



USAID
FROM THE AMERICAN PEOPLE

EVALUATION TOOLKIT

Resources



Bureau for Policy, Planning and Learning

JULY 2015

EVALUATION TOOLKIT RESOURCES

TABLE OF CONTENTS

This document assembles the up-to-date and most recent PDFs that are included as resources in the Toolkit. They are included in the order that they first appear as a resource in the Toolkit. If a resource is referenced more than once you will find it in the PDF the first time it is referenced. The following summarizes the sections of the Toolkit.

Overview

1. Evaluation Policy at USAID

2. Evaluation Throughout the Program Cycle

- Evaluation in Country Development Cooperation Strategies
- Evaluation in Mission Performance Management Plans
- Evaluation in Project Monitoring and Evaluation Plans
- Evaluation in Activity Monitoring and Evaluation Plans
- Evaluation in the Budget Cycle
- Evaluation in Portfolio Reviews

3. Planning an Evaluation

- Deciding to Evaluate
- Engaging with Stakeholders
- Determining Evaluation Purpose and Evaluation Questions
- Developing an Evaluation Scope of Work
- Developing an Evaluation Independent Government Cost Estimate
- Commissioning and Evaluation

4. Managing an Evaluation

- From Award to Approval
- Conducting an Evaluation
- Managing and Monitoring an Evaluation Team
- From Draft to Final Report

5. Sharing, Reporting, Using, and Learning

- Sharing
- Reporting
- Utilization and Learning
- Assessing the Evaluation Process and Evaluator Performance

EVALUATION TOOLKIT RESOURCES

FREQUENTLY ASKED QUESTIONS

1. What is the Evaluation Toolkit?

The Evaluation Toolkit curates the available information on Evaluation at USAID. It contains the latest guidance, tools, and templates in an intuitive and easy-to-navigate format. Resources will be regularly updated (especially when new guidance is released). In addition, the Toolkit will go through various upgrades. Additional functionality and resources will be added as users provide feedback.

The primary audience for the Toolkit is USAID staff members, but, evaluators, M&E mechanism contractors, and other parties interested in knowing more about how USAID performs and uses evaluations will also find it helpful.

2. How can I navigate through the Toolkit?

Evaluation Toolkit is navigated through an intuitive user interface. A couple of hints for navigation:

- The Overview tab provides a guide to how to use the Evaluation Toolkit.
- At the top of the page, there are five sections or tabs with information grouped according to thematic area.
- Within each section or tab there is a brief narrative introducing the content of the section.
- Sub-thematic areas are listed on the left-hand side and go more in-depth into specific areas or processes.
- Resources such as Technical Notes, Templates, Checklists or other documents related to the area or process are provided on the right-hand side.
- At the bottom of each Toolkit page are Additional Links, where the user can go in-depth into supplementary references, including webinars, books and other guidance

When you place the cursor on a resource title, hover text will summarize the purpose of the resource. If you click on the resource it will open up the Learning Lab resource and provide additional information about the date the resource was published and the link to a PDF of the resource or if it is a template a link to the word document.

3. Do I need a USAID Learning Lab account to use the Evaluation Toolkit?

No, you don't need to create an account. The benefit of creating an account is you can upload resources to Learning Lab or comment.

4. How do I find specific resources included in the Evaluation Toolkit?

All of the resources included in the Evaluation Toolkit are a part of USAID's Learning Lab website. While navigating through the Toolkit, just click through to the appropriate section you are interested in and find the link to the resource listed on the right-hand side.

EVALUATION TOOLKIT

You can also search for resources in the “search box” on the top right of the website and select the link from among the results.

5. How do I know whether a specific resource is included as part of the Evaluation Toolkit?

There are many resources about evaluation on Learning Lab. To identify those that are the latest guidance and are USAID official, start with the Evaluation Toolkit to navigate to these resources.

6. What does “USAID Only” mean?

A few resources and additional links are available only to USAID staff. These are indicated by a designation of “USAID only”.

7. What does “USAID Official” mean?

Please visit the [FAQs](#) for USAID Learning Lab for a full summary of the tags, such as “USAID Official” included in the Toolkit. Most resources linked from the Toolkit are designated as USAID Official.

