sometimes flow through respective CORs/AORs. When the partners in these relationships share a common understanding of and commitment to an impact evaluation, the chances for a successful impact evaluation are strengthened. However, this does not always occur.

INTERNAL USAID PARTNERSHIPS

Focusing first on the internal partnership between USAID's COR/AOR for an activity and the manager for an impact evaluation of that activity helps set the stage for effective collaboration when implementing partners and an evaluation team come on board. Given their separate mandates, it is not unusual for the activity COR/AOR and the evaluation manager to view interventions in somewhat conflicting ways. From the activity perspective, delivering assistance benefits on time and within budget are natural priorities. Meanwhile, an evaluation manager structures an impact evaluation to provide the strongest possible evidence of intervention effects. In the absence of adequate communication and a common vision to accomplish both sets of objectives, issues that arise between implementing partners and evaluation teams may be exacerbated.

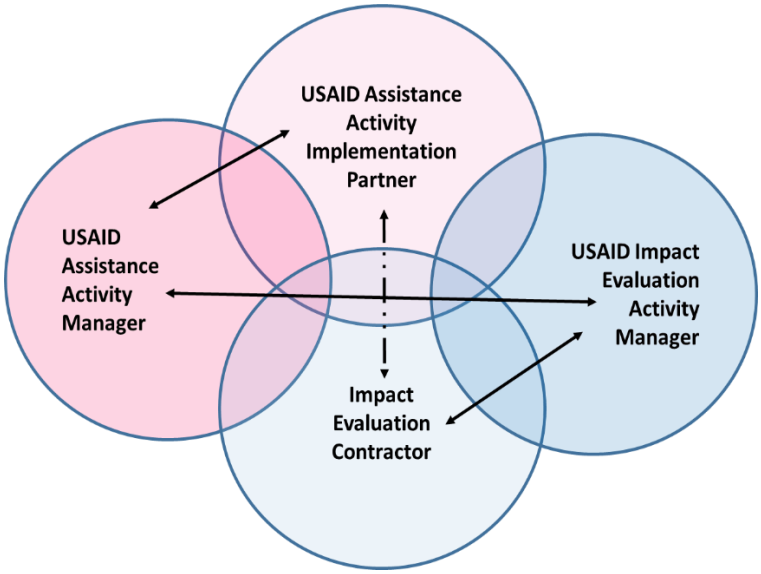
BOX 5: NOTES FROM THE FIELD – INTERNAL PARTNERSHIPS

The success of an impact evaluation comes down to the technical office in Washington and the chief of party or mission agreeing on the value of the impact evaluation.

- E3/Education Office Impact Evaluation COR

Notes from the Field

FIGURE 5: RELATIONSHIPS BETWEEN IMPACT EVALUATION PARTIES



EXTERNAL PARTNERS

All parties interviewed for this paper described the interface between implementing partners and impact evaluation teams as generally satisfactory. However, even otherwise productive relationships can experience contention between implementing partners and impact evaluation teams in situations like the following:

- Implementing partners become uncomfortable when they learn about USAID's intention to carry out an impact evaluation only after they have won the contract. A surprise of this sort can make implementing partners defensive and can render initial interactions contentious.
- Absent a startup meeting between USAID and its evaluation and implementing partners to establish working protocols for their interactions, implementing partners may see requests that come directly from the evaluation team, rather than through USAID channels, as a burden and may simply ignore them.

This paper's authors have direct experience with these kinds of issues. The first issue – when an impact evaluation comes as a surprise to the implementing partner – typically causes the greatest consternation. In many cases, USAID, its implementing partner, and the evaluation team have found ways to avoid or mitigate this challenge. In other cases, however, significant communication breakdowns or a failure to reach an agreement on how to proceed have impeded the progress of the impact evaluation.

Promising Solutions

To build effective collaboration between USAID internal partners and its external implementing partners and evaluation teams, Agency staff have introduced a number of approaches. For any impact evaluation, one or more of these promising approaches for ensuring effective collaboration may be helpful.

Strengthen USAID's Internal Partnerships

- Create opportunities for USAID stakeholders to collaboratively advance impact evaluations in which they have a shared interest.
- Elicit implementing partner input on impact evaluation plans early on.
- Work collaboratively to create a high-quality impact evaluation statements of work (SOWs)

Strengthen Partnerships with External Teams

- Use parallel solicitations to transparently inform implementation and evaluation bidders of USAID's intention to conduct an impact evaluation.
- Define the roles and responsibilities of each partner during the post-award workshop.
- Encourage external partners to undertake joint scoping visits prior to finalizing activity implementation plans and the impact evaluation design.

❖ CREATE OPPORTUNITIES FOR USAID STAKEHOLDERS TO COLLABORATIVELY ADVANCE IMPACT EVALUATIONS IN WHICH THEY HAVE A SHARED INTEREST

For most USAID evaluations it is usually possible to identify multiple internal stakeholders. In a mission these may include a technical office, program office, and possibly a gender officer or a senior staff member interested in the activity to be evaluated. Internal stakeholders may also include Washington-based technical office staff. The degree to which this group of stakeholders operates as a proactive team to facilitate an impact evaluation varies, often as a function of the effort USAID's evaluation manager exerts to foster regular communication.

Pre-evaluation clinics for impact evaluations are a valuable internal team-building activity with which DCHA/DRG, working with PPL/LER, has had considerable success. In these pre-evaluation clinics, DCHA/DRG and PPL/LER staff help mission staff prepare their upcoming activity for an impact evaluation. Using real activities at a pre-solicitation stage and providing dedicated assistance and impact evaluation training enhances the likelihood that those activities will move forward smoothly.

Page 35 in Annex A provides additional practical information on setting up this type of impact evaluation capacity building activity.

Pre-evaluation clinic participants. Credit: Jessica Benton

❖ SOLICIT IMPLEMENTING PARTNER INPUT EARLY ON

Depending on the nature of the activity USAID is planning to evaluate, the solicitation SOW prepared for implementing partner bidders may not fully specify the intervention(s) and outcomes the impact evaluation will examine. For activities that USAID expects to co-design with an awardee, or when the award will fund several grants, an opportunity may exist to encourage bidders to contribute their own ideas about the interventions on which an impact evaluation should focus. As Box 6 indicates, USAID's Evaluation Policy encourages this type of implementing partner participation in conceptualizing impact evaluations. While examples of solicitations that encourage bidders and awardees to share ideas for impact evaluations with USAID are relatively rare, the example from USAID's All Children Reading: A Grand Challenge for Development (see Box 7) illustrates how this can be done.

BOX 6: PARTNERS ARE ENCOURAGED TO SHARE IDEAS FOR IMPACT EVALUATIONS

"When opportunities exist to evaluate the impact of interventions, or to compare variations in implementation strategies, implementing partners are encouraged to bring these opportunities to the attention of the responsible technical officers. Technical officers can determine whether and how to support such impact evaluations."

- *USAID Evaluation Policy (2011)*

One way to elicit ideas from potential implementing partners about what aspects of an activity would be appropriate to examine through an impact evaluation is to ask for these suggestions when issuing a call for an expression of interest from potential bidders on an implementation contract, or to call for such ideas in a request for proposals (RFP) for potential implementing partners if one is issued. Such opportunities to solicit implementing partner ideas about what interventions warrant rigorous evaluations can have important benefits, particularly when USAID is interested in identifying approaches that may have scaling-up potential.

BOX 7: WINNING IMPACT EVALUATION IDEAS FROM ACTIVITY PARTNERS

USAID/E3's Office of Education, having invited All Children Reading grantees to submit ideas for impact evaluations of the grants they had proposed, committed to funding an impact evaluation of the best proposals that an independent evaluation team from NORC at the University of Chicago would carry out. The winning idea came from a project in Zambia that mobilizes community members to help author early grade reading materials by submitting their favorite stories and folktales by voice or digital format. The evaluation team is collecting data through an early grade reading assessment survey and learner questionnaires.

❖ WORK COLLABORATIVELY TO CREATE HIGH-QUALITY IMPACT EVALUATION SOWS

USAID's process and requirements for developing evaluation SOWs are known by many Agency technical staff. These guidelines largely focus on characteristics that are important for ensuring high-quality performance evaluations, although a few standard requirements for SOWs address impact evaluation characteristics. Other factors that need to be addressed in impact evaluations to ensure high-quality products are identified in USAID's Technical Note on Impact Evaluations (2013).

In addition to ensuring that impact evaluation SOWs address these impact evaluation-specific factors, it is also important for the members of USAID's internal impact evaluation team – often a mix of mission staff and representatives of a Washington-based technical office – share common expectations concerning the questions, design rigor, and methods to be employed. Pre-evaluation clinics, as described above and in Annex A, help ensure a meeting of the minds among internal partners. Periodic interactions among internal evaluation team members reinforce shared expectations, and it is important to collaborate on the development of an impact evaluation SOW.

To help ensure high-quality impact evaluation SOWs, it may be helpful for members of USAID's internal impact evaluation team to:

- Develop impact evaluation SOWs using a more customized impact evaluation-specific outline based on an amalgamation of USAID's current SOW requirements and good practices identified in the Technical Note on Impact Evaluations.
- Ensure that all internal impact evaluation partners are actively engaged in the review if not the development and review of an impact evaluation SOW. Neither field- nor Washington-based internal impact evaluation team members should be surprised by what is included in a final SOW for an evaluation in which they are key stakeholders. To avoid this, Agency evaluation managers should be encouraged to incorporate a highly interactive process for SOW development aimed at ensuring that agreement exists on all SOW elements, and both the field and Washington are committed to supporting any impact evaluation they co-generate or co-fund.

To support the development of high-quality impact evaluation SOWs, Box 21 on page 45 provides an expanded outline of topics and issues an impact evaluation SOW can beneficially address. This outline and other materials in Annex A reflect approaches the authors have taken with USAID impact evaluation teams to strengthen impact evaluation SOWs.

❖ CONSTRUCT PARALLEL SOLICITATIONS

USAID impact evaluation guidance calls for the inclusion of information about a planned impact evaluation in the solicitation for the activity that will be evaluated. Despite the benefits of this practice from a buy-in and cooperation standpoint, the authors of this paper observe that it is often not done. Page 51 illustrates how this can be done. This example resulted from close collaboration between USAID/West Africa and USAID/E3's Office of Forestry and Biodiversity on a regional impact evaluation. Absent this kind of language in solicitations, it is not unusual for implementing partners to be unaware of plans for an impact evaluation when they receive the award. The frequency with which implementing partners are surprised by USAID impact evaluation plans and the paucity of solicitation examples that describe those plans suggest that awareness is low regarding the Agency's requirement to include this information in solicitations.

BOX 8: PARALLEL CONTRACTS, WHEN APPROPRIATE

"If the activity is to be the subject of an evaluation, particularly an impact evaluation, information about the planned evaluation should be included in the solicitation."

- ADS 201.3.4.5

Ideally, solicitations for activity implementation and the corresponding impact evaluation would occur simultaneously, as illustrated in Figure 1. Even when solicitations are issued at different times, it is important to note in activity implementation solicitations if an impact evaluation is envisioned.

To encourage the inclusion of impact evaluation notices in activity solicitations, Box 25 on page 52 provides illustrative text for an RFP or other type of solicitation that discloses USAID's intention to undertake an impact evaluation of the activity and can be directed to share with implementation bidders.

❖ USE A POST-AWARD WORKSHOP TO REACH AGREEMENT ON ROLES AND RESPONSIBILITIES

Post-award workshops that involve implementing partner and impact evaluation teams differ from the pre-evaluation clinics indicated above, which are limited to USAID staff. Pre-evaluation clinics help build effective internal Agency teams, sometimes across bureaus, who will oversee an impact evaluation and take the lead in ensuring that post-evaluation action plans are prepared and implemented. These clinics

take place before the implementation activity is awarded. Post-award workshops, on the other hand, focus on forming a partnership between teams that implement an activity and those who evaluate it, and ensuring that these parties have a common understanding of the impact evaluation's purpose, potential design options, and timing, and are prepared to work in a coordinated manner to ensure that both activity implementation and the impact evaluation proceed smoothly and succeed.

E3/PLC, like the World Bank, is increasingly encouraging E3 offices and partner missions to hold post-award impact evaluation workshops for implementing partners and evaluation teams. These workshops take place before scoping and site selection visits or the start of work on the implementing partner's monitoring, evaluation, and learning plan. Experience indicates that if two CORs/AORs are involved – one for activity implementation and another for the impact evaluation – both should participate. In a "Notes from the Field" post on the World Bank's *Development Impact* blog, a senior impact evaluation specialist explains how these workshops orient stakeholders to impact evaluation and guide them through identifying important questions that the impact evaluation could address. The blog post's author cites reasons why such an early workshop is important, including that it: increases stakeholder ownership and reduces the chances the evaluation will "flop"; examines the political environment in which the evaluation will be conducted and has participants develop outcomes to use; and fosters awareness about what could compromise an impact evaluation or result in its failure.

Since post-award workshops are also valuable tools to prevent or address design application issues, pages 12-13 provided details on these workshops. Pages 52-57 in Annex A provide additional materials for organizing a post-award workshop, including a sample agenda.

❖ UNDERTAKE JOINT SCOPING VISITS PRIOR TO FINALIZING IMPLEMENTATION AND IMPACT EVALUATION PLANS

Another important step in the start-up phase of an impact evaluation involves finalizing decisions about where the intervention to be studied will be implemented and how a valid counterfactual will be constructed. By the time they meet, an activity implementing partner and impact evaluation team may both have ideas about treatment sites and populations, or how to select them. Site selection procedures, as well as the feasibility of other aspects of implementing partner and evaluation partner plans, are thus often a focus of their initial field trips.

Based on the authors' experiences with E3-supported studies, it appears to be beneficial for implementing partner and impact evaluation teams to jointly conduct initial scoping visits. Working in parallel in the field helps produce a shared understanding of the context in which each partner will carry out their separate-but-related tasks. Such collaborative scoping visits have effectively highlighted aspects of activity and evaluation planning where ideas converged, as well as pointed out where disagreement existed. This provides a basis for discussion about areas for compromise.

Pages 61-62 in Annex A provide additional resources for planning a scoping visit during the period allotted for preparing an impact evaluation design report.

MECHANISM AND MANAGEMENT TRANSITIONS

Five-year procurement limitations mean that some impact evaluations initiated under one mechanism will need to be completed under a different mechanism. These situations require the evaluation teams from both mechanisms and the corresponding USAID managers to work closely together to ensure a smooth transition. The authors' experiences and discussions with E3 staff and evaluators indicate that challenges to a smooth transition are most likely when:

- The design for the impact evaluation, including sample size and methods for establishing treatment and comparison groups, was not clearly or accurately documented.
- Data collection and analysis protocols (both quantitative and qualitative), including survey instruments, raw and cleaned data files, coding files, metadata, and other work products and tools, were not well organized and shared in their entirety.
- Implementation fidelity was not monitored by the original evaluation team, or the evaluation design lacked a plan and budget for this task; this results in little, if any, information on changes in the intervention during implementation reaching the impact evaluation team.
- Impact evaluations are funded only through completion of the baseline study and it is unclear, during the initial stages, how midline (if necessary) or endline data will be collected and analyzed; such evaluations run the risk of not being completed, even if mechanisms for funding later stages of the evaluation can be identified.

In addition to transfers from one mechanism to another, impact evaluations can be negatively affected when key USAID personnel relocate before an evaluation is completed and their replacements are unaware of the reason or the audience for the evaluation. This could affect the USAID staff member's level of interest or ability to ensure stakeholder buy-in over remaining stages of the evaluation.

Promising Solutions

To minimize the impact of mechanism and management transitions, advanced planning and collaborative processes have proven useful in impact evaluations conducted by E3 offices, including:

- The development and use of comprehensive transfer plans when impact evaluations must migrate from one mechanism to another; and
- A structured approach to management transitions when they occur at USAID or within evaluation teams over the period of an impact evaluation.

❖ DEVELOP A COMPREHENSIVE TRANSFER PLAN COVERING ALL PARTIES

Under the E3 Analytics and Evaluation Project, several impact evaluations that began under other mechanisms transferred to this Project to be completed, and others were transferred out for completion. Lessons learned from these experiences suggest there are benefits to developing a joint transfer plan that sets expectations about what the mechanisms sending and receiving the impact evaluation will need to do to ensure a smooth transition. A collaborative perspective and plan have advantages over a one-sided hand-off package from the mechanism that is transferring an impact evaluation for completion elsewhere.

Box 9 provides a detailed outline based on E3 Bureau experiences of the steps involved in a mechanism transition and the actions that need to be taken by USAID, the evaluation team from which an impact is to be transferred, and the follow-on evaluation team to which it will be transferred.

❖ STRUCTURED ONBOARDING WHEN INTERNAL/EXTERNAL STAFF RESPONSIBLE FOR AN IMPACT EVALUATION LEAVE OR JOIN THEIR TEAM

Activity management transitions are frequent in USAID missions, where direct-hire staff regularly rotate posts. While less frequent, USAID impact evaluation managers may change over the course of a multi-year study, and occasionally the principal investigator changes. These types of transitions can have negative consequences for ongoing impact evaluations, such as changes in expectations or schedules that affect an evaluation's completion timeline; weakened support or a diminished sense of ownership of study results; or reduced interest in promoting action based on the evaluation results.

Recognizing that poorly managed staffing transitions can result in loss of institutional memory as well as "misunderstanding of priorities, culture, roles, and responsibilities; duplication of efforts (arriving staff spending time redoing or rehashing things that have already been determined); and a disruption or discontinuation of critical relationships, both internal and external," USAID recently issued an advisory on Managing Staff Transitions Through Collaborating, Learning, and Adapting (2019). Several suggestions in this advisory seem well suited for adaptation to situations where turnover occurs within USAID or an impact evaluation team. While many of these approaches and techniques are widely known, the authors' experiences suggest that tools for facilitating management transitions that affect an impact evaluation have been piecemeal. USAID's advisory on management transitions adds value in the structured system for changing managers it recommends. Box 10 gives key suggestions for smoothing management transitions.

BOX 9: ILLUSTRATIVE ELEMENTS FOR AN IMPACT EVALUATION MECHANISM TRANSFER

USAID Actions – Identify the USAID point of contact and timeframe for working through the impact evaluation transfer with the transferring and recipient mechanisms/teams:

- Prepare a SOW for the recipient organization and share it in draft with both mechanisms/organizations involved in the transfer for comments/questions. One deliverable in the SOW should be an early post-transfer report from the recipient on its review/audit of the raw data received from the transferring organization and any issues or gaps in that data that could affect completion of the impact evaluation, including whether they can/should be addressed before further steps are taken.
- Organize transfer meetings and participate or require post-discussion written summaries of questions or issues that arise from the transferring and recipient organizations.
- Review the transferring firm's proposed list of documentation to be transferred and verify receipt of all elements with the recipient organization.

Transferring Mechanism Actions – Identify a transfer point of contact and timeframe for working through the transfer with USAID and the recipient:

- Prepare an outline of all files to be transferred to the evaluation recipient for comment and clarification
- Prepare a complete transfer file of all items covered in the transfer outline, including but not limited to: the impact evaluation solicitation from USAID; the proposal (excluding confidential financial information); the work plan; the evaluation design and notice of USAID's approval; copies of all instruments used; lists of people interviewed (excluding names of beneficiaries covered by privacy/confidentiality agreements; all interim reports (e.g., baseline; midline, implementation fidelity monitoring; quarterly progress or other reports; relevant emails where they exist); complete data set and existing analyses in formats consistent with ADS 579 inclusive of geospatial data as required for submitting to the DDL; and any other documentation the recipient organization might require to complete the impact evaluation.
- Participate in USAID-organized transfer meetings and direct follow-up communications with the recipient organization.

Recipient Mechanism Actions – Identify a transfer point of contact and timeframe for working through the transfer with USAID and the transferring mechanism/team:

- Participate in USAID-organized transfer meetings and direct follow-up communications with the transferring organization point of contact over the agreed time-frame for completing the transfer.
- Written questions, on a timely basis, concerning USAID's SOW and the transferring team's list of documents to be transferred, including requests for information not included in the transferring team's outline.
- Data review report on the recipient's review of the data package received, including any data issues identified that require attention or will affect completion of the impact evaluation (SOW deliverable).
- Recipient's work plan for completing the impact evaluation (SOW deliverable).