8. What does “Coming Soon” mean?

The Evaluation Toolkit is meant to provide the most up-to-date guidance, tools, and templates for initiating, planning, managing, and learning from evaluations. Some of these resources are in the process of being drafted or updated and are indicated as “Coming Soon”. Once they have been finalized, all of the resources cited will be available online.

9. What is the difference between the e-version and the pdf version of the Toolkit?

Both the e-version and the pdf version of the Evaluation Toolkit contain the same information. The e-version offers the advantage of being able to navigate to the resources and additional links directly from the page. It also links to WORD versions of resources that users are expected to edit, such as templates. In the pdf version, these materials are provided after the Toolkit information itself in an offline, printable pdf-only version.

10. How can I get the URL for a specific page in the Evaluation Toolkit?

The URL for the Evaluation Toolkit is usaidlearninglab.org/evaluation. If you are in a particular section it will contain this URL. If you would like to share a particular resource with a colleague, each resource page on Learning Lab has a discrete URL.

11. What do I do if I have comments or suggestions?

We welcome your input. If you have any comments or suggestions on the Evaluation Toolkit material, just click “Submit” under “Give Us Your Feedback” in the Overview section.

PROGRAM CYCLE RESOURCE

USAID EVALUATION AND MONITORING TERMS

Derived from Automated Directives System (ADS) Series 200

Evaluation Terms

I. The Basics	
Evaluation	The systematic collection and analysis of information about the characteristics and outcomes of programs and projects as a basis for judgments, to improve effectiveness, and/or to inform decisions about current and future programming. Evaluation in USAID has two primary purposes: accountability to stakeholders and learning to improve effectiveness.
Assessment	A forward-looking process that may be designed to examine country or sector context to inform project design, or an informal review of projects. It is distinct from evaluation.
II. Types of Evaluations at USAID	
i. Categorized by Questions and Method	
Impact Evaluation	Evaluations that measure the change in a development outcome that is attributable to a defined intervention. They are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change. Impact evaluations in which comparisons are made between beneficiaries that are randomly assigned to either a treatment or a control group provide the strongest evidence of a relationship between the intervention under study and the outcome measured.
Performance Evaluation	Evaluations that focus on descriptive and normative questions: what a particular activity, project, or program has achieved (either at an intermediate point in execution or at the conclusion of an implementation period); how it is being implemented; how it is perceived and valued; whether expected results are occurring; and other questions that are pertinent to program design, management, and operational decision-making. Performance evaluations often incorporate before/after comparisons but generally lack a rigorously defined counterfactual.
ii. Categorized by Policy Requirement	
Required Evaluation	<p>An evaluation for which completion fulfills either the “large project requirement” or the “pilot activity requirement” (ADS 203.3.1.3). Required evaluations must be external and managed, in most cases, by Program Office staff. Required evaluations include:</p> <ul style="list-style-type: none"> • Large Project Evaluation: Evaluation of a project that equals or exceeds in dollar value the mean (average) project size for each Development Objective (DO) for the USAID Mission/Office. Each USAID Mission/Office is required to conduct at least one evaluation of each large project it implements. Large Project Evaluations must be external evaluations to meet this evaluation requirement. For additional guidance, please see the Required and Non-Required Evaluations: Questions and Answers document available on Learning Lab. • Pilot Activity/Project Evaluation: Evaluation of an activity within a project involving untested hypotheses or demonstrating new approaches that are anticipated to be expanded in scale or scope through U.S. Government foreign assistance or other funding sources. Any activity designated as a pilot or proof of concept will fall under this requirement. Pilot Activity/Project Evaluations must be external impact evaluations to meet this requirement. If an impact evaluation is not feasible, the Pilot Activity/Project Evaluation must be an external performance evaluation.

PROGRAM CYCLE RESOURCE

Evaluation Terms

Non-required Evaluation	An evaluation whose completion does not fulfill either the “large project” or “pilot activity” requirement (ADS 203.3.1.3). Missions/Offices may conduct non-required evaluations for learning or management purposes at any point in implementation of activities, projects, or programs. As evaluations, they still must meet all procedural, reporting, and quality standards stipulated in the ADS 203 and the Evaluation Policy. Non-required evaluations may be impact or performance, internal or external.
iii. Categorized by Relationship to USAID	
External Evaluation	Evaluations that are both: <ol style="list-style-type: none"> 1. Commissioned by USAID or others, rather than by the implementing partner responsible for the activities being evaluated, and 2. Have a team leader who is an independent expert from outside the Agency with no fiduciary relationship with the implementing partner. External evaluations may include USAID staff members, but none may serve as team leaders. An evaluation with a team leader from USAID/Washington is not an external evaluation. An evaluation contracted through a subcontract of the implementing partner is not an external evaluation.
Internal Evaluation	Evaluations that are either: <ol style="list-style-type: none"> 1. Conducted or commissioned by an implementing partner concerning their own project (an implementer internal evaluation) or 2. Commissioned by USAID and led by an individual from inside the Agency (a USAID internal evaluation).
iv. Categorized by Timing	
Mid-term Evaluation¹	Evaluations that occur roughly halfway through an activity, project, or program.
Final Evaluation²	Evaluations that occur toward the end of an activity, project, or program.
III. Evaluation Planning	
Mission-wide Evaluation Plan	Identifies and tracks evaluations across the Mission and over the entire Country Development Cooperation Strategy (CDCS) timeframe. Evaluation plans should include (at minimum) the activity/project/program to be evaluated, evaluation type, possible evaluation questions, estimated budget, planned start date, and estimated completion date. It is a required component of the Mission Performance Management Plan (PMP).
Project Evaluation Plan	The evaluation portion of a Project Monitoring and Evaluation (M&E) Plan, developed during project design and included as an annex to the Project Appraisal Document (PAD). It should include a description of what type of evaluation, if any, is required for the project; a limited number of key evaluation questions; a timeline for implementing the evaluation; and an estimated budget that will be set aside from the project budget and used for the evaluation.
Activity Evaluation Plan	The evaluation portion of an Activity M&E Plan, submitted by the implementer. It should include relevant information about planned external evaluations by USAID or internal evaluations to be conducted by the implementer.
Evaluation Registry	An annex to the annual Performance Plan and Report (PPR) to be submitted by all USAID Missions/Offices and Washington Operating Units. It is an inventory of evaluations conducted during the previous year and planned evaluations and estimated budgets for the coming fiscal year, plus two out years.

¹ “Mid-term evaluation,” while a useful term, is not a policy-relevant term at USAID. USAID evaluation policy does not distinguish between mid-term and final evaluations, and there are no requirements for one that are different from the other. Mid-term evaluations are sometimes called “formative evaluations,” although their meanings are somewhat different. Like “mid-term evaluation,” the term “formative evaluation” does not have any real significance in terms of USAID policy.

² “Final evaluation,” while a useful term, is not a policy-relevant term at USAID. USAID evaluation policy does not distinguish between mid-term and final evaluations, and there are no requirements for one that are different from the other. Final evaluations are sometimes called “summative evaluations,” although, again, their meanings are somewhat different. Like “final evaluation,” the term “summative evaluation” does not have any real significance in terms of USAID policy.

PROGRAM CYCLE RESOURCE

Performance Management and Monitoring Terms

I. The Basics	
Performance Management	The systematic process of planning, collecting, analyzing, and using performance monitoring data and evaluations to track progress, influence decision-making, and improve results. Performance management activities are described at the mission level in the Mission's PMP. Performance management is one aspect of the larger process of continuous learning and adaptive management.
Performance Monitoring	The ongoing and routine collection of performance indicator data to reveal whether desired results are being achieved and whether implementation is on track. Performance monitoring continues throughout the life of an activity, a project, and a Mission's CDCS.
Other Monitoring	
Activity Oversight	The day-to-day assessment of contractor and grantee performance by a Contracting Officer's Representative/Agreement Officer's Representative (COR/AOR) or others through site visits, stakeholder meetings, and the verification of implementer inputs, outputs, and deliverables.
Contextual Monitoring	The monitoring and measurement of conditions relevant to the performance of activities, projects, and programs, such as macro-economic, social, or political conditions.
II. Plans	
Performance Management Plan	A tool to plan and manage the process of monitoring, evaluating, and analyzing progress toward achieving results identified in a CDCS and project logical framework (LogFrame) in order to inform decision-making, resource allocation, learning, and adapting projects and programs. PMPs are mission-wide documents and are distinct from Project M&E Plans and Activity M&E Plans.
Project M&E Plan	A plan for measuring progress toward planned results and identifies the cause of any delays or impediments during implementation. Missions must develop a Project M&E Plan during project design and include it as an annex to the PAD. The Project M&E Plan provides a framework for M&E that pulls together performance information from all activities contributing to a project. Project M&E plans <u>should not</u> be referred to as PMPs, which are mission-wide documents.
Activity M&E Plan	A plan for monitoring and evaluating USAID activities at the activity/implementing mechanism level. Implementers are expected to submit an Activity M&E Plan to USAID CORs/AORs within the first 90 days of an award and before major activity implementation actions begin. Activity M&E Plans submitted to USAID should include only those indicators that the Mission needs for activity management, rather than the entire set of all indicators an implementer uses for its management purposes. Activity M&E Plans should not be referred to as PMPs, which are Mission-wide documents.
Performance Plan and Report	Annual report that documents U.S. Government foreign assistance results achieved over the past fiscal year and sets targets on designated performance indicators for the next two fiscal years.