BOX 10: SUGGESTIONS FOR MANAGING USAID STAFF IMPACT EVALUATION TRANSITIONS

Adapted from USAID's Managing Staff Transitions through Collaborating, Learning, and Adapting

- Designate a "steward" for the management transition, particularly if a gap will occur between the departure and arrival of transitioning staff.
- Develop a "roadmap" for the transition, setting forth tasks and timelines, to guide USAID, implementing partner, and evaluation team staff.
- Specify tools and processes to capture and transfer knowledge, including a handover note, video, exit interview, or a round of meetings in the event of an overlap of individuals leaving and assuming a key impact evaluation management role. Encourage the departing individual to work closely with the "steward" to ensure that handover products are clear and complete.
- Structure learning opportunities for the new arrival. Consider organizing a "refresher" workshop, much like the kickoff workshop. If a set of roles and responsibilities were documented for USAID, the implementing partner, and the evaluation team at the kickoff, share these relationship agreements as well as how collaborations among these actors have worked to date.

EVALUATION PROCESS MANAGEMENT AND QUALITY CONTROL

While some of the impact evaluations with which the authors are familiar have been implemented smoothly, other have encountered difficulties of the kinds already discussed. In some cases, necessary steps for managing impact evaluations may have been missed or inappropriately sequenced. USAID's Technical Note on Impact Evaluations provides staff with a comprehensive overview of the purpose and methods for these studies. However, Agency staff who were relatively new to impact evaluation when assigned to manage one may also have needed a step-by-step guide through the phases of an impact evaluation, and a primer on managing relationships between Agency staff, implementing partners, and impact evaluation teams.

Promising Solutions

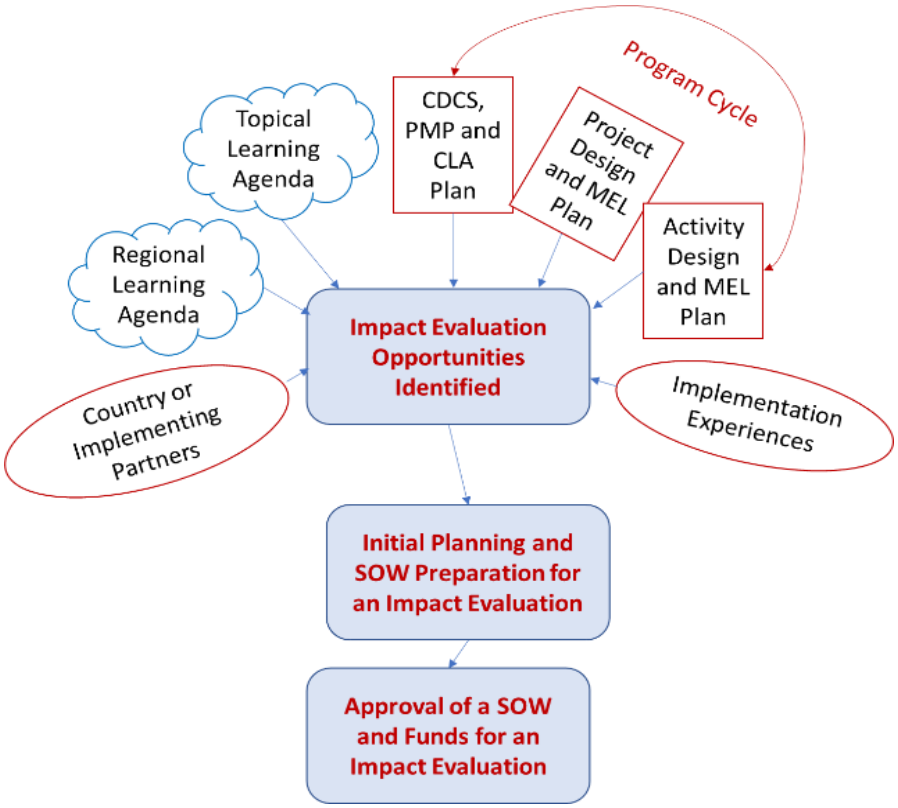
- Engage all relevant parties when deciding to undertake an impact evaluation. Organize stakeholder consultations (as described earlier) or other opportunities to systematically review the interest of various parties in what could be learned from an impact evaluation. Gauge their willingness to facilitate its implementation. Identify funding sources for conducting the evaluation and collectively agree on a "go" or "no-go" basis before expending additional time and resources.
- Identify an Agency impact evaluation manager soon after reaching a decision to proceed. Convey to this individual and his/her superiors that the manager will be responsible not only for overseeing the impact evaluation consistent with Agency expectations, but also for facilitating ongoing dialogue among evaluation stakeholders to ensure their continued buy-in and the evaluation's attention to their needs for updates and evidence from this type of investment.

❖ ENGAGE ALL RELEVANT STAKEHOLDERS IN AN INITIAL DECISION TO UNDERTAKE AN IMPACT EVALUATION

In USAID missions, the decision to undertake an impact evaluation can be influenced by many factors, as Figure 6 illustrates, and can occur as early as the start of a five-year programming cycle. This is well before activity-level hypotheses about specific interventions are usually developed. Opportunities for conducting impact evaluations can arise based information from a variety of sources, in addition to research in support of a country development cooperation strategy (CDCS). They are then captured in a mission's performance management plan or in a learning agenda prepared by a mission or a Washington-based

office. Revisions to these long-term plans may occur as project appraisal documents through which results to be achieved are prepared and SOWs for implementing activities are developed.

FIGURE 6: IDEAS FOR IMPACT EVALUATIONS CAN ARISE FROM MANY SOURCES



A decision to identify a potential impact evaluation in a performance management plan may occur long before an evaluability assessment is carried out for the activity that will implement the intervention that potential impact evaluation listing described. Ideally, USAID will review the merits of various opportunities for conducting impact evaluations with all Agency stakeholders who might need to support or otherwise be involved in an impact evaluation before proceeding.

BOX 11: NOTES FROM THE FIELD – DECIDING TO UNDERTAKE AN IMPACT EVALUATION

Notes from the Field

"In our bureau, we put together a learning agenda based on consultations and a literature review. From here, we developed six key themes and priority questions under each. From there, we did a request for concept papers. The initial response was enough to have resulted in 27 impact evaluations in 5 years. From these responses, various designs were culled, refined, and funded."

- Senior USAID program advisor

Guidance on organizing stakeholder consultations provided earlier in this volume (as well as later in Annex A), may be useful regardless of whether these conversations occur as a learning agenda or CDCS is being prepared; during project or activity design; or closer to the start of an impact evaluation, through an evaluability assessment or other means. The list of additional stakeholder references in Table 3 may also be helpful in relation to this task.

❖ IDENTIFY AND EMPOWER AN AGENCY MANAGER FOR EVERY IMPACT EVALUATION

E3 offices participate in impact evaluations in various ways. Some E3-funded impact evaluations were designed and implemented from Washington and involved data collection in one or more countries with little to no mission interaction. Others have resembled joint ventures between an E3 office and a mission, with E3 staff active in the design of an activity and its evaluation. Whenever impact evaluations involve two operating units, such as an E3 office and a mission, it is important to build a common understanding and expectations about a future impact evaluation. Identifying the USAID team that will manage the impact evaluation is an important step for this task. It involves naming an evaluation manager, normally from the operating unit that is funding the evaluation, and identifying points of contact in collaborating operating units.

Annex A is a compendium of resource materials created to help USAID impact evaluation managers carry out their responsibilities. It provides basic management tools including an impact evaluation process management checklist builds on the Agency's more generic evaluation process management checklist.

The checklist on page 31 includes additional checkpoints that are specific to impact evaluations. The first section of Annex A also provides impact evaluation managers with templates for preparing a work plan covering all impact evaluation phases. Other approaches and tools in this annex are keyed to the main phases of an impact evaluation: planning, implementation and reporting, and dissemination and utilization.

USAID
FROM THE AMERICAN PEOPLE

RESOURCES FOR USAID IMPACT EVALUATION MANGERS

Annex A to *Impact Evaluation: Critical Challenges/Promising Solutions*

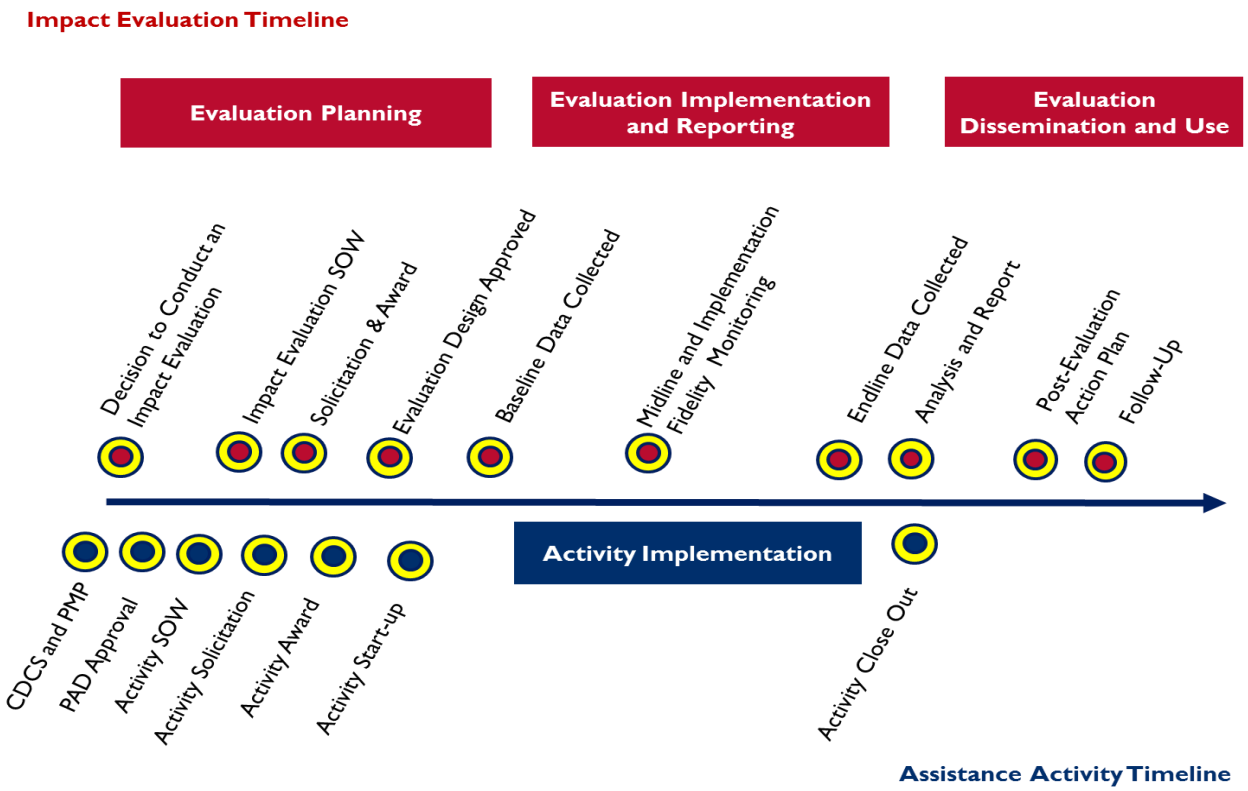
AUGUST 2019

This publication was produced for review by the United States Agency for International Development. It was prepared for the E3 Analytics and Evaluation Project by Molly Hageboeck, Jacob Patterson-Stein, and Irene Velez of Management Systems International, A Tetra Tech Company.

ANNEX A: RESOURCES FOR USAID IMPACT EVALUATION MANAGERS

This annex to *Impact Evaluation: Critical Challenges/Promising Solutions* is an extended response to the report's finding that impact evaluation mangers need consolidated resources on impact evaluation process management and quality control. The annex includes a variety of impact evaluation-specific approaches and tools organized by phases of the impact evaluation process described in Figure 7.

FIGURE 7: ILLUSTRATIVE IMPACT EVALUATION TIMELINE



IMPACT EVALUATION PROCESS MANAGEMENT CHECKLIST

USAID's Evaluation Toolkit includes an evaluation process management checklist that was designed to help Agency staff responsible for overseeing an evaluation do that job well. The checklist has no features specific to impact evaluations. For this annex, the report's authors customized USAID's existing checklist to make it a valuable resource for impact evaluation managers. Changes in USAID's existing checklist are most noticeable in the impact evaluation-specific standards provided in the accompanying text for a range of checklist tasks and milestones. In addition, the main opportunities for quality control over the course of an impact evaluation are highlighted.

IMPACT EVALUATION PROCESS MANAGEMENT CHECKLIST

(Adapted from USAID's Evaluation Process Management Checklist [2015])

Key: ☑ = Quality control opportunity

Task/Milestone	Target Date
Planning	
Preliminary decision made to explore undertaking an impact evaluation	
Evaluability assessment conducted (optional), inclusive of: • Stakeholder analysis to determine information needed, priority questions, likelihood of use • Theory of change (developed or refined) ☑ • Existing data assembled, including performance indicators and data (pre-evaluation clinics can facilitate this process)	
Decision made to proceed with impact evaluation	
Evaluation manager/COR or designee named	
Evaluation manager's impact evaluation management work plan developed	
Stakeholder input obtained, as appropriate (if evaluability assessment was not done)	
Main impact evaluation parameters determined (e.g., purpose, questions, timing)	
Mission-wide evaluation plan updated (if appropriate)	
Initial evaluation communication and dissemination plan developed	
Evaluation SOW drafted, including standards for design and final report ☑	
Evaluation independent government cost estimate drafted	
Peer review of evaluation SOW against impact evaluation-specific SOW outline (Box XX) ☑	
Final evaluation SOW approved	
Mechanism selected (for external evaluations)	
Solicitation issued (for external evaluations)	
Technical evaluation of proposals (review) ☑	
Evaluation contract awarded (external) or evaluation team selected (internal)	
Disclosure of conflict of interest forms received	
Design	
Post-award orientation ("kickoff") meeting with evaluation team ☑	
Evaluation work plan submitted by evaluation team	
Evaluation background review/literature review submitted (if requested)	
Post-award evaluation workshop with evaluation team and activity implementation team (recommended)	
Scoping visit to determine feasibility of design options (if possible); report submitted	
Evaluation design report submitted (based on design report standards/outline included in evaluation SOW (Box MM)	
Evaluation design shared with partner country stakeholders and implementing partners	
Internal USAID peer review of impact evaluation design report ☑	

Task/Milestone	Target Date
External peer review conducted by impact evaluation expert(s) ☑	
Evaluation design approved (after modifications if required)	
In-brief for mission and/or evaluation stakeholders (if requested)	
Implementation and Reporting	
Stakeholder engagement to select treatment and comparison sites	
Stakeholder engagement during baseline planning activities	
Out-briefing from baseline data collection (if requested)	
Submission of baseline draft report	
Internal USAID peer review of draft report ☑	
Submission of baseline final report	
Acceptance of baseline final report by USAID mission/operating unit	
Stakeholder engagement during implementation fidelity monitoring activities and reports ☑	
Mid-evaluation briefing/periodic briefing (if requested)	
Out-briefing (if requested)	
Stakeholder engagement during endline planning activities	
Out-briefing from endline data collection (if requested)	
Evaluation team's preliminary briefing on evaluation findings, conclusions, recommendations; feedback provided by USAID, other stakeholders ☑	
Submission of draft evaluation report	
Internal USAID peer review of draft report ☑	
External peer review conducted by impact evaluation expert(s) ☑	
Statements of differences by funders, implementers, and/or members of the evaluation team received	
Submission of final evaluation report	
Acceptance of final evaluation report by USAID mission/operating unit	
Dissemination and Utilization	
Evaluation report dissemination plan updated	
Evaluation report disseminated	
Evaluation report submitted to DEC following COR approval	
Evaluation data submitted to the Development Data Library (DDL) (if applicable)	
Evaluation contractor performance assessed (CPAR system) (if applicable)	
USAID post-evaluation review of findings, conclusions, and recommendations	
Post-evaluation action plan approved	
Actions in post-evaluation action plan completed	
Evaluation summary data entered into the evaluation registry of the performance plan and report (PPR)	

IMPACT EVALUATION MANAGEMENT PLAN

In addition to a checklist of tasks to be completed, a mangement plan for an impact evaluation might usefully include:

- A list of deliverables and standards for the acceptance of those deliverables. The discussion of impact evaluation SOWs in this annex includes a list of suggestions for the kinds of deliverables evaluation managers may want an impact evaluation team to produce. As to standards for impact evaluation deliverables, the annex provides sample outlines for impact evaluation design reports and final reports, which identify what those deliverables need to include.
- A schedule that shows when tasks will be initiated and completed, and when deliverables will be submitted. A Gantt chart can be prepared in Word or Excel that summarizes task and deliverable schedules for an evaluation manager and the team. Bidders can be asked to include a Gantt chart in in their submission and to periodically update it in their quarterly reports, or a USAID evaluation manager can create their own.

	Year 1				Year 2				Year 3				Year 4			
	Q 1	Q 2	Q 3	Q 4	Q 1	Q 2	Q 3	Q 4	Q 1	Q 2	Q 3	Q 4	Q 1	Q 2	Q 3	Q 4
Planning																
Task/Milestone																

- A summary of the roles and responsibilities of USAID staff who are members of the internal impact evaluation team. A simple way to capture this information is by extending the number of columns shown on the evaluation manager's checklist. For example:

Planning	Target Date	Evaluation Team Leader (External)	USAID Evaluation Manager	Mission M&E Officer
Task/Milestone				

BOX 12: ADDITIONAL IMPACT EVALUATION RESOURCES

USAID impact evaluation managers and others new to this type of evaluation can use a variety of methods to become more familiar with the technical aspects of this type of study, including:

- Impact Evaluation in Practice by Paul J. Gertler, Sebastian Martinez, Patrick Premand, Laura B. Rawlings and Christel M.J. Vermeersch, Inter-American Development Bank and World Bank, 2nd edition, 2016. USAID used an earlier edition of this book in the Agency's Evaluation for Evaluation Specialists course.
- Running Randomized Evaluations: A Practical Guide by Rachel Glennerster & Kudzai Takavarasha, 2013, Princeton University Press. This volume is used as the basis for some of the online courses J-PAL provides through edX.
- Handbook on Impact Evaluation: Quantitative Methods and Practices by Shahidur R. Khandkerm, Gayatri B. Koolwal and Hussain A. Samad, 2010, The World Bank.
- Experimental and Quasi-Experimental Design by William R. Shadish, Thomas D. Cook, and Donald T. Campbell, 2002, Houghton Mifflin Company. An online version is available.

Some USAID offices have also produced materials on conducting impact evaluations on specific topics or sectors that may be useful to include as handouts or pre-workshop reading material:

- USAID's MEASURE Evaluation Project's two-page summary: "What Sample Size Do I Need for an Impact Evaluation?"
- USAID's MicroLinks project has a seven-part "primer series" for impact evaluations of private-sector programs.
- Working with the multi-donor Building Evidence in Education working group, USAID/E3's Office of Education published the sector-specific Generating Evidence in Education: Impact Evaluations guide.
- For agricultural projects, USAID's Bureau for Food Security collaborated with the Millennium Challenge Corporation on the preparation of it "principles into practice" guide for Impact Evaluations of Agricultural Projects.
- The Land and Property Rights Impact Evaluation Tool developed by the E3/Office of Land and Urban is another sector-specific impact evaluation guidance coming out of the E3 Bureau.

IMPACT EVALUATION PLANNING PHASE

When planning an impact evaluation, USAID evaluation managers can take steps to avoid or minimize the negative effects of the technical and management challenges described in the main volume of this report. This is particularly true when an impact evaluation manager is assigned early and given enough authority to organize an internal team of stakeholders who help ensure these studies are technically sound and useful. This section describes illustrative approaches and tools for the impact evaluation planning phase.

Approaches and Tools

This section describes eight approaches that USAID impact evaluation managers may wish to consider for the planning phase of an impact evaluation, prior to the selection of an evaluation team:

- Hold a pre-evaluation clinic to prepare USAID staff for impact evaluations.
- Conduct an evaluability assessment, if needed.
- Create and implement a stakeholder engagement plan.

- Create a preliminary impact evaluation dissemination plan.
- Prepare an impact evaluation SOW.
- Review and approve the impact evaluation SOW.
- Select an appropriate mechanism (if the evaluation team will be external).
- Issue a solicitation for parallel contracts, if appropriate.

The main volume of this report already introduced several of these approaches. In those cases, a reference to that section is provided where appropriate in this resource annex.

❖ HOLD A PRE-EVALUATION CLINIC

Pre-impact evaluation clinics differ from the post-award workshops organized to bring together the implementing partner and impact evaluation teams that will examine an activity or strategy. USAID staff hold pre-evaluation clinics before the activity is awarded and an evaluation team is selected. These clinics, which page 20 of this report described briefly, provide technical assistance and opportunities for team building among USAID staff who expect to oversee such studies. Agency personnel organized and led sessions at these clinics. This approach for moving ideas about impact evaluations toward operational plans help ensure technical strength in impact evaluation designs as well as improve internal partnerships between offices or across bureaus to support this type of evaluation. Box 13 provides information on these clinics from USAID staff blog posts.