PROGRAM CYCLE RESOURCE

Types of Monitoring Indicators

I. Categorized by What They Measure	
Performance Indicator	Measures a particular characteristic or dimension of or activity-, project-, or strategy-level results based on a Mission's CDCS Results Framework or a project's LogFrame.
Context Indicator	Measures conditions relevant to the performance of activities, projects, and programs, such as macro-economic, social, or political conditions; critical assumptions of a CDCS; and the assumptions column of project LogFrames. Context indicators do not directly measure the results of USAID activities.
II. Categorized by Method of Data Collection	
Quantitative Indicator	Indicator based on mathematical quantities, typically taking the form of a count value, a mean or median, or a percentage or ratio.
Qualitative Indicator	Indicator based on subjective criteria. Qualitative indicator data are quantified to more effectively measure the result and mitigate subjectivity. Approaches include: <ol style="list-style-type: none"> 1. Rating Scale Indicator: A measurement device that quantifies a range of subjective responses on a single issue or single dimension of an issue. 2. Milestone Indicator: A type of indicator that measures progress toward a desired outcome by dividing the progress into a series of defined steps. The simplest form of a milestone indicator is a binary indicator of whether a particular discrete result has or has not been achieved.
III. Categorized by Complexity	
Simple Indicator	Measure of a single quantity or single dimension of a result, typically phenomena with clear boundaries.
Composite Indicator (Index)	Combines two or more indicators into a single measure.
IV. Categorized by U.S. Foreign Assistance	
Standard U.S. Foreign Assistance Indicator	An indicator from the State Department Office of Foreign Assistance Standard List of Foreign Assistance Indicators for use in the annual PPR.
Custom Indicator	Any non-standard foreign assistance indicator created by an Operating Unit to adequately capture its key results or achievements in the PPR.

PROGRAM CYCLE RESOURCE

Performance Monitoring Indicator Terms

Performance Baseline	The value of a performance indicator at the onset of implementation of USAID-supported activities, projects, or strategies that contribute to the achievement of the relevant result.
Performance Target	The specific, planned level of result to be achieved within an explicit timeframe with a given level of resources.
Performance Actual	The actual value of a performance indicator achieved within an explicit timeframe.
Indicator Disaggregate	A breakdown of the performance indicator data into component groups (e.g. sex, age, or other category) for more detailed analysis.
Performance Indicator Reference Sheet (PIRS)	<p>A document capturing all reference information about a performance indicator. At minimum, reference data for each indicator in a PIRS include:</p> <ul style="list-style-type: none"> • The definition of the indicator and unit of measure; • Its link to the Results Framework and LogFrame; • Whether and how the data must be disaggregated (by sex, age, or other category); • Data source and method of data collection, construction, and/or analysis; • Reporting frequency; • Known data quality limitations relative to the five standards of data quality; • Date of last Data Quality Assessment (DQA) and name of the DQA reviewer; • Responsible office and individual for collection and analysis; and • Any changes to the indicator reference data over time.
Performance Indicator Tracking Table (PITT)	A spreadsheet, database, or other information technology solution that serves as a repository of indicator data and, preferably, enables analysis of performance indicator data collected for PMPs and Project M&E Plans. The data tables must be updated, at minimum, on an annual basis.

Data Quality Terms

Data Quality Standards	<p>Standard criteria for determining the quality of performance monitoring data for evidence-based decision-making and credible reporting. The five standards of data quality are:</p> <ol style="list-style-type: none"> 1. Validity: Data should clearly and adequately represent the intended result. 2. Integrity: Data collected should have safeguards to minimize the risk of transcription error or data manipulation. 3. Precision: Data should have a sufficient level of detail to permit management decision-making; e.g., the margin of error is less than the anticipated change. 4. Reliability: Data should reflect stable and consistent data collection processes and analysis methods over time. 5. Timeliness: Data should be available at a useful frequency, should be current, and should be timely enough to influence management decision-making.
Data Quality Assessment	An examination of the quality of performance indicator data in light of the five standards of data quality to ensure that decision-makers are fully aware of data strengths and weaknesses and the extent to which data can be relied upon when making management decisions and reporting.
Evidence	Factual basis for programmatic and strategic decision-making in the program cycle. Evidence can be derived from assessments, analyses, performance monitoring, and evaluations. It can be sourced from within USAID or externally and should result from systematic and analytic methodologies or from observations that are shared and analyzed.

PROGRAM CYCLE RESOURCE

Relevant Strategy and Planning Terms

I. Types of Logical Models	
Results Framework	A graphical representation (typically supplemented by narrative) of the development hypothesis that includes the CDCS Goal, DOs, intermediate results (IRs), sub-IRs, and performance indicators.
Project Log Frame	A rigorous methodology used for project design that focuses on the causal linkages between project inputs, outputs, sub-purpose, purpose, and project goal.
II. Programming Hierarchy	
Program	A set of projects typically at the DO level. "Program" can also have more generic meanings, such as "program funds" or host-country government "program."
Project	A set of executed interventions, over an established timeline and budget, intended to achieve a discrete development result (i.e., the project purpose) through resolving an associated problem. It is explicitly linked to the CDCS Results Framework.
Activity	A sub-component of a project that contributes to a project purpose. It typically refers to an award (such as a contract or cooperative agreement) or a component of a project (such as policy dialogue) that may be undertaken directly by Mission staff.
III. Components of Logical Models	
Input	A resource, such as technical assistance, commodities, training, or provision of USAID staff, that is used to create an output. It appears at the lowest level of a project LogFrame.
Output	A direct result of inputs. They are the tangible, immediate, and intended products or consequences of an activity within USAID's control or influence.
Outcome	The conditions of people, systems, or institutions that indicate progress or lack of progress toward achievement of project/program goals. Outcomes are any result higher than an output to which a given output contributes but for which it is not solely responsible. Outcomes may be intermediate or end outcomes, short-term or long-term, intended or unintended, positive or negative, or direct or indirect.
Result	A significant, intended, measurable change in a development condition affecting people, systems, or institutions. Results are outputs and outcomes, including Goals, DOs, IRs, sub-IRs, Project Purpose, and Project Outputs as specified in a Mission's CDCS or project LogFrame.



TECHNICAL NOTE

Impact Evaluations

Monitoring and Evaluation Series

This Note defines impact evaluations and discusses design and key planning considerations.

Technical Notes

are published by the Bureau for Policy, Planning and Learning and provide key concepts and approaches to USAID staff and partners related to the Program Cycle. This Technical Note supplements USAID ADS Chapter 203 and replaces TIPS 18, Rigorous Evaluations.

INTRODUCTION

This Note defines impact evaluations, explains when they should be commissioned according to USAID policy and describes different designs for quasi-experimental and experimental impact evaluations. The USAID Automated Directives System (ADS) 203 defines **impact evaluations** as *those that measure the change in a development outcome that is attributable to a defined intervention. Impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change.*

Decisions about whether an impact evaluation would be appropriate, and what type of impact evaluation to conduct, are best made early during the project design phase. Some impact evaluation designs can only be implemented if comparison groups are established and baseline data is collected before an intervention begins. Although they are most effective and sometimes only possible when planned before program implementation, impact evaluations can sometimes be used to measure changes that occur either during or after program implementation. In most cases, an expert should be consulted in advance to determine whether an impact evaluation will be feasible.

This note outlines key considerations that USAID staff and evaluators should take into account when planning for and designing impact evaluations. Those commissioning an evaluation should include the evaluator when making decisions about an intervention's targeting and implementation, and consider issues related to logistics, time and cost. Therefore although impact evaluations are a powerful tool to answer key questions about a particular intervention, they should be used selectively and only when appropriate in terms of purpose and funding.

WHAT IS AN IMPACT EVALUATION?