BOX 13: PRE-IMPACT EVALUATION CLINICS

These clinics were initiated by DCHA/DRG in collaboration with PPL/LER. The third clinic, in Washington in 2016, built on positive lessons learned during earlier clinics in Tanzania and Thailand. As Jessica Benton Cooney described the Washington-based clinic in her Learning Lab post, it brought together "mission representatives from Peru, Macedonia, Honduras, Guatemala, and El Salvador teamed up with academics, impact evaluation experts, and USAID/Washington staff to design five new impact evaluations." As explained in a post on the earlier Tanzania clinic, an invitation to participate in a clinic comes at the end of a competitive process. For Tanzania, Ms. Cooney wrote: "A total of 38 applications from field missions were submitted for participation in the clinic, and seven projects were selected based on how promising they were for designing evidence-based DRG activities that lend themselves to rigorous impact evaluation. Participating missions included: DRC, Haiti, Malawi, Mali, Nepal, Senegal, and West Bank and Gaza." While originally developed as a mechanism to foster high-quality impact evaluations of democracy and governance activities, nothing prevents other bureaus and even missions from copying this approach to strengthen impact evaluation development in their areas or soliciting PPL/LER's assistance to structure and conduct a clinic.

❖ CONDUCT AN EVALUABILITY ASSESSMENT

Page 8 of this report described evaluability assessment as an approach for ensuring that an activity, project, or strategy is ready to be the focus of an impact evaluation. This assessment has a long history in U.S. government domestic agencies, particularly the U.S. Department of Health and Human Services, and it was adapted for USAID in Planning for Cost Effective Evaluation with Evaluability Assessment (2008). It was initially applied for impact evaluations in an Evaluability Assessment of Profit Zambia, where an evaluability assessment was conducted prior to an impact evaluation baseline study. In 2015, PPL/LER issued guidance on Conducting an Evaluability Assessment. USAID's Learning Lab also offers the Overseas Development Institute guide, Evaluability Assessment for Impact Evaluations: Guidance, Checklists, and Decision Support (2015), as a resource for conducting this process.

An evaluability assessment involves a series of steps through which a design team develops answers to three key questions, shown in Box 14. The sub-sections below provide resource materials specific to each

main step of an evaluability assessment. These steps include a stakeholder analysis, the review or development of a theory of change for the evaluation, and an examination of the availability of documents relevant for operationally defining and measuring the outcomes of interest for the evaluation. Aspects of these steps can be carried out sequentially or simultaneously. It may sometimes be practical to combine the stakeholder analysis step with an effort to engage at least some stakeholders in the development of a theory of change for the activity to be evaluated.

BOX 14: EVALUABILITY ASSESSMENT FOCUS QUESTIONS

1. Would an impact assessment be useful?
   - Are there specific stakeholder needs that an impact evaluation would satisfy and can a study be designed to meet those needs?
2. Is it plausible to expect impacts?
   - Do stakeholders share a clear understanding of how the program operates and are there logical links from program activities to intended impacts? Does a "theory of change" for the activity exist?
3. Is it feasible to measure impacts?
   - Is it possible to measure the intended impacts of the activity, given the resources available for the impact assessment and the program implementation strategy? Do indicators for the activity already exist and have data on these measures previously been collected?

- Adapted from Planning for Cost-Effective Evaluation with Evaluability Assessment, 2008

Stakeholder Analysis

A stakeholder analysis, inclusive of an initial consultation with key stakeholders about the specific evaluation to be undertaken, provides an evaluation team with information about the demand for an impact evaluation, the specific questions stakeholders hope will be addressed, and the relationship of each stakeholder to the evaluation. Table 4 identifies a range of possible stakeholder groups to consider when developing a customized list for a specific impact evaluation. The organizations to be represented should be considered first, and then specific units and individuals within those units should be identified.

A stakeholder analysis of the type described here for evaluability assessments is also the first step in a broader process often referred to as stakeholder engagement. While these terms sound similar, the former is but one element of a more comprehensive process that conceptually extends across all phases of an impact evaluation. Stakeholder engagement is described on pages 41-42.

TABLE 4: ILLUSTRATIVE TYPES OF IMPACT EVALUATION STAKEHOLDERS

• USAID technical office staff	• Other involved U.S. government entities
• USAID senior managers	• Bilateral or multilateral donors
• Implementing partner team	• Private donors (e.g., Gates Foundation)
• Country partner – ministry level	• Assistance beneficiaries - institutions
• Country partner – local government	• Assistance beneficiaries - individuals
• Country partner(s) – civil society	• Partner country citizens
• Country partner(s) – private sector	• U.S. Congress/American people

There are a variety of templates in the evaluation literature for documenting stakeholders and their relationship to an evaluation. The sample in Table 4 builds on ideas included in a worksheet posted on USAID's website that is used in one of its evaluation courses. The example provided in Box 15 builds on that model and illustrates the level of detail useful for describing specific stakeholders. For any evaluation, USAID would use only the rows in this table it identified as being directly relevant and should be addressed.

TABLE 5: STAKEHOLDER ANALYSIS – IDENTIFICATION BY RELATIONSHIP TO AND ROLE IN THE IMPACT EVALUATION

Stakeholder Groups	Intended User of the Impact Evaluation Results				Facilitator of the Impact Evaluation			
	Direct Client – Evaluation Funder	Primary Intended User of Results	Secondary User Affected by Results	Observer Interested in Results	Active Support for Evaluation Required	Needs to Interact Positively with Evaluation	Aware and Needs to Avoid Interfering	Observer with No Role
USAID Program Office								
USAID Technical Office								
USAID Senior Managers								
Implementing Partner Team								
Country Partner – Ministry Staff								
Country Partner – Ministry High-Level								
Country Partner – Local Government								
Country Partner(s) – Civil Society								
Country Partner(s) – Private Sector								
Other Involved U.S. Government Entities								
Bilateral or Multilateral Donors								
Private Donors (e.g., Gates Foundation)								
Assistance Beneficiaries – Institutions								
Assistance Beneficiaries – Individuals								
Partner Country Citizens								
U.S. Congress/ American People								

BOX 15: EXAMPLE OF STAKEHOLDER ANALYSIS MATRIX FROM MEASURE EVALUATION

Name of stakeholder organization, group, or individual National, regional, or local?	**Stakeholder description** Primary purpose, affiliation, funding.	**Potential role in the issue or activity** Vested interest in the activity.
Government sector		
MOHSP National: Dr. Finar	Responsible for approving changes to the FP service delivery and service delivery guidelines.	Vested interest. Will approve the service delivery change if the evidence supports it.
MOHSP Regional: Drs. Baku, Ramannonsoa, Raharison, Andriatsiferananarivo, Rafara, Razanakoto	Responsible for implementing and assuring quality of service delivery changes in the regions.	Vested interest. Will support service delivery change with national approval.
National Commodities Storehouse: Dr. Anisoa	Responsible for ordering commodities.	Vested interest. Commodity provision.
Professional Associations		
Malagasy National Association of Doctors: Dr. Ranomenjanahary	National network to unite physicians to work on important public health issues.	Vested interest.

USAID's Collaborating, Learning, and Adapting Toolkit is another resource on stakeholder consultations, analysis, and engagement. It includes short guidance and case competitions where stakeholders played a role.

Box 16 provides additional stakeholder consultation and analysis resources. The USAID MEASURE Evaluation manual is particularly useful for understanding the specificity with which it is useful to describe stakeholder groups and identify which specific individuals represent each group.

BOX 16: ADDITIONAL STAKEHOLDER CONSULTATION AND ANALYSIS RESOURCES

Several resource documents from USAID and other sources provide useful guidance for organizing stakeholder consultations and keeping key stakeholders engaged throughout an impact evaluation:

- "Who Matters to You? Mapping Your Stakeholders," a USAID Learning Lab post
- USAID CLA Toolkit resource, Collaboration Mapping: A Facilitation Guide (2018), based on a collaboration between the LEARN project and USAID/Rwanda.
- USAID's MEASURE Evaluation Project, "Stakeholder Data Use and Dissemination Planning Tool: An Example from a Research Study in Haiti: Final Report" (2019).
- USAID Resource Guide: "Best Practices for Stakeholder Engagement in Biodiversity Programming" (2018)
- United Nations Evaluation Group, "Principles for Stakeholder Engagement" (2017)
- USAID evaluation training resource on evaluation stakeholder analysis
- Tools for Data Demand and Use in the Health Sector, a manual from USAID's MEASURE Evaluation Project
- The Centers for Disease Control booklet, Engage Stakeholders

Theory of Change

Theory of change, which the main volume of this report briefly discussed on page 9, is a broad term that subsumes a variety of methods for describing and visually displaying the cause-and-effect logic on which the success of development assistance initiatives depends.

- At the country strategy level, USAID uses a results framework diagram in a CDCS to depict cause-and-effect expectations.
- At the project level, USAID uses a logical framework to display how the achievement of an intermediate result at the Purpose level and the Outputs that support it will contribute to the realization of a Development Objective identified in a results framework at the Goal level.
- USAID sometimes uses other varieties of logic models to express cause-and-effect relationships that are building blocks of programs and other initiatives. USAID's Learning Lab also has a blog post on theory of change. Box 18 provides additional resources on theory of change.

While theories of change are frequently found in USAID documents, the focus of a mission-level results framework, for example, is usually on results at the sector- or country-level. Those theory of change diagrams do not always reach down far enough to describe the specific intervention in an activity on which a proposed evaluation would focus. Thus, the emphasis on theory of change in an evaluability assessment for an impact evaluation is on a plausible causal chain from the specific intervention of interest and the specific outputs and outcomes USAID hypothesized would result from that intervention, including those that may be out of direct influence of the activity's implementer. Often, an evaluability assessment team will find that even if a country-level results framework and a project-level logical framework exist, they do not articulate the specific intervention of interest and the causal chain toward its intended results. In these situations, USAID staff, the activity implementing partner, and the evaluability assessment team – or the evaluation team if it is already on board – will need to diagram a theory of change specific to an evaluation's scope that will be helpful to conduct the evaluation.

When working with stakeholders to develop evaluation theories of change, the E3 Analytics and Evaluation Project team often diagrams an evaluation's theory of change horizontally, rather than

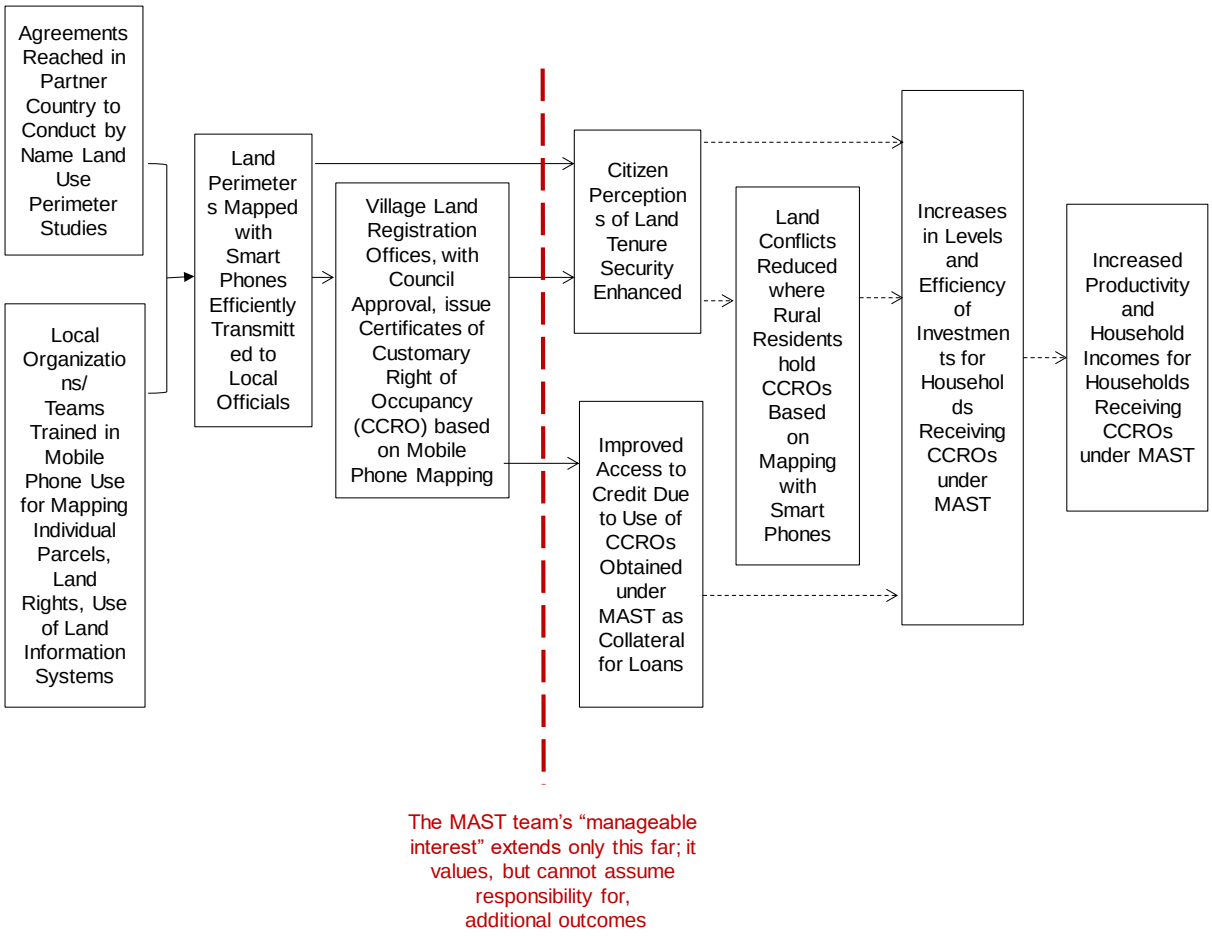
vertically, to indicate that the diagram is specific to the evaluation and does not replace an official results framework or logical framework for the activity in which the intervention of interest is embedded.

Figure 3 on page 9 of the main volume of this report provides a theory of change diagram for a USAID women's entrepreneurship activity in India through which business training, market linkages activities, and gender-sensitive joint counseling are provided to women entrepreneurs of small and medium enterprises to increase business growth, profitability, entrepreneurial leadership. Box 17 provides another theory of change diagram, along with background information, on a land tenure activity. Both diagrams are specific to the evaluations USAID undertook in that both the intervention and its intended outcomes are visible.

BOX 17: THEORY OF CHANGE – TANZANIA LAND TENURE ACTIVITY

For USAID's Mobile Application to Secure Tenure (MAST) activity in Tanzania, the E3 Analytics and Evaluation Project worked with USAID staff to articulate the cause-and-effect logic diagram representing Agency expectations about how its cellphone-based efforts to establish land boundaries and facilitate issuance of certificates of customary rights of occupancy would induce increased farmer investments in their holdings and farming practices. A horizontal diagram was used to distinguish this evaluation-specific theory of change from the broader results framework to which the MAST activity contributed.

As this evaluation theory of change was being prepared, discussions with USAID brought out questions and concerns about whether conflicts over land in the area where MAST was active would discourage farmer investment even after certificates of land occupancy were issued. This led to the insertion of a hoped-for conflict reduction result while recognizing that in some areas, this was not expected to be an issue. While conflict reduction was not an explicit intended result of MAST, discussions about its role led the impact evaluation planners to include data collection on this issue in a draft impact evaluation concept paper.
The red dotted line in this example is not a standard element of a theory of change diagram; the activity implementation team inserted it to highlight differences between the results its contract called for and the downstream outcomes USAID hoped would result from MAST and wanted to measure under its planned impact evaluation.



The resources in Box 18 provide additional guidance on preparing these graphic models of hypothesized cause-and-effect relationships between interventions and outcomes in USAID activities and strategies.

BOX 18: ADDITIONAL THEORY OF CHANGE RESOURCES

USAID's Learning Lab has several logic model and theory of change resources that may prove useful, as do other sites such as Better Evaluation. These include but are not limited to:

- "Theories of Change" USAID Learning Lab, and the Overseas Development Institute report referenced
- Theory of Change. Patricia Rogers, a methodological brief for UNICEF
- USAID's How To Note: Developing a Project Logic Model
- "What is this thing called 'Theory of Change?" USAID Learning Lab
- A Better Evaluation website page on theory of change and program logic models
- Additional USAID Learning Lab resource list on theory of change

Existing Data and Planned Data Collection on Activity Performance Indicators

The final element of an evaluability assessment, gathering relevant documents, focuses on monitoring information for a performance evaluation. It is unlikely that significant documentation will exist for a prospective impact evaluation that will work in parallel to its activity or intervention. But if it does, page 57 of this annex addresses existing information relevant for an impact evaluation in its description of the type of literature review needed as part of an evaluation design report.

❖ DEVELOP A STAKEHOLDER ENGAGEMENT PLAN

Compared to a stakeholder analysis, which is often primarily a data-gathering exercise, stakeholder engagement is a partnership-building process undertaken across multiple phases of an evaluation, or to support the design and implementation of an activity or strategy. A stakeholder analysis is a first step towards developing a stakeholder engagement plan.

Table 6 illustrates the phases and range of stakeholder involvement an impact evaluation stakeholder engagement plan might subsume. The manual from USAID's MEASURE Project includes a worked example of a similar template. USAID's Resource Guide on Best Practices for Stakeholder Engagement in Biodiversity Programming also includes several approaches that can be adapted to build stakeholder engagement in the evaluation process.

Box 16 includes resources with additional suggestions for preparing a stakeholder engagement plan.

TABLE 6: STAKEHOLDER ENGAGEMENT WORK PLAN BY IMPACT EVALUATION PHASE

Stakeholder Groups	Planning				Implementation			Dissemination/Use	
	Engage in Decision to Proceed with IE	Provide IE Status Briefings on a Schedule	Engage in Preparing or Approving IE SOW	Engage in Preparing or Approving IE Design	Brief or Share Early Products (Baseline)	Engage in Pre-Report Briefing on Results	Engage in IE Report Review and Approval	Engage through Direct Dissem-ination	Engage in Post-Evaluation Action Planning
USAID Program Office									
USAID Technical Office									
USAID Senior Managers									
Implementing Partner Team									
Country Partner – Ministry Staff									
Country Partner – Ministry High-Level									
Country Partner – Local Government									
Country Partner(s) – Civil Society									
Country Partner(s) – Private Sector									
Other Involved U.S. Government Entities									
Bilateral or Multilateral Donors									
Private Donors (e.g., Gates Foundation)									
Assistance Beneficiaries – Institutions									
Assistance Beneficiaries – Individuals									
Partner Country Citizens									
U.S. Congress/American People									

❖ CREATE A PRELIMINARY IMPACT EVALUATION DISSEMINATION PLAN

A dissemination plan for an impact evaluation specifies the stakeholders with whom USAID's impact evaluation manager and other internal team members will reach out to through a customized communication strategy, and how study findings and conclusions will be shared with them. Delivering the impact evaluation report to the DEC is an Agency requirement, but that alone does not constitute a dissemination plan.

While dissemination of an impact evaluation's results does not occur until after the study is completed, USAID recommends that preliminary versions of a study's dissemination plan occur early. In most cases it should be possible to draft an initial plan soon after an impact evaluation's questions have been selected and the evaluation's stakeholders have been identified.

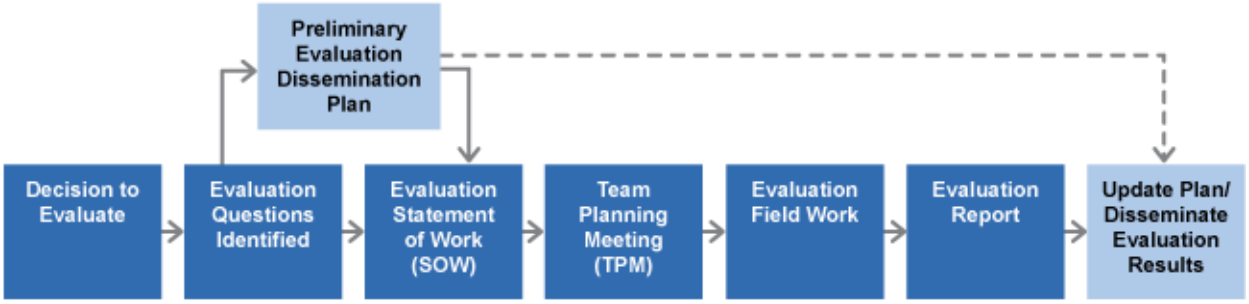
BOX 19: START EVALUATION DISSEMINATION PLANNING EARLY

"Evaluation managers should begin initial planning for dissemination at the early stages of an evaluation – after a decision to evaluate has been made but prior to completing the evaluation SOW. This way, dissemination activities can be written into the evaluation team's SOW and appropriately resourced and budgeted."

- ADS 201 Additional Help: Developing an Evaluation Dissemination Plan

Figure 8 provides a notional view of when a preliminary dissemination plan should be created in relation to the development of an impact evaluation SOW and to the preparation of a final version of a study's dissemination plan.

FIGURE 8: PRELIMINARY AND FINAL EVALUATION DISSEMINATION PLANS



The ADS guidance for Developing an Evaluation Dissemination Plan has an embedded link to a sample template for this purpose. USAID does not require that evaluation dissemination plans be posted on a preliminary or final basis, for either performance or impact evaluations. Accordingly, the percentage of evaluations for which written versions exist could not be calculated, nor were examples located by the authors of this report. The timing of a preliminary evaluation dissemination plan is nonetheless important. As Figure 8 shows, when its elements are known prior to the development of an impact evaluation SOW, related products and services can be included as deliverables in the agreement that funds the evaluation.

BOX 20: ADDITIONAL EVALUATION DISSEMINATION PLANNING RESOURCES

Other useful `guidance documents on evaluation dissemination planning include:

- CDC Coffee Break: Creating an Effective Dissemination Plan, Centers for Disease Control (2015)
- Tip Sheet: Disseminating Evaluation Results, RTI
- UNIFEM Guidance Note on Developing an Evaluation Dissemination Strategy, 2009
- Effectively Communicating Evaluation Findings, Center to Improve Project Performance (2017)

❖ PREPARE AN IMPACT EVALUATION SOW

USAID's basic guidance on the preparation of an evaluation SOW includes:

- ADS 201 Mandatory Reference on Evaluation Statement of Work Requirements (2016), which identifies 20 discrete requirements a SOW is expected to address;
- How-To-Note: Evaluation Statements of Work (2016); and
- Evaluation Statement of Work Checklist and Review Template (2017).

For examples of how various aspects of a SOW have been addressed in completed evaluations, the slightly older volume Evaluation Statements of Work: Good Practice Examples (2011) may also be helpful.

To further support the development of impact evaluation SOWs, this section provides impact evaluation-specific suggestions for topics that warrant attention in an impact evaluation outline. A number of these are topics identified in USAID's Technical Note: Impact Evaluation (2013), which are unique to this type of evaluation and are not reflected in USAID's broader evaluation guidance.

SOW elements for which these supplementary impact-evaluation suggestions are provided include:

- An outline of topics to address in an impact evaluation SOW;
- Suggestions on the evaluation's purpose, audience, and intended uses (mandatory SOW requirement 5);
- The evaluation questions to be addressed (mandatory SOW requirements 6 and 9);
- Ensuring that gender considerations receive attention (mandatory SOW requirement 7);
- Impact evaluation deliverables (mandatory SOW requirement 11); and
- Impact evaluation reporting requirements (mandatory SOW requirement 12).

SOW Coverage Outline, Including Impact Evaluation-Specific Topics

Impact evaluations differ from performance evaluations in numerous ways. Key among these are the design descriptions for impact evaluations and topics that address what will be done to ensure the rigor and reporting standards envisioned in USAID's Technical Note on Impact Evaluations. Accordingly, the first tool presented here in Box 21 is an illustrative outline of topics for a comprehensive impact evaluation SOW. Evaluation managers can further customize these topics to fit the study they will oversee.

BOX 21: SUGGESTED IMPACT EVALUATION SOW ELEMENTS

Adapted from ADS 201 Mandatory Reference on Evaluation Statement of Work Requirements (2016)

- Identify that this is a SOW for an impact evaluation
- Describe the specific strategy, project, activity, or intervention to be evaluated, including appropriate award numbers, award dates, funding levels, and implementing partners.
- Briefly describe the country or sector context; specific development problem or opportunity; and the development hypotheses, theory of change, or intervention that addresses the problem.
- Identify any existing and relevant strategy, project, or activity documents or performance information sources that are available, with special attention to monitoring data.
- State the purpose, audience(s), and anticipated uses of the evaluation, inclusive of USAID and non-USAID stakeholders (e.g., country partners, implementing partners, civil society).
- Identify a small number (i.e., no more than five) of proposed evaluation questions that are answerable with empirical evidence and relevant to future programming decisions or learning.
    - In an impact evaluation, identify questions about measuring the change in specific outcomes attributable to a specific USAID intervention (see Box 23 for examples of question wording).
    - Identify the evaluation questions requiring sex-disaggregated data, the use of gender-sensitive data collection methods, and analysis of differential impacts on males and females.
- Specify proposed data collection and analysis method(s) or request that prospective evaluators propose data collection and analysis method(s) that will generate the highest quality and most credible evidence on each evaluation question – taking time, budget, and other practical considerations into account.
    - In an impact evaluation, require specific experimental or quasi-experimental methods or request the prospective evaluators propose experimental or quasi-experimental methods.
- Specify the evaluation deliverable(s) and their timelines and logistics, including (at minimum):
    - An impact evaluation design report. Fully specify the technical criteria the design report must meet, including key questions, methods, data collection instruments, and a data analysis plan.
    - An impact evaluation final report. Fully specify the technical criteria the report must meet (e.g., the technical criteria in Box 34; all elements described in ADS 201mah, USAID Evaluation Report Requirements; reference USAID's How-To-Note: Preparing Evaluation Reports, evaluation report template, and evaluation report checklist and review protocol).
    - Datasets – and supporting documentation such as code books, data dictionaries, and the scope and methodology used to collect and analyze the data – to be submitted to the DDL, consistent with the technical requirements in guidance referenced in Box 36.
    - Additional deliverables as required to enable implementation of the Agency's anticipated dissemination plan for the impact evaluation.
- Clarify requirements for reporting and dissemination (for an impact evaluation, see above including technical criteria in Box 34).
- Clarify requirements for identifying, reporting on, and fulfilling institutional review board requirements for research in the countries and sectors the impact evaluation will examine.
- Clarify requirements for complying with USAID privacy standards when collecting and storing evaluation data.
- Include the Criteria to Ensure the Quality of the Evaluation Report to communicate USAID's quality criteria to evaluators.
- Clarify expectations about the methodological and subject matter expertise and composition of the evaluation team, including expectations concerning local evaluation team members and evaluation specialists.
- Require evaluation team members to provide a written disclosure of conflicts of interest and require key personnel to submit their disclosures with the proposal.
- Describe the intended participation of USAID staff, implementing partners, national counterparts, or beneficiaries in the design or conduct of the evaluation.
- Specify the evaluation's period of performance.
- Address scheduling, logistics, security requirements, and other support
- Include illustrative information about the level of effort expended.
- Be accompanied by an independent government cost estimate, if applicable.

SOW Purpose, Audience, and Intended Uses (Requirement 5)

USAID's basic evaluation SOW guidance on this section is clear: for completed impact evaluations there is often a lack of specificity in introductory language on these topics. USAID is often treated as an entity; the interests and intended uses for the mission's technical office, at the senior staff level, or for USAID/Washington are not broken out, nor are the specific interests and intended uses by country partners, implementing partners, or other stakeholders specifically identified in evaluation reports. It is not clear how these topics are treated in impact evaluation SOWs, as less than 10 percent of completed USAID impact evaluations on the DEC provide their SOWs as an attachment.

ADS 201 notes that a required element for every USAID evaluation SOW and final report is to "state the purpose of, audience for, and anticipated use(s) of the evaluation." Unfortunately, the cursory treatment this requirement receives, including in impact evaluations posted on the DEC, often results in a vague picture of the stakeholders who are expected to use impact evaluation findings to guide their decisions. The following paraphrased description from a 2018 USAID impact evaluation illustrates this point.

> *"Hence, the ultimate purpose of this evaluation is to provide policymakers, the government, and aid organizations the inputs required to make an informed decision on replicating or scaling up this program. More broadly, this study is also intended to contribute to the literature on [this type of intervention]."*

The evaluation SOW should precisely state who the stakeholders are (e.g., which USAID bureaus and offices commissioned the impact evaluation, and what other USAID operating units, partner country government entities, implementing partners, civil society organizations, non-governmental organizations, universities, private sector organizations, etc. are considered stakeholders or potential users of the evaluation). Existing information about the timing and nature of decisions these stakeholders expect the impact evaluation to inform also help set the stage for stakeholder consultations prior to an impact evaluation or help build a protocol that will obtain this kind of information. The performance evaluation example in Box 22, from USAID's Evaluation Statements of Work: Good Practice Examples, illustrates how this type of information can facilitate stakeholder consultations at the start of an impact evaluation.

BOX 22: EXAMPLE OF EVALUATION AUDIENCE AND INTENDED USES

Example 13: Audience and Intended Uses of an Evaluation

The audience of the evaluation report will be the USAID/Mozambique Mission, specifically the health team, the Africa Bureau, and the implementing partner. An Executive Summary and recommendations will be provided to the MOH. USAID will use the report to make changes to its current strategy of providing support to the central level and to share lessons learned with other stakeholders; Chemonics and its subcontractors will learn about their strengths and weaknesses and adjust their programs accordingly; and the MOH will learn more on how to better benefit from Chemonics TA. It is expected that the PVO partners and the Provincial Health Directorates (DPSs) will have the opportunity to discuss how the Chemonics Forte Saúde project assisted them and how this type of project could better assist them in the future to meet goals.

Absent better descriptions of the intended audience for an impact evaluation, it is almost impossible for USAID's impact evaluation manager or other internal team members to organize an appropriate set of pre-evaluation stakeholder consultations.

SOW Evaluation Questions to be Addressed (Requirements 6 and 9)

Two types of questions are important to consider when preparing an impact evaluation SOW:

1. Questions about causality, which are central to any impact evaluation; and
2. Questions that emerge in discussions with those most likely to use impact evaluation findings to inform policy and programming decisions and other stakeholders prior to the development of a SOW. These questions can identify further evidence that decision-makers may need to act on an impact evaluation's findings about the effectiveness of interventions and the attribution of results to USAID's investments.

Questions about Causality

To help impact evaluation managers standardize impact evaluation SOWs with respect to questions about causality, Box 23 describes appropriate ways to structure these questions, consistent with USAID policy and several examples included in the Technical Note on Impact Evaluations. While there are other ways in which questions about causality can be written, the guidance and examples in this box illustrate the kinds of question forms USAID expects. At minimum, the examples provided should make clear that the evaluation question section should not be written as a list of hypotheses instead of a list of questions, or as a narrative paragraph in which what is to be learned is included but not presented in the form of a clear list of questions.

BOX 23: STRUCTURING IMPACT EVALUATION QUESTIONS

Following the ADS definition of an impact evaluation, the simplest version of an impact evaluation question asks:

What was the effect of the intervention (X) on the outcome(s) of interest compared to results in its absence? For example:

- What was the effect of the activity's early grade reading intervention on children's reading scores compared to its absence? (single intervention, comparison group, single outcome example)
- What effect did the presence of drip irrigation have on corn yields and farm family earnings compared to farms without this technology? (single intervention, sequential outcomes from the same cause)
- Compared to participants in the control group, do participants who are randomly assigned to receive the program have higher mean values on the following post-intervention outcomes: entrepreneurial leadership, business growth, business knowledge/practices, and social/business networks? (single intervention; multiple outcomes)

When an impact evaluation will test more than one intervention, the question structure may change to:
How effective were each of the interventions compared to each other and to no intervention?
For example:

- What was the effect of the children's reading intervention on reading scores when the intervention included afterschool tutorial sessions; parental coaching at home; or a tablet provided to each child for their independent study use, compared to no intervention? (three versions (arms) of the intervention, comparison group, single outcome)
- What were the effects of the nutrition intervention alone, compared to the sanitation intervention alone and to a combined intervention that provided both nutrition and sanitation assistance? (two different interventions, combination of the two, comparison group, single outcome)

Additional Questions on Which Decision-Makers May Seek Evidence from Impact Evaluations

USAID's purpose in undertaking impact evaluations is to provide decision-makers in the Agency and among country partners with credible evidence on which to base policy and programming decisions. In practice, USAID, other donors, and impact evaluation experts at 3ie, J-PAL and other entities that conduct and serve as repositories for impact evaluations are finding that strong evidence about the effects of development assistance interventions, while necessary, are not always enough to trigger policy or programming changes, even where the resources to do so exist. Examinations of why the evidence that impact evaluations produce is not being reflected in decisions as extensively as the evaluation community had hoped to highlight the fact that some decision-makers feel they also need complementary evidence on other questions, including:

- The cost of shifting from current practice to a new approach or intervention. Put another way: how cost-effective is the innovation an impact evaluation examined compared to either current practice or other relevant alternatives?
- The portability of the intervention that was evaluated: what is known about the intervention and the context in which it was examined that would make it likely or unlikely that it would be equally effecting in a different setting? Simply put, what did the evaluation learn that is pertinent for understanding the external validity of its findings?
- The unintended consequences of the intervention the impact evaluation examined. What positive or negative side effects were detected?
- The sustainability of the intervention that was studied: what did the evaluation learn about the likelihood of sustainability or the factors that would affect sustained delivery or sustained benefits?

These questions rarely appear in USAID impact evaluation SOWs. However, some are receiving increased attention in organizations that produce impact evaluations – particularly the first two questions, which are further discussed below.

Cost-Effectiveness

The cost-effectiveness of an intervention studied through an impact evaluation is emerging as the most critical question for organizations and individual decision-makers who are moving toward a more evidence-based approach to their development assistance practice. J-PAL, for example, has begun adding a cost-effectiveness element to its impact evaluations. Some of the Millennium Challenge Corporation's impact evaluations also include cost-effectiveness in their scope. USAID/E3's Office of Education is laying the ground work for moving in this direction by improving its ability to determine the cost of education interventions, on a unit cost basis, to compare options and foster evidence-based decision-making. Box 24 includes an excerpt from a recent J-PAL publication explaining its decision to incorporate cost-effectiveness more systematically into its impact evaluations.

Portability (also known as External Validity)

Research into the transferability of impact evaluation findings to new locations and institutions, which lies at the heart of a scaling-up strategy, is proving more difficult than are efforts to incorporate cost-effectiveness analysis as an impact evaluation best practice. Research by Stanford visiting professor and founder of Aid Grade, Eva Vivalt, reported over several years (updated in 2019) on what we can generalize from impact evaluations, based on meta-analyses of 20 types of international development interventions. She demonstrates important differences in effect sizes found in multiple studies testing close to the same interventions as well as differences in the design and implementation of studies that may contribute to such variations. Other studies reach similar conclusions, including those reviewed in the 3ie Evaluations with

Impact volume. Examining the question of generalizability by looking at USAID Development Innovation Ventures (DIV) grants, Duflo and Kremer (2015) found that 28 percent of effective interventions tested through DIV grants had reached more than 100,000 people in less than 4 years. Low-cost interventions were more likely to be scaled up than other interventions, interventions that had been the subject of an impact evaluation were more likely to have been scaled up than others, and most scaling up occurred in countries where the innovation had been tested while little scaling up occurred in other countries – even though some of the best-known scaling-up examples (e.g., cellphones, deworming, oral rehydration salts, bed-nets) have proven scalable beyond their original sites.

BOX 24: THE ARGUMENT FOR EXAMINING COST-EFFECTIVENESS IN IMPACT EVALUATIONS (A SYNOPSIS)

"In the last 15 years, there has been a sharp increase in the number of rigorous evaluations of the impact of development programs in a host of fields. … One of the major objectives of such studies is to provide evidence to policymakers on what works and does not work in the fight against poverty, so they can use scientific evidence to determine which policies and programs to adopt and invest in. But it can be very difficult for policymakers to compare results from different programs and their evaluations, performed in different countries, in different years, and that use different instruments to achieve the same outcome. … As a result, policymakers may decide to ignore such evidence altogether and go back to relying on their instincts on what works or does not work, or selectively choose studies that support their instincts or predetermined choices. … One way to encourage policymakers to use the scientific evidence from these rigorous evaluations in their decision-making is to present evidence in the form of a cost-effectiveness analysis that compares the impacts and costs of various programs. … Cost-effectiveness analysis by itself does not provide enough information for a policymaker to make an investment decision, but such analysis does provide a very useful starting point for researchers and policymakers to collaborate in assessing the efficacy of the different programs and their relevance to the specific situation. Moreover, when cost-effectiveness analyses have been done with data at a highly disaggregated level, where assumptions about key factors such as program take-up or unit costs are made explicit, it is much easier to perform sensitivity analysis. This sort of sensitivity analysis gives policymakers an idea of how cost-effective a similar program might be in their situation by varying key assumptions to reflect their context."

- *Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education.* Iqbal Dhaliwal, Esther Duflo, Rachel Glennerster, Caitlin Tulloch. J-PAL, August 10, 2011.

The portability of impact evaluation findings remains a serious question for both impact evaluation specialists and Agency decision-makers who may wish to use impact evaluation evidence to inform programming across a broader range of locations. To this end, Evaluations with Impact includes a list on pages 16-17 of dimensions along which the generalizability of impact evaluation results may fail. This list may prove useful to USAID evaluation teams and mangers when trying to interpret the implications (and limitations) of any specific impact evaluation for decisions about replication or scaling-up.

The discussion above is not a recommendation for USAID to require that impact evaluations include questions beyond the basic causality questions they are mandated to address. Rather it shows that some decision-makers are making known their interest in and need for evidence beyond proof of superior effectiveness to change their policies and programming practices, and that some impact evaluation providers are responding to those needs, particularly with respect to adding a cost-effectiveness component where possible.

SOW Impact Evaluation Gender Considerations (Requirement 7)

Collecting evaluation data on a gender-specific basis and incorporating evaluation approaches to allow evaluators to detect and report on gender-specific and gender-differential effects of USAID-funded interventions is an established requirement, and needs to be factored into plans for impact evaluations. Principal investigators recruited from universities who are first-time USAID evaluators are often unaware of the Agency's expectations in this regard. USAID impact evaluation managers in these instances have a special obligation to raise awareness on this point and ensure that impact evaluation teams are familiar with resources such as Engendering Evaluation at USAID early in the process. Putting this topic on the agenda for the evaluation's kickoff workshop, discussing this requirement, and distributing USAID's guide at that time is ideal. Also worth sharing with impact evaluation teams is the Overseas Development Institute's Addressing Gender in Impact Evaluation.

SOW Impact Evaluation Deliverables (Requirement 11)

The minimum set of deliverables USAID prescribes for any evaluation in its evaluation SOW requirements include a design report, draft and final versions of the impact evaluation report, and the datasets developed in the course of the evaluation. Impact evaluation deliverables may include additional items. USAID encourages that SOWs include a deliverables list with all products identified in a preliminary impact evaluation dissemination plan (see page 43) that are needed to support the dissemination and utilization of an evaluation. These additional deliverables often include things like two-page briefs and slide decks but might also include social media posts or a short video. Including them in the SOW list of deliverables helps ensure that the evaluation budget will include the funds needed for their preparation.

With respect to the requirement for a design report and draft and final versions of the impact evaluation report, USAID's 2011 Evaluation Policy and its Technical Note on Impact Evaluations set a new bar for the strength of evidence concerning what changed as the result of USAID's interventions under an activity or strategy. They also raised expectations about the professionalism with which plans for and the results of such evaluations are documented. These expectations are mentioned with respect to questions addressed by impact evaluations in USAID's statement of evaluation SOW requirements, and described more fully in the Technical Note on Impact Evaluations. In this technical note, USAID explicitly states or alludes to the need for impact evaluation design reports and final reports to place the study in its broader context, indicating whether prior research has addressed the same or similar questions about causality; explain the impact evaluation design chosen, how the counterfactual was designed, what its expected minimum detectable effect size is and why that parameter was chosen; and describe what is the resulting power of the study design, including presentation of the model or other description of the study's parameters and expectations about attrition and other limiting factors – and, in the final report, the actual effect size found and its statistical and practical significance. While the technical note describes the range of topics impact evaluations need to address and is aligned with other statements of professional norms, it does not provide USAID staff and evaluators with a consolidated statement or checklist of what must be covered in a design or final report.

Based on the above, and direct experience with USAID staff on impact evaluations, this annex suggests that Agency impact evaluation managers can improve the quality of the impact evaluations they oversee by including illustrative outlines for design and final report deliverables as attachments to impact evaluation SOWs. This annex provides examples of these two attachments that can be copied into SOWs:

- An illustrative outline of the topics to be covered in an impact evaluation design report that would encourage evaluators to present their design in a manner consistent with professional norms and USAID's Technical Note on Impact Evaluations is provided in Box 30 on page 63.
- An illustrative outline of the topics to be covered in an impact evaluation's draft and final reports that is consistent with professional norms for reporting impact evaluations and USAID's Technical Note on Impact Evaluations is provided in Box 34 on page 73.