Impact evaluations are useful for determining the effect of USAID activities on specific outcomes of interest. They test USAID development hypotheses by comparing changes in one or more specific outcomes to what would have happened in the absence of the intervention, called the *counterfactual*. Impact evaluations use a *comparison group*, composed of individuals or communities where an intervention will not be implemented, and one or more *treatment groups*, composed of project beneficiaries or communities where an intervention is implemented. The comparison between the outcomes of interest in the treatment and comparison group creates the basis for determining the impact of the USAID intervention. An impact evaluation helps demonstrate *attribution* to the specific intervention by showing what would have occurred in its absence.

Most interventions track changes in key outcomes through performance monitoring, but comparing data from performance indicators against baseline values demonstrates only whether change has occurred, but does not establish what actually caused the observed change. *Confounding factors* include interventions run by other donors, natural events (e.g. rainfall, drought, earthquake, etc.), government policy changes, or natural changes that happen in an individual or community over time. Due to the potential effects of confounding factors, USAID managers cannot claim that their interventions actually caused the observed changes or results. In some cases, the intervention does cause all observed change. In these cases, the group receiving USAID assistance will have improved significantly while a similar, non-participating group will have stayed roughly the same. In other situations, the target group may have already been improving, and the intervention helped to accelerate that positive change. Or, intended outcomes may appear to be negative (for instance, during an economic downturn), but comparison groups fare even worse. Impact evaluations are designed to identify the effects of the intervention of

interest in all of these cases, where both the target group and non-participating groups may have changed, but at different rates. By identifying the effects caused by an intervention, impact evaluations help USAID, implementing partners, and key stakeholders learn which approaches are most effective. This is critical for determining future development programming and resource allocation.

Note that the term "impact evaluation" involves a specialized meaning of the word "impact." In common usage, "impact" could refer to high level results or long-term outcomes from an intervention. However, "impact evaluation" implies a structured test of one or more hypotheses underlying an intervention. Impact evaluations are characterized by a specific evaluation design (quasi-experimental or experimental) in order to answer a cause-and-effect question. These methods can be used to attribute change at any program or project outcome level, but typically focus on one specific activity. Impact evaluations typically collect and analyze quantitative data, but should also be informed by qualitative data collection methods as long as they are used to gather information from both treatment and comparison groups.

QUESTIONS FROM USAID-FUNDED IMPACT EVALUATIONS

- What is the added value of the use of sports in workforce development programs for at-risk youth in Honduras and Guatemala? To what extent are program effects stronger or weaker for female, higher risk, younger, or less educated participants?
- To what extent were neighbors of beneficiaries positively or negatively affected by a livelihoods program in Ethiopia?
- Does training traditional leaders on human rights and peaceful conflict mitigation result in improvements in community leadership and dispute resolution? To what extent do top-down, horizontal, or bottom-up social pressures change the behavior of local leaders?

WHEN SHOULD IMPACT EVALUATIONS BE USED?

Impact evaluations *answer cause-and-effect questions* about intervention effects. While impact evaluations are sometimes used to examine the effects of only one intervention or project approach, they are also extremely useful for answering questions about the effectiveness of alternative approaches for achieving a given result, e.g., which of several approaches for improving farm productivity, or for delivering legal services, are most effective. Missions should consider using impact evaluations strategically to answer specific questions about the effectiveness of key approaches. *External validity* - the extent to which evaluation results can be generalized to other settings, such as when an intervention is scaled up or attempted in other regions - is an important consideration for impact evaluations. Ways to ensure external validity include carrying out multiple impact evaluations across Missions on a similar topic or approach and making sure that the evaluation measures the effects of an intervention on different types of beneficiaries (across gender, age, socioeconomic groups, or other relevant factors). It is important for Missions to consult sector experts and coordinate with their Regional and Pillar Bureaus to ensure that they are contributing to a Bureau-wide learning and evaluation strategy.

Impact evaluations require strong performance monitoring systems to be built around a clear logical framework. The development hypothesis should clearly define the logic of the project, with particular emphasis on the intervention (independent variable) and the principle anticipated results (dependent variables), and provides the basis for the questions that will be addressed by the impact evaluation.