In addition, page 76 provides those who are responsible for creating impact evaluation SOWs with an explanation and list of official guidance documents on transferring evaluation data sets to the DDL. This list can also be copied into impact evaluation SOWs to help ensure evaluation teams fully understand the expectations associated with their SOW's statement of required deliverables.

❖ REVIEW AND APPROVE AN IMPACT EVALUATION SOW

USAID's requirement for internal peer reviews of evaluation SOWs was introduced in the 2011 Evaluation Policy and reaffirmed in its 2016 update of ADS 201. USAID also issued an Additional Help Program Cycle Guidance for the Evaluation SOW Peer Review Process in 2016, and the Learning Lab contains guidance on using an Evaluation SOW Checklist and Review Template to support such reviews. In addition, this section suggests USAID evaluation managers consider reviewing impact evaluation SOWs against Box 21.

Recognizing that impact evaluation duration may not match the period of performance under available procurement mechanisms, procurement options warrant careful review before an operating unit decides how it will fund an impact evaluation. In parallel to choosing a mechanism, USAID staff should ensure that impact evaluation SOWs are portable. To this end, SOW sections on evaluation management and deliverables can specify when and how the impact evaluation team should document evaluation planning and implementation protocols to facilitate a smooth transition, if necessary.

❖ SELECT AN APPROPRIATE MECHANISM (IF TEAM WILL BE EXTERNAL)

The duration of an impact evaluation is one of its distinguishing features. While most performance evaluations are completed within months, impact evaluations generally take a few years, ending after the activity being evaluated terminates. Some impact evaluations are conducted under separate, parallel contracts in which evaluation duration is matched to the period required to determine the intervention's impact. Other impact evaluations are initiated under multiple evaluation mechanisms such as Agency-wide indefinite quantity contracts or bureau- or office-level evaluation support contracts. In addition to the Agency's array of contracts and agreements, USAID impact evaluation managers may have access to mission-level monitoring, evaluation, and learning support activities through which impact evaluations can be funded, or to U.S. government-wide ordering agreements for both large and small businesses. In all cases, impact evaluation managers are well-advised to consult with the program and contracting offices that support the operating unit in which they work when considering what mechanism will best suit the type of impact evaluation they are planning, and its anticipated scope, duration, and budget parameters.

❖ ISSUE AN IMPACT EVALUATION SOLICITATION

For impact evaluations, USAID encourages the use of parallel solicitations where appropriate that alert implementing contract bidders of a planned impact evaluation and make impact evaluation bidders aware of the key aspects and timeline for implementation activities. While this is not always feasible, it can encourage early interaction between these two teams and help avoid problems that compromise the work of either team. Box 25 provides illustrative text for an implementing partner solicitation in situations where parallel solicitations are used.

BOX 25: ILLUSTRATIVE RFP LANGUAGE FOR IMPLEMENTATION BIDDERS ON AN IMPACT EVALUATION

Request for Task Order Proposals SOL-624-15-000001 for the West Africa Biodiversity and Climate Change (WA-BiCC) Activity for USAID/West Africa (USAID/WA)

"An important component of the WA-BiCC project is to identify which interventions work, and why, to facilitate learning and scaling up across the region. A core objective of the WA-BiCC approach is that regional and national decision- makers use evidence to inform policy and management decisions for biodiversity conservation, climate change mitigation, and adaptation. As a result, USAID/WA will identify opportunities and implement an impact evaluation over the course of the WA-BiCC project. It is expected that the implementer of the WA-BiCC project implementing partner will not only accommodate, but and proactively support, the implementation of an impact evaluation over the life of the project.

Specifically, USAID/WA, in consultation with the USAID/E3 Bureau, shall evaluate approaches to the conservation, restoration, and/or sustainable management of coastal and mangrove ecosystems, and gather evidence to understand the biodiversity, climate change mitigation, and climate adaptation outcomes these interventions provide. USAID/E3 will provide an analysis of some (not all) potential interventions, as well as a potential framework for impact evaluation of some of these interventions. This document is not intended to be prescriptive. However, it is the intention of USAID to rigorously evaluate the impact of the project (or components thereof) within the structure of an experimental or quasi-experimental design. The WA-BiCC contractor will be expected to accommodate modifications to its approaches, implementation schedule, and monitoring framework to facilitate such an evaluation. This may include modifications, including, but not limited to, the selection, timing, location, and way activities are implemented and the monitoring methodology used. It is expected that the contractor will collect certain data (or facilitate the collection of such data by other partners selected by USAID). The purposes served by these data include: the selection of intervention and/or control or comparison sites; to understand baseline conditions and post-intervention outcomes; to understand biophysical and socio-economic variables which that may identify the conditions in under which project approaches are (or are not) successful; or other purposes, as appropriate. USAID/WA expects to count on independent evaluation expertise to advise the project and carry out aspects of the evaluation; however, the WA-BiCC contractor will need to coordinate some aspects of project implementation and monitoring with USAID and the external evaluation contractor. Expectations and coordination will be discussed, agreed to and articulated during the WA-BiCC annual work planning process.

Given that USAID intends to conduct an impact evaluation of certain biodiversity and climate change approaches in coastal areas, the offeror shall identify (within the discussion of their technical approach) those interventions and associated higher-level biodiversity and climate change outcomes (not outputs) which that it feels are particularly suitable for impact evaluation. The technical approach will suggest an implementation and monitoring approach which that the contractor estimates will facilitate the impact evaluation of those primary outcomes, as well as any secondary or co-benefits (e.g., socio economic, health, governance, etc.) which that the contractor evaluates may result from the project. The impact evaluation will be implemented under a parallel mechanism that will be separated, arranged, and supervised by USAID.

USAID evaluation policy also requires that baseline data from both target group and comparison group members be collected prior to the start of the intervention under study. To this end, bidders should expect that upon award, USAID will facilitate discussions between its implementing partner and impact evaluation team that result in the establishment of a parallel start-up period, during which baseline data will be collected on the activity's performance indicators and the impact evaluations outcome variables, and that both baselines will be delivered to USAID before initiation of the activity's deliverance of the intervention under study is authorized."

IMPACT EVALUATION DESIGN PHASE

Approaches and Tools

This section discusses five approaches and related tools that are useful to support the evaluation design process, one a team has been selected:

- Organize a kickoff workshop for the impact evaluation and implementing partner teams.
- Require a review of the literature relevant to the evaluation.
- Identify impact evaluation design options and their costs.
- Undertake a scoping visit to validate impact evaluation design feasibility.
- Provide an illustrative evaluation design report outline and oversee its development and approval.

❖ ORGANIZE AN EXTERNAL PARTNER IMPACT EVALUATION KICKOFF WORKSHOP

E3/PLC, like the World Bank, is increasingly encouraging E3 offices and partner missions to hold post-award impact evaluation workshops for implementing partners and evaluation teams. These workshops take place before scoping and site selection visits occur or work on the implementing partner's monitoring, evaluation, and learning plan begins. Experience indicates that if two CORs/AORs are involved – one for the activity implementation and another for the impact evaluation – it is important for both to participate.

The post-award workshop should take place as early as possible upon award of both mechanisms. In the early months of the E3 Analytics and Evaluation Project, several potential impact evaluations faced critical challenges because kickoff workshops were delayed and missions or implementing partners lost interest or moved ahead to select sites and make other activity rollout decisions without involving the impact evaluation team. On the other hand, a couple of impact evaluations did undertake early kickoff workshops that increased collaboration between implementing partner and evaluation teams and resulted in evaluation designs that were aligned to the practicalities of the interventions and adequate to infer causality. These experiences prompted E3/PLC and the Project team to more formally integrate a kickoff workshop into its process for all new impact evaluations. Mission staff working with the E3 Bureau on impact evaluations under the Project now receive information on the kickoff workshop soon after discussions begin. This includes a description of the Project's structured approach for these workshops and a sample workshop agenda (see Table 7) that contains several standard elements, including:

- An orientation to impact evaluation for the activity implementing partner, mission staff, and country stakeholders, if appropriate.
- A discussion of the purpose of the assistance activity; USAID's hypotheses about the intended activity outcomes; the implementing partner's strategy, plans, and schedule for delivering the activity; and questions about activity effects that the impact evaluation might be able to answer.
- The evaluation team's initial ideas about how to construct a counterfactual that could support reaching conclusions about the relationship between the activity's interventions and observed changes in outcomes that are of interest to USAID and other stakeholders. If a theory of change diagram does not exist for the activity, this workshop is a good time to develop and reach agreement on one.
- A discussion about the roles and responsibilities of all evaluation stakeholders, and a review of the current timeline and next steps.

By the end of a post-award impact evaluation workshop, or through follow-up discussions, it is helpful for stakeholders to agree on and document their roles and responsibilities, as well as communication protocols, over the course of the evaluation. The importance of clarity concerning roles and responsibilities

increases as the number of actors rises. Impact evaluations of mission-funded activities that are undertaken by or with the help of an E3 office are inherently complex, as such arrangements may involve two USAID activity managers (one for the activity and another for the impact evaluation), as well as two external entities that each may report to a different activity manager.

TABLE 7: ILLUSTRATIVE OUTLINE FOR A POST-AWARD IMPACT EVALUATION WORKSHOP

Time	Day 1	Day 2
9:00 – 10:30	• Welcome • Introductions **Discussion: Why evaluate? What types of questions can different evaluations answer?** • Workshop Objectives • Agenda Review	**Evidence Review** • What has previous research in the sector shown? • What are some ongoing evaluations in the sector? • What isn't known – evidence gaps? **Group Session:** ➢ Refine Evaluation Purpose and write hypothesis as Evaluation Questions
	Break	**Break**
11:00 – 12:30	**Impact Evaluation Introduction** • Impact Evaluation Methods • Impact Evaluation in Practice	**Theory of Change** • Brief presentation guiding Theory of Change **Groups Session:** • Develop Theory of Change
	Lunch	**Lunch**
1:30 – 3:00	**Project Overview Presentation** • WA-SSD: Project Description and Implementation Plans	**Outcomes Measures Discussion** • Identifying the dependent variables for an impact evaluation and sources of data **Groups Session:** ➢ Revise outcome measures of interest and data sources
	Break	**Break**
3:30 – 5:00	**Group Session:** ➢ Evaluation Purpose: What are the goals for this evaluation? ➢ Specific hypotheses to be tested: (What do we want to learn from this project?) **Wrap-Up**	**Group Presentation Overview Discussion** **Group Session:** • Continue working on linking outcomes to TOC to Evaluation Questions • Start discussion on target population, eligible beneficiaries, and plan for roll-out of activities **Wrap-up**
	Session Ends	**Session Ends**

Time	Day 3	Day 4
9:00 – 10:30	**Compared to What?** • Why a counterfactual matter • How to randomize (unit of treatment assignment) • Other options: How equal are the target (treatment) groups and the counterfactual (non-treatment) groups? **Group Discussion: Is randomization feasible?** ➢ Evaluation Design Decision Tree	**** Group Presentation of Evaluation Design ****
	Break	
11:00 – 12:30	**Threats to Impact Evaluation** • Sample size considerations • Attrition, low take-up • Contamination and spillovers **Group Discussion:** ➢ Identify potential threats to the impact evaluation ➢ How can these threats be minimized or prevented?	**Next Steps for the impact evaluation Design – Discussion:** • Scoping activities: Uncertainties in the feasibility of impact evaluation design • Coordination between impact evaluation and implementing partner team • Baseline before the intervention starts – treatment and comparison groups
	Lunch	
1:30 – 3:00	**Group Session:** ➢ Evaluation Design: Is randomization feasible? Why or why not? Key factors to enable randomization? Other options? ➢ Minimizing threats to the validity of the evaluation	**Next Steps for the impact evaluation Implementation Stage – Discussion:** • Parallel Timelines: Creating a detailed schedule • Coordination between impact evaluation and implementing partner team
	Break	
3:30 – 5:00	**Group Session:** ➢ Finalizing Questions and Evaluation Design Method ➢ Beyond an impact evaluation: Other questions of interest and how these can be answered (performance evaluation component)? **Wrap-up**	**Dissemination Plan:** • Sharing evaluation findings: What? Who? When? How? • Timeline of important policy milestones **Wrap-up**

The authors' experiences suggest that a strong relationship between USAID's activity managers, based on a shared perception of the value of the impact evaluation, bodes well for this type of enterprise through all phases. This is also why a written summary at the end of an impact evaluation workshop or kickoff meeting can be helpful later as a reference point for handling issues that arise, or as new staff become involved. However, the E3 Analytics and Evaluation Project, which has organized some of these workshops, has found agreements about roles and responsibilities of the actors involved in an impact evaluation – while conceptually simple – can be hard to execute. Both the activity implementing partner and evaluation partner are already working under a USAID award and additional written agreements are viewed from both perspectives as potentially in conflict with existing requirements.

In a separate examination of the roles and responsibilities of impact evaluation actors, the E3 Office of Land and Urban developed an impact evaluation primer that lists the collaboration steps between USAID, the activity implementing partner, and the evaluation team, along with the lead responsible party for each step. The primer can guide an initial workshop and serve as a stand-in for a written workshop summary if questions come up later about roles.

For impact evaluations initiated with assistance from DCHA/DRG, a memorandum of understanding (MOU) is generated between the mission where an impact evaluation originates and the evaluation team. This MOU spells out the roles and responsibilities of the partners as well as those of DCHA/DRG. An MOU that emerges from a start-up workshop gives DCHA/DRG something to share with a mission director to help ensure continued support throughout the evaluation. The E3 Bureau has also experimented with MOUs for impact evaluations, but the practice is not yet institutionalized as well as with DCHA/DRG.

BOX 26: NOTES FROM THE FIELD – EVALUATION MANAGERS

"I presented about two years ago at an internal impact evaluation interest group meeting saying I thought a key factor for success was to have the same manager for the implementation pilot and the impact evaluation. I think that's why we succeeded."

- Experienced E3 Bureau Impact Evaluation COR

Some E3 offices have also produced materials on conducting evaluations in their fields that may be useful to include as handouts or pre-workshop reading material:

- Working with the multi-donor Building Evidence in Education working group, the E3 Office of Education published the sector-specific Generating Evidence in Education: Impact Evaluations.
- The Land and Property Rights Impact Evaluation Tool developed by the E3 Office of Land and Urban is another example of sector-specific impact evaluation guidance.
- The E3 Office of Global Climate Change produced a video on climate change adaptation evaluation methods and posted it to YouTube, to share this information with the practitioner community.
- Also, USAID's Bureau for Food Security collaborated with the Millennium Challenge Corporation to prepare its *principles into practice* guide, "Impact Evaluations of Agricultural Projects."

❖ REQUIRE A LITERATURE REVIEW RELEVANT TO THE EVALUATION

As an initial activity, the evaluation team that USAID selects should undertake a brief review of the relevant literature and other available documentation to determine findings of prior studies that have examined similar interventions for addressing similar problems, and whether those studies have already answered any of the questions in the evaluation SOW. Also, of interest from this type of review is information on

the size of the effects that previous studies have reported, as this information is needed to inform choices about the size of the sample needed for the impact evaluation that is being designed during this phase. The literature review should place the planned evaluation in the stream of preceding research and show what new evidence it expects to add to knowledge about the type of intervention and outcomes that will be studied. Box 27 provides an example of this type of literature scan from an evaluation design report under the E3 Analytics and Evaluation Project.

BOX 27: SAMPLE LITERATURE OVERVIEW FOR AN IMPACT EVALUATION DESIGN REPORT

The theory of change for the MAST project, which used mobile devices to map and certify land in Tanzania, described in this section is based on a review of the relevant literature, as well the evaluation team's understanding of the intended interventions.

The existing literature describes several underlying mechanisms whereby interventions, such as MAST, generate economic benefits by addressing issues related to weak property rights. Where rights to land are insecure, landholders perceive risks of conflicts related to land or expropriation by more powerful government or non-government actors. This risk weakens investment incentives by reducing the expected payoff to certain investment types. Investments that are immobile or that pay off over time (e.g. wells, soil conservation measures) are less attractive if it is uncertain that land rights will be maintained during the time during which returns to the investment accrue (Besley 1995). Activities such as MAST aim to reduce these perceived risks and the propensity for landholders to make sub-optimal investment decisions, leading to an increase in more productive investments that are longer-term or tied to the land.

Insecure property rights may also have adverse impacts by inducing cultivators to allocate resources to defend property rights when the resources could be put to more productive use. Sjaastad and Bromley (1997), among others, have noted that certain types of visible investment, such as non-productive tree planting (i.e., along boundaries), frequently occur to strengthen recognition of property rights under informal tenure systems. Field (2007) finds that insecure property rights can result in reallocation of labor to watch over property and prevent competing claims. Where property rights are secure, investment and labor allocation strategies can maximize returns. The evaluation team's scoping activities have not found much evidence of unproductive, defensive investments in the Tanzanian context, but the evaluation will nonetheless consider this possible channel of impact.

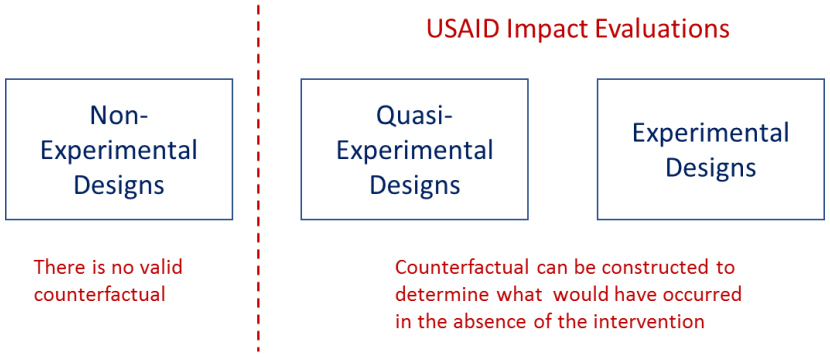
Weak property rights also limit the transferability of land. Where land cannot be easily transferred from one user to another, farm sizes will tend to be sub-optimal, as households are unable to adjust their land endowments to match their endowments of other factors and access to capital. In addition, the value of investments in land cannot be recouped in the land market if rights to land cannot be sold, which Brasselle et al. (2002) term "the realizability effect." On a broader scale, limited transferability may also reduce labor productivity by limiting the propensity of workers to migrate from areas where land-related employment is scarce to areas where land-related employment is abundant. In the case of Tanzania, land markets are limited and constrained by government restrictions and requirements for permission related to land transfers (USAID 2011). The MAST project may contribute to a broader process of deepening land markets and improving transferability of land, but the evaluation team does not expect observable shorter-term outcomes in these areas due to the limitations of Tanzanian land markets.

The literature also recognizes that the formalization and strengthening of property rights can allow landowners to use their property as collateral to obtain financial institution loans, improving access to credit. This argument was popularized in Hernando de Soto's *The Mystery of Capital* (2003). In the context of MAST activities, this impact may be limited because of weak formal land rights markets and other factors that make collateralized lending unattractive to lenders (e.g., higher returns from other types of lending, lack of experience with foreclosure). However, some impact to credit access may occur if offering land certificates, known as CCROs, as collateral serves as a lender commitment. In the evaluation team's initial scoping efforts, landholders widely cited access to credit as a potential benefit of obtaining CCROs.

❖ IDENTIFY IMPACT EVALUATION DESIGN OPTIONS AND THEIR COSTS

For most impact evaluation questions, there is more than one design that could be used to obtain an answer. The term evaluation design, as used in this volume, refers to the research design or evaluation architecture to construct the counterfactual, not the specific method used to collect or analyze data. When addressing questions about causality, most practitioners categorize research approaches as falling into one of three categories, as shown in Figure 9. In this report, variations of experimental and quasi-experimental designs are the only options discussed for questions about causality.

FIGURE 9: VALID EVALUATION DESIGNS FOR USAID IMPACT EVALUATIONS



Tools for Clarifying Impact Evaluation Design Options

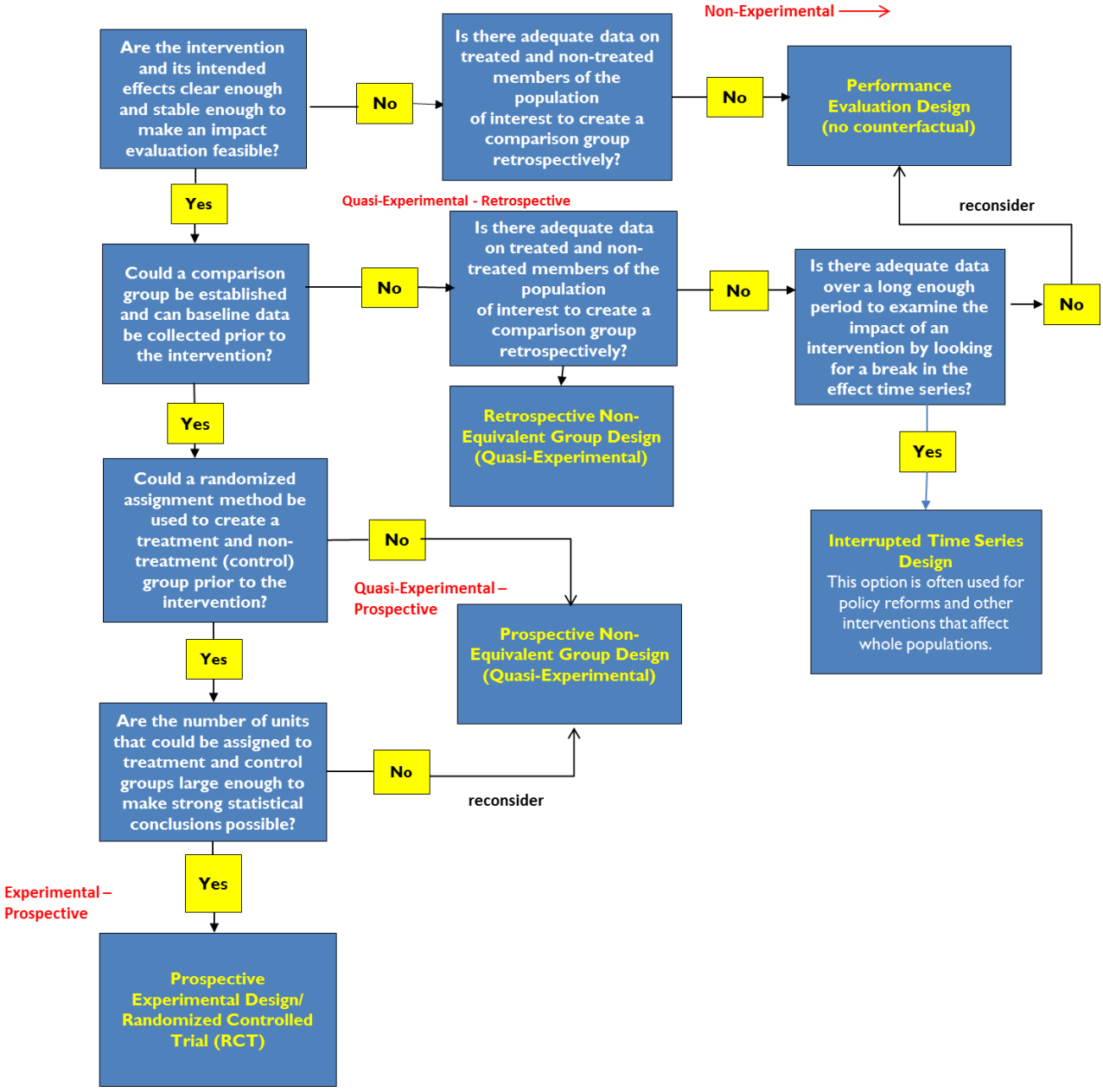
A helpful tool for considering impact evaluation design options is an impact evaluation decision tree, as shown in Figure 10. This diagram can help USAID staff and other stakeholders understand the sequence of questions to ask when trying to determine whether an impact evaluation would be feasible for a specific intervention. A USAID Evaluation for Evaluation Specialists course handout offers additional questions to consider when assessing the feasibility of an impact evaluation.

Estimating the Costs of Impact Evaluation Design Options

Before settling on an impact evaluation design, USAID's impact evaluation manager can ask to see the options that are being considered by an evaluation team and an estimate of the cost of each option along with the pros and cons. Some evaluations under the E3 Analytics and Evaluation Project provide this information in a Concept Paper prior to finalizing the research design. The example in Box 28 from a Project Concept Paper illustrates how the costs of design options can be outlined even without using dollar estimates. If a valid comparison of design options can be made, this step in an evaluation design process is worth including.

Where a full presentation of all options is not needed, a simpler mechanism can be used to share information on the likely costs or level of effort associated with the best evaluation design options, to arrive at a final choice.

FIGURE 10: IMPACT EVALUATION DECISION TREE (FOR EVALUATIONS WHERE AN EQUIVALENT COMPARISON GROUP WOULD BE IDEAL)



BOX 28: COMPARING COST IMPLICATIONS OF IMPACT EVALUATION DESIGN OPTIONS

When impact evaluation design options are first discussed, the advantages and disadvantages of each option are often discussed from a technical perspective on questions such as selection bias. When discussions about an impact evaluation initiated under the E3 Analytics and Evaluation Project seemed deadlocked on technical points, the Project team clarified the pros and cons of each option by providing USAID with information on their cost implications. This quickly reframed the discussion about options to include cost-effectiveness.

Household Survey Design Option	Number of Household Per Village	Number of Villages	Number of Data Collection Rounds	Total Number of Household Surveys
1. Regression Discontinuity Design (RDD) without an economic outcomes (productivity and incomes) module	Treatment Villages: 500 Comparison Villages: 500 divided equally among the number of comparison villages	Treatment: 1 Comparison: 10	2 (Baseline at 0 months & endline at 24 months)	2,000
2. RDD with an economic outcomes module (productivity and incomes)	Treatment Villages: 1,000 Comparison Villages: 1,000 divided equally among the number of comparison villages	Treatment: 1 Comparison: 10	2 (Baseline at 0 months & endline at 24 months)	4,000
3. Propensity Score Matching (PSM) design without an economic outcomes module	Treatment Villages: 1,500 Comparison Villages: 3,500 divided equally among the number of comparison villages	Treatment: 1 Comparison: 10	3 (Baseline at 0 months, midline at 12 months & endline at 24 months)	15,000

❖ UNDERTAKE A SCOPING VISIT TO VALIDATE IMPACT EVALUATION FEASIBILITY

Another important step in the start-up phase of an impact evaluation involves the finalization of decisions about where the intervention to be studied will be implemented and how a valid counterfactual will be constructed. Both the activity implementing partner and the evaluation team may have ideas about treatment sites and populations, or how they should be selected, by the time they meet. Site selection procedures, as well as the feasibility of other aspects of implementing partner and evaluation partner plans, thus are often a focus of initial workshops and consultations.

Based on the authors' experiences, it appears beneficial for an activity implementing partner and impact evaluation teams to join forces for their initial scoping trips. Working in parallel in the field helps produce a shared understanding of the context in which each partner will carry out their separate but related tasks. Scoping visits carried out collaboratively by evaluation teams and implementing partners have proven effective in highlighting aspects of activity and evaluation planning where ideas converged, as well as pointing out where disagreement existed, thus providing a basis for discussion about areas for compromise.

As shown in the illustrative list in Table 8, a scoping visit, either alone or with an implementing partner, offers evaluators an important opportunity to "ground truth" and refine initial ideas about the optimal design and methods for an impact evaluation.

TABLE 8: ILLUSTRATIVE IMPACT EVALUATION SCOPING VISIT TASKS

• Determine the feasibility of a counterfactual • Assess impact evaluation design options • Select an assignment method • Estimate likely attrition	• Determine sample size and power • Evaluate local survey research capacity • Estimate logistical challenges • Initiate institutional review board process, if needed

The number of design features to check during a scoping visit will vary from study to study. When both the implementing partner and evaluation teams participate, it is best for both teams to approach the scoping visit with an open mind and discuss their views on various issues that arise in a collaborative manner, seeking the best overall solutions.

- In one activity, the evaluator may require data from individual units (e.g., families, farms, school children) but due to the nature of the intervention perhaps the treatment assignment unit is at a cluster-level (e.g., villages, classrooms), which would drive up an evaluation's required sample size and may stretch the resources of the implementing partner.
- In another activity, the presence of conflict in the region and the implementing partner's draft work plan may block the evaluator's hope of randomly assigning villages to treatment and control groups, forcing the evaluator to choose among quasi-experimental options. Other factors, such as conflict-induced mobility, may drive up likely attrition estimates, again raising the sample size estimate, and logistical challenges such as seasonal issues may stretch the amount of time required to complete data collection rounds. In addition, an evaluation team may learn that institutional review board clearance is required prior to initiating field research, and that it will need to hire a local expediter who can represent the team at various stages of this process.

Each situation will likely be unique and by the end of the scoping visit, the evaluator may have a plan for executing the impact evaluation that is solid but differs greatly from what was envisioned prior to the trip.

BOX 29: NOTES FROM THE FIELD – COMMUNICATION

Notes from the Field

"We had one impact evaluation where everyone was on board early and understood what we were doing. Randomization was not complex for that impact evaluation. … On a different impact evaluation, it took a while to get everyone on the same page due to communication styles that did not work well together. In this case, they were used to working with communities that already have buy-in to their work, so the idea of randomization was difficult. Not being able to pitch their benefits to the local government was tough for the implementing partner, since this is how they usually operate. We had an issue with a sub-partner (who was eventually replaced) that consistently violated agreements between the impact evaluation and the implementing partner … which did a lot to undermine the impact evaluation."

- USAID senior advisor

❖ SUPPORT PREPARATION, REVIEW, AND APPROVAL OF THE DESIGN REPORT

This section provides support for the preparation of an impact evaluation design report in the form of an illustrative outline that incorporates a variety of topics and professional standards that go beyond the outline and checklists USAID uses for performance evaluations. The section also describes USAID's process for reviewing evaluation design reports considering the impact evaluation-specific elements of the illustrative outline.

Illustrative Impact Evaluation Design Report Topic Outline

USAID's Technical Note on Impact Evaluations discusses many of topics related to professional standards for impact evaluation designs and final reports but does provide a condensed format that would make it

easy for USAID impact evaluation managers to guide the development of a design report. To meet that need, Box 30 provides an illustrative topic outline.

BOX 30: ILLUSTRATIVE ELEMENTS TO BE COVERED IN AN IMPACT EVALUATION DESIGN REPORT

Adapted from the USAID Technical Note on Impact Evaluations and other standards such as the What Works Clearinghouse and publications guides

- Evaluation purpose, audience(s), and anticipated uses (including USAID and non-USAID stakeholders such as country partners, implementing partners, civil society).
- Background on the development problem or opportunity and USAID assistance provided (activity, project, country strategy).
- List of specific questions the impact evaluation will address (about causality, other); list must match the evaluation SOW list of questions.
- Literature review summary of prior relevant research, including evidence on questions this evaluation asks and pertinent information for calculating an appropriate sample size for this evaluation (e.g., effect sizes found for similar interventions elsewhere).
- Clear presentation of the evaluation design and methods to be used, including all study instruments (attached in an annex).
- *For evaluation questions about causality:*
  - Identification of the intervention(s) to be examined, including variations to be examined (multiple arms) and their target groups/areas.
  - Detailed (operational) description of the intervention(s) to be examined (amount, frequency).
  - Specification of outcomes the intervention(s) are expected to affect.
  - Indicators or other means by which the status of each intended outcome will be determined.
  - Type of impact evaluation proposed (experimental, quasi-experimental).
  - Description of how appropriate control/comparison groups will be constructed.
    - For experimental designs, control group unit of assignment and randomization method (e.g., lottery).
    - Specific type of quasi-experimental design and comparison group selection method (e.g., matching, regression discontinuity).
    - Whether/how quasi-experimental design choices will minimize selection bias.
  - Proposed sample size (calculated number of units required to confidently identify changes in the outcomes) and how it was calculated (show equation).
  - Parameter estimates used, including:
    - Minimum detectable effect size and its basis.
    - Confidence/significance level (.90, .95, other).
    - Power level used (.80 or .90), and basis if other than .80 as costs rise going up.
    - Variance.
    - Compensation made for anticipated attrition, and basis.
    - Other design effect adjustments to sample size (inter-cluster correlation, other).
- Identification of any other questions the impact evaluation will address and full methods description for each (e.g., cost-effectiveness, indications of likely scalability and/or sustainability, portability of intervention).
- Description of institutional review board requirements for research in the countries and sectors the impact evaluation will examine and how the evaluation has/will comply with those requirements.
- Description of study measures to protect the privacy of intervention beneficiaries and others from whom data will be collected.
- Data analysis plan for the study, organized by evaluation question and data set, including equations and other analysis specifications and anticipated analytic outputs, including analyses of gender-specific data and gender-specific or differential effects.
- Study limitations.
- Impact evaluation duration, number of data collection rounds, and cost.
- A plan for the transfer of the impact evaluation data to the DDL.

Impact Evaluation Design Report Reviews

When planning to review an impact evaluation design report, USAID's evaluation manager may consider organizing both internal and external peer reviews. The former ensures that USAID's basic expectations are met while an external peer review may provide important insights if the design is complex or the subject matter needs to be considered by technical experts who may not be available among Agency staff.

Internal Review of Impact Evaluation Designs

Once a draft is prepared, USAID's evaluation manager should organize an appropriate review process. Internally, it is appropriate and helpful to engage the same group of individuals who reviewed the impact evaluation SOW to review the design report. USAID does not have an official protocol or checklist for reviewing an evaluation design report. ADS 201 does however require that impact evaluation design reports be shared quite broadly, as shown in Box 31.

BOX 31: SHARING IMPACT EVALUATION DESIGN REPORTS

"Except in unusual circumstances, the design will be shared with country-level stakeholders as well as with the implementing partners for comment before being finalized. After finalization of the design, it must be shared with the relevant implementing partners and funders and be made available upon request to development actors in a format deemed appropriate by the mission or Washington [operating unit]."

- ADS 201

In lieu of such a protocol, the illustrative outline in Box 30, along with USAID's Technical Note on Impact Evaluations, may be adequate for an internal review.

External Peer Review of Impact Evaluation Designs

USAID does not currently have a formal process or protocol for conducting external peer reviews of impact evaluation designs. To carry out external peer reviews, the E3 Analytics and Evaluation Project developed a simple template provided in Box 32 that it refined based on feedback from reviewers. To identify appropriate potential peer reviewers, the Project team examined the literature and talked with individuals in the USAID technical office that oversees the intervention being evaluated. The Project team engaged two individuals to conduct the external peer review using the protocol in Box 32. The reviewers returned their completed peer review response to the team, which responded to the comments and then shared the reviews with USAID's evaluation manager. The team then revised the evaluation design to address comments received in both the internal and external peer reviews.

Another approach for external peer reviews would be to use the peer-review process of a well-regarded journal, which for experimental and quasi-experimental designs includes the peer-review protocol of the Publications Manual of the American Psychological Association.

BOX 32: IMPACT EVALUATION DESIGN PEER REVIEW TEMPLATE

Please answer each of the questions below to address where the Evaluation Design Proposal is successful in meeting these criteria, but more importantly where the Proposal needs to be improved or expanded on, as well as, any additional options or alternatives that should have been considered.

Each question below should be answered concisely but thoroughly. Please elaborate on key points using your own expertise or past studies that have provided insights or challenges into different components of the evaluation design.

Peer Reviewer Name: _____

Date: ___ / ___ / ___

Motivation for the Impact Evaluation:
1. Is the impact evaluation designed to address a specific policy problem so that results can be used to inform future decision-making and/or programming?
2. Does the Evaluation Design Proposal show where the evaluation is situated in the existing literature and how it will contribute to the current state of evidence?
Evaluation Questions:
3. Are the main evaluation questions stated as testable hypotheses about cause-and-effect relationships between the program and the anticipated outcomes?
4. Are all the evaluation questions consistent with the theory of change?
Theory of Change:
5. Does the theory of change clearly outline the causal pathways from the interventions to outputs to intermediate and final outcomes?
Key Outcome Variables:
6. Is each outcome variable adequately specified (i.e. clearly defined to gather reliable data that measures the construct they were designed to measure)?
7. Are important intermediate outcomes, within the scope of the evaluation, included to help explain the underlying causal mechanisms? Are there any missing outcomes that should be included?
Evaluation Design:
8. If this is a randomized experiment, does the design ensure that randomization will be conducted at the appropriate level?
9. If this is not a randomized experiment, does the quasi-experimental approach proposed select the comparison group in such a way that raises concerns about selection bias? Are there alternative methods that could be considered that would reduce selection bias further?
10. Does the evaluation design raise concerns that spillovers or contamination into the control group during implementation could compromise the evaluation?
11. If the intervention is one in that involves selection into treatment (for example, voluntary take-up) does the evaluation methodology account for this?
Power Calculations and Sample Size:
12. Is the appropriate approach to calculate power or sample size used given the evaluation design and outcomes measures of interest? Are estimates for key parameters explicit and consistent with typically used standards?
13. Is the expected minimum detectable effect size reported? Is the basis for this estimate explicit and justified?
14. If this is a clustered design, do the power calculations account for the intra-cluster correlation (ICC)? Is the basis for this estimate explicit and justified?
15. Do the power calculations explicitly account for imperfect take-up and attrition, when appropriate?
Implementation Fidelity and Monitoring Plan
16. Is there a proposed plan to monitor implementation fidelity, compliance, and external events to maintain integrity of the evaluation design?

Data Collection Methods:
17. Are the data collection methods appropriate for each of the outcome measures of interest?
18. Is the qualitative component likely to enhance the quantitative component? Does the scope of the qualitative component seem like the best approach for the study?
Data Analysis Plan:
19. Is the estimation strategy (i.e. intent-to-treat or treatment-on-the-treated) and treatment effect equation (i.e., difference-in-differences, ANCOVA, post-specification) clearly laid out, including additional methods to conduct appropriate robustness checks?
20. Does the Evaluation Design Proposal include discussion of appropriate statistical corrections including control variables, standard errors, and clustering?
21. Does the analysis plan for integrating quantitative and qualitative data and findings seem adequate?
Timeline:
22. Is the timeline for implementation of the project and evaluation aligned so that a) baseline data can be collected before the interventions start and b) enough time passes to expect changes in the outcome variables before endline measurements are collected?
Limitations:
23. Based on this review, are there any serious risks and/or limitations that should be addressed before moving forward with the evaluation?
Publication:
24. Does this evaluation design meet the standards for publication?
Additional Comments:
25. Do you have additional comments on other important factors?

IMPACT EVALUATION IMPLEMENTATION PHASE

The classic model of the implementation phase of an impact evaluation involves the establishment of treatment and comparison or control groups, consistent with the study's design; pre-testing instruments; training the data collection team; gathering baseline, possibly midline, and endline data using a cross-sectional or panel study design; analysis focused on determining the effect size of an intervention; and the preparation of the evaluation report. At every step in this process, errors can occur, and publications on impact evaluation devote lengthy sections to mistakes (in question construction, translation, enumerator training and supervision, data transfer and cleaning, and analysis) and how to avoid them. Bookending this process are important opportunities for engaging evaluation stakeholders in field locations that technical guides on survey research, observation, and focus groups often overlook.

Given the wealth of information available on field research design and management, including the World Bank's Power of Survey Design: A User's Guide for Managing Surveys, Interpreting Results, and Influencing Respondents and the study-specific data collection and analysis requirements of impact evaluations USAID funds, this section focuses selectively on a few opportunities during implementation for making an important difference in the quality and use of impact evaluations.

Opportunities for improving stakeholder engagement in the utilization of impact evaluation findings as well as for ensuring that study data and reports are valued for their quality exist throughout the implementation phase. Rather than attempting an extensive literature review on data collection and analysis of best practices and problem-solving techniques, this section focuses on opportunities unique to USAID and similar development assistance agencies or, less frequently, are the focus on impact evaluation guidance.

Approaches and Tools

Approaches and tools for addressing technical and management challenges can emerge during impact evaluation implementation. The approaches and tools here are organized in two sections, the first focused on field work aspects and the second focused on data analysis and reporting.

Field Work Aspects of Implementation

USAID impact evaluation managers play an important role in this phase of an evaluation when they:

- Facilitate mission in-briefings and a completion/pre-departure briefing;
- Oversee instrument pre-tests; and
- Encourage implementation fidelity monitoring between an impact evaluation's baseline and endline

Data Analysis and Reporting Aspects of Implementation

Similarly, impact evaluation managers help facilitate a high-quality evaluation report when they:

- Review baseline data quality in near real-time;
- Foster compliance with high standards for impact evaluation reports; and
- Supervise transfer of evaluation data to the DDL

Field Work Aspects of Implementation

❖ FACILITATE MISSION BRIEFINGS

At USAID, it is standard practice for a performance evaluation team, often comprising a mix of local and expatriate evaluators, to meet with mission staff, including the mission director or their deputy, at the start of the evaluation and when the fieldwork phase ends. Established as a courtesy and guarantee of the mission's awareness of what external teams were doing and learning, these visits have long served as opportunities to fine-tune evaluations to meet mission needs and for previewing and gaining insights into the feasibility of evaluator recommendations. With impact evaluations, more opportunities exist to meet with mission staff over the life of the study, but it is not clear from known impact evaluations if these opportunities are being actively pursued and used to better align with mission decision-making processes or to more generally build buy-in for a study and the utilization of evidence it produces.

Viewed this way, an evaluation team's most important visits with mission staff, particularly at the senior level, may be the initial scoping visit and debriefs following baseline data collection and subsequent data collection rounds. During these stages, a study design can still be modified or its early findings can be shared. Field-based kickoff workshops for implementing partners and evaluators, along with their respective USAID CORs/AORs/activity managers, as described in the Impact Evaluation Design Phase section, offer another opportunity to build impact evaluation stakeholder engagement, including with country partners, civil society representatives, and other local stakeholders. Missions vary considerably in the degree to which country partners and stakeholders are asked to actively participate in evaluation planning processes. As this paper's authors have learned from experience, including country partners can yield positive surprises. In Cambodia, a chance response in one meeting between the evaluation team and Mission staff revealed strong interest in an impact evaluation in one section of the government that was not previously known to have a strong interest in the evaluation questions and results.

Under the E3 Analytics and Evaluation Project, most impact evaluation work plans incorporate repeated opportunities for engaging stakeholders, building interest in impact evaluations findings, and sharing early results, although in some cases these opportunities for dialogue have been harder to establish. Other

USAID bureaus that support a sizeable number of impact evaluations, including DCHA/DRG and the Bureau for Food Security, reported similar experiences with efforts to build stakeholder engagement and direct involvement over the course of impact evaluations.

❖ OVERSEE INSTRUMENT PRE-TESTS

When data collection instruments are piloted or pre-tested, their developers sometimes learn that the way in which a question is worded, or translated, or even the cross-cultural understanding of a concept can make a question incomprehensible to the intended respondent. For simple surveys where what is to be measured is straightforward, the course corrections required after instruments are pre-tested also tend to be simple. However, this is not always the case for the kinds of outcome variables on which impact evaluations seek answers. As the Framing Flaws section of the main volume indicated, intended outcomes of interventions to be studied that are not well-specified, or lack metrics for assessing when their status has changed, can be an impact evaluation's Achilles' heel.

When outcomes are not operationally defined at the start of an impact evaluation, instrument pre-tests may offer a chance to catch and address that problem. This kind of situation occurs when outcomes are described in ways that people involved in an evaluation think they understand and fail to recognize that others do not. Terms like capacity, resilience, trust, quality of life, responsiveness, and empowerment all have this characteristic, yet fit nicely into causal statements of the "if X happens, then _____ will improve" variety. Even when evaluators try to break such terms down into their various dimensions when developing instruments, the data that emerge may not add up to what USAID hoped to learn.

Where there is doubt about whether impact evaluation respondents and USAID or the evaluation team understand a term or the concept behind it in the same way, a substantial investment in instrument pre-testing can be valuable. This may include piloting alternative versions/wordings of questions to find common terms and meanings, or less formal but highly interactive ethnographic dialogues with people who live near and have the same level of education as those who will eventually be interviewed. As this kind of investment in measurement accuracy is not standard practice for survey research teams at USAID, such efforts to ensure alignment may not be undertaken, even when they are needed.

❖ ENCOURAGE IMPLEMENTATION FIDELITY MONITORING AND REASSESSMENT

Implementation fidelity monitoring is a formal element in some impact evaluations of health and education programs in the U.S. Informally, it is part of many other domestic and international impact evaluations. From an evaluation perspective, it is important to understand whether intervention delivery conformed to activity plans and adhered to the agreed evaluation design requirements, as evaluation reports tend to follow their approved designs and report whether outcomes were attributable to what the activity design promised to deliver. Moreover, measuring implementation fidelity can provide a more thorough understanding of the theory of change and an intervention's contribution to outcomes. Evaluators need to know about any deviation in the promised delivery, so they can report attribute findings accurately (e.g., the intervention delivered as planned and to assigned areas, or some modification to that planned intervention or assigned areas). Otherwise, it cannot be determined whether a lack of impact is due to poor implementation or inadequacies inherent in the interventions themselves.

In discussions with evaluators who undertake impact evaluations for USAID, the authors found that funds for implementation fidelity monitoring are rarely provided, and USAID evaluation SOWs seldom call for a specific process to check on the fidelity of the activity being evaluated to its original design. When an impact evaluation calls for implementation fidelity monitoring, evaluators generally focus on the degree to which activity implementation adhered to the intervention plan with respect to the numbers and types of recipients reached and delivery locations.

Attention to what was delivered (e.g., volume, frequency, duration) is also normal for implementation fidelity monitoring, as is participant responsiveness (take-up and dropout), crossovers and spillovers, and site-specific variations in how the intervention was implemented. Measurement procedures vary, but generally involve a comparison between (1) what the monitor observes or is told by providers and recipients and (2) the activity implementation team's written plans and agreed evaluation design requirements.

Data Analysis and Reporting Aspects of Implementation

❖ REVIEW BASELINE DATA QUALITY IN NEAR REAL-TIME

To independently verify the quality of impact evaluation baseline data and reports it was receiving for evaluations it had funded, and to highlight information that might be immediately useful to USAID evaluation managers, the E3 Office of Land Tenure and Resource Management collaborated with the E3 Analytics and Evaluation Project to develop a data review protocol. The protocol enables a structured assessment of ongoing impact evaluations, examining the quality of the evaluation's design and associated data. The protocol, provided in Box 33, is a set of questions that an experienced impact evaluator who was not involved in the impact evaluation will answer The items in the protocol consider a range of critical components and characteristics of high-quality impact evaluations. This differentiates the protocol from standard review checklists as it allows for a scale of quality rather than a binary mark of completeness. There are two levels of the protocol: an overarching measure of quality and specific questions related to the design and data quality. This tool can be adapted to review impact evaluation baseline or endline data reports in any sector and assists in the management of quality impact evaluations. The Project team applied this protocol to six baseline studies, one midline study, and two endline studies for land tenure impact evaluations in Africa.

The Cloudburst Group and USAID conducted an impact evaluation in Zambia to measure the effect of securing property rights on the adoption of climate-smart agricultural practices. *Credit: The Cloudburst Group*.

BOX 33: IMPACT EVALUATION DATA REVIEW PROTOCOL

Description of Scoring for Impact Evaluation Review Protocol

I	II	III	IV
Fully satisfactory: the evaluation meets high standards on this criterion, and no revisions are suggested.	**Satisfactory with minor issues**: the evaluation is broadly on track to meet high standards of quality and generate findings consistent with the expectations of the design report. However, some revisions are suggested, and/or the evaluation will have some minor shortcomings in terms of the power, precision, or scope of the findings that were anticipated in the design report.	**Raises concerns**: the outcomes of the baseline on this criterion suggest significant concerns for the evaluation. While the evaluation may still be able to produce meaningful findings, the analytical outputs will be significantly weaker and/or narrower than what had been envisioned in the design report. Substantial revisions may be required.	**Raises major concerns**: the outcomes of the baseline on this criterion indicate that the evaluation design has been compromised to the extent that it may not be worth carrying out the remainder of the evaluation.

Illustrative Items from Impact Evaluation Review Protocol

Key Issues (scored according to table above)	Rating (I-IV)
1. The indicators from the design report are included and appropriately measured in the survey.	
2. The questions reflect an appropriate set of control variables for the analysis.	
3. The documentation of the dataset is clear, complete, and user-friendly.	
4. There are no errors or inconsistencies in the data that raise questions about its quality.	
5. There were no problems that occurred during the data collection process that raise concerns about the quality of the data.	
6. The results of the baseline survey do not raise any concerns about the statistical power of the evaluation.	
7. The results of the baseline survey do not raise any concerns about the balance between the treatment and control/comparison groups.	
8. The baseline report adequately addresses the data collection process, statistical power, and balance issues.	
9. The presentation of descriptive statistics in the baseline report is thorough and informative.	

Detailed Issues Related to Data Quality and Design

Detailed Issues (Scored Yes, Yes with Minor Exceptions, No, N/A)	
Dataset Quality	**Rating**
Were the files easily convertible to commonly used statistical software packages such as STATA or SPSS?	
Do the variables and variable labels match across files? (if applicable)	
Does the codebook describe all variables in each dataset?	
Does each question on every instrument have a corresponding variable?	
Are any other variables identified and labeled?	
Is the construction of composite variables explained thoroughly?	
Does each variable have the correct question number either in the label or in the variable name?	
After tabulating or running frequency checks on every variable, is the dataset free of any inconsistencies?	
Is the structure of each variable logical and easy to understand?	
For example, are the "check all that apply" questions presented as binary variables for each option?	
Does each dataset have a logical unique identifier?	
For multiple identifiers, can they be linked together in a logical manner?	
Is the dataset free of duplicate cases?	
Are skip patterns recognized with skip-pattern codes?	
Are the following coded uniquely by question: refused, legitimate skips, "don't know", and missing (other)? These will be referred to as "disposition" codes.	
Based on the descriptive statistics, do all binary and categorical variables appear to include only the allowable values?	
Is the overall response rate for each variable reasonable?	
Have all potentially identifiable characteristics of the data been adequately removed to protect the confidentiality of all respondents?	
Design Implications	
Are the indicators from the design report included in the baseline report and appropriately measured?	
Do the questions reflect an appropriate set of control variables (such as demographics, SES, etc.) for the analysis? Are there additional control variables that should have been included?	
Are the wording, organization, and clarity of the questionnaires enough?	
Do the results of the baseline survey confirm the power calculations from the design report?	
Are the treatment and comparison/control adequately balanced along all key outcomes and covariates?	
Does the report provide a clear description of the data collection process?	
Were the supervision and quality control measures adequate to provide confidence in the dataset?	
Were any unexpected challenges that arose during the data collection process addressed in a way that will not raise issues for the eventual analysis?	
Are descriptive statistics and cross-tabulations presented thoroughly and clearly?	
Is the formatting informative and appropriate? (I.e. Is the sample size reported for every table?)	

❖ FOSTER COMPLIANCE WITH HIGH STANDARDS FOR IMPACT EVALUATION REPORTS

To meet high standards for impact evaluation reports, evaluation teams need a combination of guidance and feedback.

Provide Impact Evaluation Teams with Adequate Guidance to Meet Standards

USAID has extensive evaluation report guidance, including the Evaluation Report Requirements, the How To Note: Evaluation Reports, and the Evaluation Report Review Checklist and Template. In addition, evaluations are expected to:

- Comply with branding and marking norms (see USAID's Graphic Standards and Partner Co-Branding Guide, ADS 320: Branding and Marking), the sample Evaluation Report Template, sample evaluation covers, conflict of interest disclosure forms, and guidance on preparing statements of differences.
- Ensure reports meet requirements for those who are disabled (see ADS 302mak, USAID Implementation of Section 508 of the Rehabilitation Act of 1973).
- Accept limitations on disclosure, in some cases (see ADS 201mae, Limitations to Disclosure and Exemptions to Public Dissemination of USAID Evaluation Reports).

While adequate for performance evaluations, these guidance materials are not sufficiently impact evaluation-specific to help teams meet high standards for this subset of evaluations. The Agency's main resource on impact evaluation is its Technical Note on Impact Evaluations, which expects that all USAID impact evaluations will comply with its expectations for:

- A clear explanation of the hypotheses tested through the evaluation;
- How the treatment and comparison groups were constructed, including the unit of analysis and unit or level at which randomization occurred (individual or cluster), and the planned sizes of these groups as well as their actual sizes (net of dropouts);
- The power calculation level that the study samples met at the design stage (e.g., .80, .90);
- For cluster-design studies, the design effect or intra-cluster correlation coefficient used and the basis for that choice from prior data sources or similar study designs;
- The minimum detectable effect size the evaluation selected to gauge an effect and the basis for that choice in the literature or from specific prior knowledge;
- A discussion on the study's dropout rate, as well as spillovers and crossovers, and whether these were measured and their likely impact analyzed (e.g., the effect of a high dropout rate in the treatment group and the study's power calculations); and
- The actual effect size detected during data analysis for each dependent variable examined.

Box 34 provides a suggested outline for a final impact evaluation report based on these and other professional norms. To encourage professional reporting on impact evaluations, the outline can be attached to an evaluation SOW as an annex or circulated at an impact evaluation team kickoff workshop.

BOX 34: ILLUSTRATIVE ELEMENTS TO BE COVERED IN AN IMPACT EVALUATION FINAL REPORT

Adapted from the USAID Technical Note on Impact Evaluations and USAID Evaluation Report Requirements

- Abstract (250 words).
- Executive summary (2-5 pages).
- Indication, as early as the title page and executive summary, that this is an *impact evaluation* report.
- Evaluation purpose, audience(s), and intended uses including USAID and non-USAID stakeholders (e.g., country partners, implementing partners, civil society).
- Background on the development problem and USAID assistance provided (activity, project, country strategy).
- List of the specific questions the impact evaluation addressed (about causality, other); list must match evaluation SOW list of questions).
- Brief literature summary of the extent and findings of prior research on the same or similar interventions.
- Clear presentation of design and methods used (summary in report plus methods annex). This section should include any modifications or deviations from the original evaluation design, particularly changes to the implementation of the interventions that may have affected the validity or robustness of the evaluation design.

For evaluation questions about causality:

- Identification of the intervention(s) examined.
- Specification of the outcomes of interest.
- Indicators or other means by which the outcomes of interest were measured.
- Type of impact evaluation undertaken (experimental, quasi-experimental).
- Description of how appropriate control/comparison groups were constructed.
    - For experimental designs, control group unit of assignment and randomization method (e.g., lottery).
    - Specific type of quasi-experimental design and comparison group selection method (e.g., matching, regression discontinuity).
    - Whether/how quasi-experimental design choices minimized selection bias.
    - Sample size (calculated number of units required to confidently identify changes in the outcomes) and how it was calculated (*show equation*).
    - Parameter estimates used, including:
        - Minimum detectable effect size and its basis.
        - Confidence/significance level (.90, .95, other).
        - Power level used (.80 or .90).
        - Variance.
        - Compensation made for anticipated attrition, and basis.
        - Other design effect adjustments to sample size (e.g., inter-cluster correlation).

For other evaluation questions:

- Description of institutional review board approvals obtained.
- Description of study measures taken to protect the privacy of intervention beneficiaries and others from whom data are collected.
- Data analyses undertaken, organized by evaluation question, including equations and other analysis specifications; and analytic output (in appendices), including analyses of gender-specific data and gender-specific or differential effects.
- Study limitations.
- Findings (clear answer/detailed findings for each evaluation question), including for questions about causality:
    - The actual attrition rate from the intervention, for each intervention-outcome combination studied.
    - Intervention implementation fidelity (amount, frequency, planned/actual unit cost of delivery); spillovers, cross-overs.
    - The effect size (impact) found for each intervention examined in relation to each outcome that intervention was expected to affect.
    - The statistical and practical significance of the effects (impact) found.
- Conclusions
- Recommendations
- Annexes: instruments and other annexes as required

❖ PEER REVIEWS OF DRAFT AND FINAL IMPACT EVALUATION REPORTS

Responsibility for the preparation of an impact evaluation final report lies with the external team that conducted the evaluation. The quality of final reports, however, also reflects the effort USAID staff put into making external evaluators aware of Agency evaluation reporting requirements, including the responses to evaluation questions and the systematic approach to presenting findings, conclusions, and recommendations. USAID contributes to the quality of evaluation reports through the quality reviews it conducts for these products.

Internal Peer Reviews of Impact Evaluation Reports

USAID requires that draft evaluation reports be subject to a peer review by Agency staff. An ADS 201 *Additional Help* document provides guidance on Managing a Peer Review of a Draft Evaluation Report, and USAID's Criteria to Ensure the Quality of the Evaluation Report advises peer reviewers to determine if individual reports meet high-quality standards. Another available tool is the evaluation report checklist included in PPL/LER's Meta-Evaluation of Quality and Coverage of USAID Evaluations (pages 153-155). A variation on this checklist is in USAID's Evaluation Toolkit.

Evaluation reports are reviewed twice at USAID. They are first reviewed at the draft stage to ensure that all the questions in the SOW were addressed; assess the degree to which the evaluation conformed to its design report; and understand the findings and any problems encountered in obtaining them or updates on the design stage statement of expected study limitations. This process is laid out in Managing the Peer Review of a Draft Evaluation Report. The second review examines the final report for compliance with all USAID expectations, including the report's adherence to USAID's evaluation report template and its checklist against ADS evaluation report requirements and USAID Evaluation Policy, as summarized in the Agency's Evaluation Report Review and Template guidance.

External Peer Reviews of Impact Evaluation Reports

The resources described above, particularly the Evaluation Report Review and Template, are helpful to ensure evaluation reports meet ADS guidelines. While this checklist focuses on a report's compliance with Agency guidance, it does not provide enough information to assess the quality of the evaluation and analysis itself. Thus, in addition to an internal peer review, it is beneficial to conduct an external peer review to assess the implementation of the evaluation design, the analytical methods, and the presentation and interpretation of findings. This helps ensure quality and can strengthen the credibility of the findings to facilitate utilization of the evaluation results for decision-making.

Box 35 provides a template that the E3 Analytics and Evaluation Project developed and used for external peer reviews of impact evaluation reports, following the same process discussed for the peer review of the design report. However, the peer review template for the final report includes additional items on fidelity to the evaluation design, data analysis, and the presentation and interpretation of results.

BOX 35: IMPACT EVALUATION REPORT PEER REVIEW TEMPLATE

The primary goal of the Impact Evaluation Report is to provide USAID with information about the impact of its programs for evidence-based decision-making. Through a Peer Review process, USAID would like to gather expert opinion about the quality of this evaluation report. The following document contains a series of questions reflective of USAID criteria representative of a technically sound and successful Evaluation Report. As the reviewer, you are being asked to answer the questions below to the best of your ability, honestly, and concisely but thoroughly. The feedback you provide will inform the Agency on the strengths and weaknesses of the report to ensure the findings are applied appropriately. Please keep in mind that the evaluation report was prepared for a general USAID audience, rather than a technical or academic one, and therefore may not contain the level of detail required for a peer-reviewed publication.

Peer Reviewer Name: _____ **Date:** ___ / ___ / ___

Utilization for the Impact Evaluation
1. Does the impact evaluation's design and analytical approach appropriately address the evaluation questions so that results can be used to inform future decision-making on programming? Please elaborate if yes or no, and what could have been better approaches or methods.
Evaluation Design Implementation
2. What are the key strengths and weaknesses in the evaluation design and methodology? Please elaborate your answer for each of these components: a. Design: b. Data Collection Methods: c. Analytical Models:
3. Does the report provide sufficient information to assess whether attrition or incomplete participation into treatment pose a threat to the evaluation's estimation of impact? Are details on sample size across data collection rounds and their implications for statistical power clearly stated?
4. Does the report specify adequate steps taken to mitigate common threats to the integrity of the evaluation, such as non-equivalence at baseline, non-compliance, and spillover?
Findings
5. What finding or result do you see as the study's most important contribution?
6. Reviewing the specific findings for the thematic areas, based on your expertise, how well did the study answer the evaluation questions regarding the outcome measures of interest?
7. If the study were repeated, what changes would you recommend (if any), and how would these alterations improve the results?
Limitations
8. Are there limitations to the evaluation design, estimation strategy, or data collection approach that are not mentioned, but should be considered and accounted for in the report?
Interpretation
9. Are there conclusions within the context of the evaluation questions and findings presented that are not considered or presented in the evaluation report? What additional analysis would need to occur to draw these conclusions?
Publication
10. While noting that this is not a report prepared for publication, does this evaluation have the potential to meet the standards for peer-reviewed publication?
Additional Comments
Do you have additional comments?

❖ SUPERVISE DATA TRANSFER

In parallel to preparing an evaluation report, USAID evaluation teams are expected to prepare a data transfer package to submit to the DDL shortly after the final evaluation report is approved. USAID policy requires that these evaluation datasets include supporting documentation such as code books, data dictionaries, scope, and methodology used to collect and analyze the data.

To this end, it is important for all USAID evaluation teams to be aware of this requirement at the start of the evaluation and comply with all provisions by the time the final evaluation report is approved. It is important for at least one team member to fully understand and be able to implement actions required by the U.S. government's memorandum for heads of agencies, M-13-13, Open Data Policy—Managing Information and as an Asset, and ADS 579: USAID Development Data.

USAID staff and evaluation teams may also find the items in Box 36 useful for this task.

BOX 36: ADDITIONAL EVALUATION DATA SUBMISSION RESOURCES

- Project Open Data
- Guidance on Submitting Geospatial Data Assets to the USAID Development Data Library
- DDL Registration Form
- Frequently Asked Questions on USAID's Open Data Policy

IMPACT EVALUATION DISSEMINATION AND UTILIZATION PHASE

While the work of an external evaluation team often ends with the submission of a final report, USAID's commitment to an evaluation extends well beyond this milestone. On a post-evaluation basis, USAID's impact evaluation manager plays a leading role in the dissemination of evaluation findings and often organizes USAID's review of evaluation recommendations. This is when decisions are made about which recommendations to adopt and an action plan is developed laying out responsibilities and deadlines for their implementation. Tracking the status of commitments included in a post-evaluation action plan may also fall to the impact evaluation manager or the program office that commissioned the evaluation. In this final phase, capturing information on whether an impact evaluation was utilized and how its utilization affected development outcomes may also warrant attention.

BOX 37: FOCUS ON DEVELOPMENT EFFECTIVENESS

"In the end, the measure of our success will not be predicated on the number of evaluations done, or stored within a database, or even solely upon the quality of the findings. We'll be successful if, and when the evaluation work of USAID contributes to greater development effectiveness."

- USAID Evaluation Policy (2011)

USAID's 2011 examination of evaluation utilization at the Agency[1] found gaps in post-evaluation practices in this phase, relative to expectations set forth in ADS 201 and its evaluation resource,

[1] Overall, that study found evaluation utilization to be strong, with 93 percent of evaluations reportedly leading to knowledge gains and modifications of existing activities or contributing to the design of follow-on activities or a mission's CDCS.

Utilizing and Learning from Evaluations (2015). As Box 38 shows, a large gap existed between USAID's dissemination of evaluation findings to Agency staff and dissemination of findings and recommendations from the same evaluation to implementing partners and host-country counterparts. However, 19 percent of study respondents indicated that implementing partners had made changes in their activities because of USAID evaluations, and 8 percent reported partner country utilization of evaluation results.

BOX 38: EVALUATION DISSEMINATION BY USAID

"USAID's dissemination of evaluation results through briefings, report dissemination, and other events was reported to be strong to USAID staff (97 percent of the survey responses), but weaker in reaching USAID implementing partners (76 percent of responses reported this type of dissemination) and country partners (43 percent reportedly disseminated to country partners)."

- Evaluation Utilization at USAID, page viii

With respect to compliance with post-evaluation actions aimed at promoting evaluation utilization, survey respondents and group interview participants indicated that ADS-recommended post-evaluation review meetings and decisions on the recommendations to adopt occurred for around half of all evaluations. Post-evaluation action plans were prepared for about 40 percent of evaluations. Action tracking of the implementation status of adopted recommendations occurred for about 20 percent of USAID evaluations.

Approaches and Tools

USAID staff may want to consider four clusters of approaches and tools for the post-evaluation phase of an impact evaluation:

- Update and implement impact evaluation dissemination plans.
- Synthesize and distribute findings from multiple evaluations.
- Develop a post-evaluation action plan and track action status.
- Capture and share impact evaluation utilization evidence and stories.

❖ UPDATE AND IMPLEMENT IMPACT EVALUATION DISSEMINATION PLANS

USAID's Evaluation Policy and ADS 201 assign responsibility for the development and implementation of evaluation dissemination plans to missions and Washington-based operating units. Such plans are the mechanism through which USAID expects staff to "promote transparency and learning by disseminating evaluations when the evaluation report has been completed. Missions and [operating units] should follow the dissemination plan developed during the evaluation planning stage [and] should openly discuss evaluation findings, conclusions, and recommendations with relevant partners, donors, and other development actors" (ADS 201). USAID also produced a guide on Developing an Evaluation Dissemination Plan (2016).

Page 43 of this annex discussed preliminary dissemination plans for impact evaluations as a planning tool to help ensure that post-evaluation products and services that USAID will need (e.g., oral briefings, two-page briefs for senior managers, slide decks staff can use to do their own presentations) are listed among the impact evaluation SOW's deliverables. Dissemination plans need to be updated as the completion date for an impact evaluation nears. While there is no fixed rule with respect to when a dissemination plan must be finalized, USAID sometimes focuses on this task soon after receiving the draft evaluation report.

Awareness and use of evaluation results beyond the country where an evaluation was undertaken were reported as negligible, as was the use of USAID evaluations to help develop or update policies for sectors and problems on which the Agency works.

USAID and other donor guidance included in Box 20 provides ideas and examples for preparing dissemination plans. Another helpful source for dissemination ideas is the USAID Biodiversity and Development Research Agenda, This document describes dissemination activities involving social media, which E3's Office of Forestry and Biodiversity describes as being as important for reaching a general audience as peer-reviewed publications are for reaching the academic community.

Box 39, which comes from USAID's guidance on Developing an Evaluation Dissemination Plan, includes helpful questions and ideas to review prior to finalizing an impact evaluation dissemination plan.

BOX 39: USAID TEMPLATE FOR DEVELOPING AN EVALUATION DISSEMINATION PLAN

Audience	Goal	Tool/Medium	Forum	Responsible Party	Timing
Identify stakeholders by asking" "Who is likely to be affected by the evaluation and its results? Who is likely to be interested?"	Are we simply pushing out information? Hoping to effect change? Contributing to the knowledge base?	These may include reports, briefs, presentations, blog posts, meetings, facilitated discussions, videos, journal articles, press releases, graphics, emails, and Listservs.	Are there existing networks or venues through which findings should be disseminated? Or will communications be distributed directly to target audiences?	Who is responsible?	Is there a deadline?

Updating an impact evaluation dissemination plan provides USAID's evaluation manager with an important opportunity to re-engage Agency, country partner, and other stakeholders who represented the intended user community for an evaluation. Given the length of many impact evaluations, it is almost inevitable that some individuals who represented likely user organizations will have moved on, or the agendas and evidence for decision-making needs will have changed. Some may have forgotten about the impact evaluation. To ensure the likelihood of an impact evaluation's results getting wide attention and sparking interest across a broad range of stakeholder groups, some fanfare in advance of the report's release may be in order. The Evaluation Utilization at USAID study revealed that Agency staff have not consistently invested in post-evaluation efforts of this sort, and in many cases have done no more than submit a copy of the final evaluation report to the DEC.

An enumerator discusses impact evaluation findings with a community member. *Credit: Inzy Media.*

In recent years, USAID's attention to learning from and sharing knowledge gained from evaluations stimulated dissemination and other post-evaluation activity aimed at promoting utilization. But the number of Agency operating units and evaluation managers who prioritize building an audience for impact evaluation results and actively promoting the utilization of evaluation results remains quite low, in the authors' experience. There are, however, exceptions.

As part of an impact evaluation for the E3 Office of Land Tenure and Resource Management, an evaluation team from the Cloudburst Group organized enumerator presentations of study findings in the communities from which it collected data as part of an inclusive dissemination strategy. This story is summarized in Getting Data Back to Communities.

Other examples of innovative Agency dissemination approaches are described below:

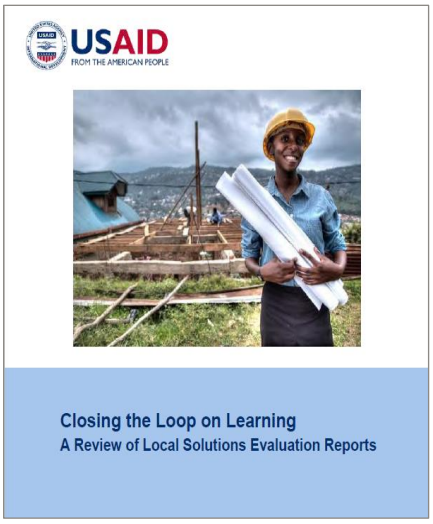
- Focusing on impact evaluations and illustrating its vision of widespread evaluation dissemination, the E3 Office of Land and Urban created a Land Tenure Evidence Hub, where it has posted synopses of impact evaluations and draws Agency-wide lessons about what it is learning. Going beyond print media to disseminate the findings of its land tenure impact evaluations, the office condensed many lessons into a presentation for a large USAID staff conference and posted it online. (Click on the photo to watch.)

- Another dissemination example that could be adapted to explain evaluation findings is a two-minute movie that describes the intervention and its impact, posted on USAID's YouTube channel. (Click on the photo to watch.)
- Beyond USAID, 3ie regularly uses Facebook to disseminate information about evaluations, methods, and events. Registering impact evaluations at 3ie's repository where evaluation findings, synopses, and other dissemination products can also be archived offers another approach for disseminating evaluation lessons.

❖ SYNTHESIZE AND DISTRIBUTE FINDINGS FROM MULTIPLE EVALUATIONS

The dissemination of findings from individual impact and performance evaluations is an important first step. Ensuring that important evaluation findings reach a broad audience over an extended period requires attention to dissemination that goes beyond an individual evaluation, and often requires an effort to shorten and sharpen evaluation results messaging and synthesize what has been learned from multiple evaluations. In this arena, the E3 Bureau is an Agency leader in developing products that cover all evaluations in the sectors it supports, as well as for impact evaluations specifically. E3/PLC has produced annual sectoral syntheses of evaluation findings since 2012 and institutionalized this process within the bureau. Its Sectoral Synthesis of 2013-2014 Evaluation Findings covered 117 evaluations (mostly performance evaluations), and the 2015 edition examined 92 evaluations.

These volumes have been well received by Agency leadership. Also, the Bureau for Global Health Bureau has produced roughly 15 highly rigorous systematic reviews of findings from experimental and quasi-experimental studies on the effects of innovative interventions relevant to health and family planning programming. The peer-reviewed journal articles that describe these studies are available on the DEC. To encourage the production of evaluation syntheses on more topics, PPL/LER commissioned Learning from Evaluation Syntheses (2019), a technical report on evaluation syntheses. One of the 13 synthesis efforts this volume examined in depth was the 2016 Bureau for Food Security evaluation synthesis of findings related to USAID's Feed the Future learning agenda. The same year, PPL/LER commissioned the Closing the Loop on Learning: A Review of Local Solutions Evaluation Reports evaluation synthesis.



Building on the technical report described above, USAID issued the Discussion Note: Making Evidence Accessible through Evaluation Syntheses (2019). This Note includes information for USAID managers on conducting future evaluation syntheses to foster learning and support evidence-based decision-making.

❖ DEVELOP A POST-EVALUATION ACTION PLAN AND TRACK ACTION STATUS

Since 2003, USAID guidance (now ADS 201.3.5.18) has called for post-evaluation meetings to review recommendations and, for those that are accepted, assign responsibility for implementation along with a schedule for implementation action. Discussions with E3 staff indicate that while some have participated in the kind of post-evaluation decision-making reviews the ADS recommends, others have not. The ADS also calls for the creation of a post-evaluation action plan. USAID has posted supplementary guidance on holding post-evaluation reviews and the preparation of Post-Evaluation Action Plans (2017).

Ensuring that such action plans are implemented is a second important step. PPL/LER's Evaluation Utilization at USAID study identified five Agency operating units that, as of 2016, were using electronic action plan trackers to follow up on evaluation recommendations that the Agency committed to implement in post-evaluation meetings. Box 40 includes a sample action tracker template from the guidance on Post-Evaluation Action Plans. As this example shows, action trackers are conceptually simple. The intention in USAID's guidance is for all missions and Washington-based operating units to install and use action tracking tools to monitor the implementation of accepted recommendations for all their completed evaluations, as this process is equality appropriate for performance and impact evaluations.

BOX 40: POST-EVALUATION ACTION PLANS ENCOURAGE THE UTILIZATION OF IMPACT EVALUATION FINDINGS

Post-Evaluation Action Plan Template

Evaluation Title:
Evaluation Completion Date:
Date of Management Response:

No.	Management and Program Actions Needed	Reason for Action	Individual Responsible for Completing Action	Budget Allocated (if applicable)	Date for Completion of Action	Status of Actions (as of date)
1						
2						
3						
4						
5						

The authors are unaware of any missions that have made their evaluation action trackers public, but some international development organizations do this. Notably, the World Bank's Management Action Record Database can be accesses by any user of its website, as Box 41 illustrates. Its color-coded ratings are a user-friendly feature of this action tracker. The European Commission also has an evaluation recommendations action tracker that is publicly accessible via its website.

❖ CAPTURE AND SHARE IMPACT EVALUATION UTILIZATION STORIES

USAID does not have published protocols for: gathering and sharing information about how impact evaluation results have been utilized to inform evidence-based decision-making; determining which impact evaluation findings have resulted in programming or policy changes; or assessing whether and how interventions that impact evaluations have proven to be effective have been scaled up.

Good practices in this regard include a modest effort to document examples of the effects of USAID evaluations in the Agency's Strengthening Evidence-Based Development: Five Years of Better Evaluation Practice at USAID – 2011-2016, from which the example in Box 42 is drawn.

BOX 41: WORLD BANK ACTION TRACKER FOR FOLLOWING UP ON EVALUATION RECOMMENDATIONS

Management Action Record Database

IEG | WORLD BANK GROUP

The Management Action Record (MAR) is a tool that facilitates the annual follow up on the adoption of IEG's recommendations by the Bank Group.

In 2012, World Bank Group Management and IEG agreed to adopt specific, measurable actions in response to IEG's findings and recommendations. As the result, IEG rates the progress of Management's Action Plans for each evaluation for four years during the annual update cycle.

Since 2015, to capture the progress of implementation of the Action Plans more accurately, WBG and IEG management introduced a new rating and started using a five point rating scale. This change introduced "Complete" as the highest rating accounting for actions that are completed and ensured that "High" is used for actions that are nearing the 90% of completion status.

MAR Ratings Legend: Complete High Substantial Moderate Negligible Not Accepted Not Rated

Search | Report Titles | Apply | Clear Results

	Report Title	Organization	Recommendation	Tracking Since	Accepted	IEG Rating	Management Rating	View Data
1	Behind the Mirror: A Report on the Self-...	IFC	Expand voluntary evaluations that respond to...	2017	Yes			Read More
2	Behind the Mirror: A Report on the Self-...	World Bank	Expand voluntary evaluations that respond to...	2017	Yes			Read More
3	Behind the Mirror: A Report on the Self-...	IFC	Formulate a more systematic approach to...	2017	Yes			Read More

More extensive efforts to document the utilization of evaluation findings, including from impact evaluations, that might serve as models for the future include:

- The World Bank's Influential Evaluations report (2004) used a case study approach to document changes resulting from the use of evaluation findings.
- Esther Duflo and Michael Kremer's Which Innovations Reach More Than 100,000 or One Million People? Evidence from the Development Innovation Ventures Portfolio report (2015) documents instances of scaling up solutions based on 24 randomized controlled trials and other evaluations conducted for USAID DIV grants.
- The J-PAL website space Evidence to Policy includes 17 case studies of impact evaluation utilization that led to policy change.

These reports illustrate that when evidence from rigorous evaluations of donor assistance is acted upon and demonstrably increases aid effectiveness, there is value in investing in high-quality evaluations. Communicating to Congress and the public how the utilization of evaluation findings improves USAID and country partner development effectiveness could also bolster support for the Agency's work. Strengthening Evidence-Based Development: Five Years of Better Evaluation Practice at USAID drew attention to Agency efforts to capture utilization stories about how USAID-funded evaluations, particularly impact evaluations, have been utilized. Box 42 reproduces one example from that volume.

BOX 42: USING EVALUATION FINDINGS CAN LEAD TO PROGRAM AND POLICY CHANGE

USING EVALUATION RESULTS TO CHANGE GOVERNMENTAL POLICY

USAID/Ethiopia Supporting Highly-Vulnerable Children

Program Description

Since 2011, USAID has funded the Yekokeb Berhan program in Ethiopia to support highly-vulnerable children and their families. Working with children who have lost one or more parents, the program builds a network of volunteers and community-level caregivers to provide for the needs of these children. One of the key areas of focus is in health, including vaccination coverage as well as HIV testing.

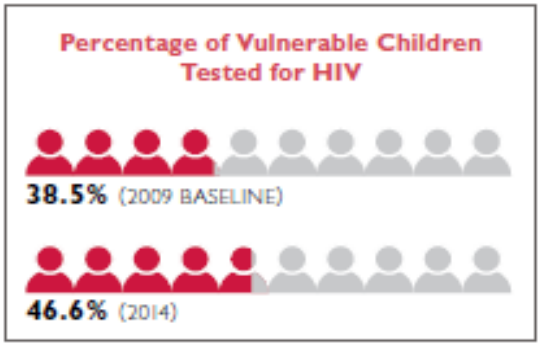
A 2014 mid-term evaluation collected evidence to understand how the program was performing.

Results

The evaluation found that USAID funding increased the percentages of the most vulnerable children in Ethiopia who had been tested for HIV, were provided with their test results, and were on antiretroviral therapy.

However, despite this positive progress, still less than half of the vulnerable children population had been tested for HIV and knew their status. This was an important finding because these children are at an elevated risk for contracting HIV.

USAID went beyond measuring program performance to using its evidence to help change the Government of Ethiopia's guidance on HIV testing for vulnerable children.



Source: *Yekokeb Berhan Program for Highly Vulnerable Children in Ethiopia: Mid-Term Evaluation Report* (June 2014)

Actions Taken

USAID staff expressed to Ethiopian government counterparts their suggestions that highly-vulnerable children should be prioritized for counseling and testing services.

The Federal Ministry of Health and Ministry of Women and Children's Affairs agreed, and in 2014, the National Guidelines for Comprehensive HIV Prevention were revised. The new version includes helpful guidance on disclosing HIV status results to testers, including vulnerable children. As a result of this policy change, HIV testing for vulnerable children has been prioritized both for USAID programming and for the Government of Ethiopia.

To this end, several initiatives might be considered, including:

- Routinely include stories about the effects of evaluation utilization on development programming and partner country policies in Agency reports on evaluation practice improvements.
- Introduce more detailed guidance on requirements for describing the utilization of evaluation results in USAID's annual PPR exercise.[2]
- Work with USAID DIV to more effectively document and disseminate the results of the randomized controlled trials and other impact evaluations of pilots funded through its grants, including posting randomized controlled trial reports to the DEC as special evaluations, if not external impact evaluations under USAID's Evaluation Policy definition.[3]

[2] Evaluation Utilization at USAID (2016) examined PPRs for several years and found that while some provided no information on evaluation use, most others presented only short and vague descriptions (e.g., "used to improve ongoing activity.")

[3] Notably, the results of rigorous of USAID-funded randomized controlled trials of DIV grant pilot projects are not published as USAID impact evaluations and posted on the DEC, or widely reported in Agency evaluation practice updates. This depresses the number of rigorous impact evaluations that USAID appears to have completed.

- Provide USAID staff with illustrative resource material to capture the effects of evaluation utilization as they follow up on post-evaluation action plans, per ADS 201.3.5.18.

Notes from the Field

BOX 43: NOTES FROM THE FIELD - EVALUATION UTILIZATION

"We have seen impact evaluations utilized, especially by governments. One of our biggest impact evaluations in Ghana – it was a nationwide program – and the local government got very involved and interested in how infrastructure projects can be completed efficiently. The government has really taken up this study's findings. Similarly, in Haiti, we did a pre-trial detention impact evaluation where we showed that it was costlier to keep people in detention before their trials and if government would just provide representation it would be cheaper. My favorite stories are where the host country uses these findings. We are also using them in USAID. For example, in Uganda, we had this impact evaluation that tested SMS and health, and we learned how to implement it better."

- USAID evaluation methodologist

To help the E3 Bureau track what happens to recommendations in evaluation reports, the E3 Analytics and Evaluation Project developed a protocol for contacting offices and missions approximately six months after delivery of the final report for an evaluation the Project completed. It seemed that the best way to learn about how evaluation findings and recommendations had been utilized was face-to-face or telephone interviews with individuals who had served as evaluation managers for specific studies or staff of the technical office involved in the activity that had been evaluated. For interviews conducted six months after an evaluation ends, the Project team determined that interviews should be preceded by an email that requests an interview and provides a few questions to guide the in-person or phone discussion. A summary of the evaluation to be discussed and its URL were included in the emails used to set up these post-evaluation interviews. Box 43 includes the relevant evaluation utilization questions from that protocol.

BOX 44: SAMPLE POST-EVALUATION EMAIL AND QUESTIONS ON EVALUATION UTILIZATION

Regarding preparation for the interview we have requested, it would be helpful if you could review the evaluation (especially its recommendations), and then think about the following:

- How were those recommendations useful or not? (Which recommendations were accepted by USAID for action?)
- What actions have been taken or are planned because of the evaluation?
- How have the evaluation findings and recommendations been used in programming or policy-level decisions by USAID or by USAID's implementing partners or country counterparts?
- Who else has seen, used, or commented on the evaluation to you, an what did they say they had done because of the evaluation's findings?
- We can go from there in our conversation. Some of these may be overlapping questions, but it's just to get you thinking. I will also be asking about whether the evaluation met your needs and if you can think anything that should be done differently in future evaluations to make their utilization more likely.