Impact evaluations are always most effective when planned before implementation begins. Evaluators need time prior to implementation to identify appropriate indicators, identify a comparison group, and set baseline values. In most cases they must coordinate the selection of a treatment and comparison group with the implementing partners. If impact evaluations are not planned prior to implementation the number of potential evaluation design options is reduced, often leaving alternatives that are either more complicated or less rigorous. As a result, Missions should consider the feasibility of and need for an impact evaluation prior to and during project design. On the other hand, interventions should not be evaluated too early in their “start-up phase,” when the implementation details of the intervention are still being worked out. A good way to account for startup issues is to conduct a small pilot in a few communities (not included in the evaluation) before working with and conducting an evaluation of the full sample.

WHEN TO CONDUCT IEs

ADS 203 states that “any activity within a project involving *untested hypotheses* or demonstrating *new approaches that are anticipated to be expanded* in scale or scope through US Government foreign assistance or other funding sources will, if feasible, undergo an impact evaluation... Any activity or project designated as a ‘pilot’ or ‘proof of concept’ will fall under this requirement.”

The World Bank has published the following guidelines for when an impact evaluation is appropriate:

- Is the intervention *INNOVATIVE*? Is it testing a new, promising approach?
- Is the intervention *REPLICABLE*? Can it be scaled up or can it be applied to a different setting?
- Is the intervention *STRATEGICALLY RELEVANT*? Is it a flagship intervention that requires substantial resources; covers, or could be expanded to cover, a large number of people; or could generate substantial savings?
- Is the intervention *UNTESTED*? That is, is very little known about the effectiveness of the intervention globally or in the specific context in which it is implemented?
- Is the intervention *INFLUENTIAL*? Will results be used to inform key policy decisions?

(*Impact Evaluation in Practice*, p. 11)

While impact evaluations do require advanced planning and significant attention to detail, they need not be impossibly complex, particularly since many of the most common questions and challenges can be anticipated and minimized with advanced planning. In all cases, USAID staff must coordinate between the evaluator, the implementer, and other stakeholders to identify an appropriate comparison or control group.

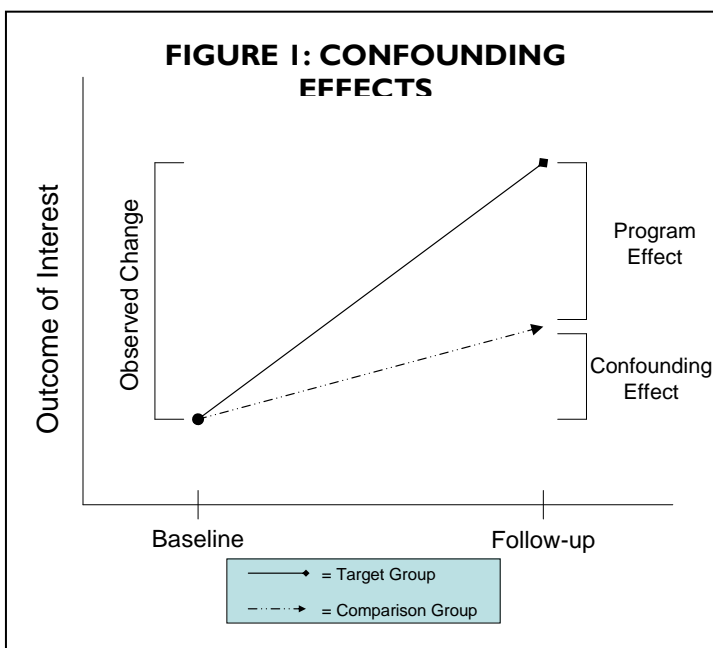
Finally, impact evaluations are not appropriate for all situations. They often involve extra costs for data collection and always require high levels of attention to detail, coordination, and time during intervention implementation. The potential extra costs should be considered against the information needs when determining whether an impact evaluation is appropriate. Performance evaluation may be more appropriate for answering other types of evaluation questions. For example, a USAID manager may be more interested in describing a process or analyzing ‘why’ and ‘how’ observed changes, particularly unintended changes, were produced. Questions generated in these cases may be more effectively answered using other evaluation methods, including participatory evaluations or rapid appraisals. Similarly, there are situations when impact evaluations, which use comparison or control groups, will not be advisable or even possible. For example, assistance focusing on political parties can be difficult to evaluate using impact evaluations, as this type of assistance is typically offered to all parties, making the identification of a comparison group difficult or impossible. Other methods may be more appropriate and yield conclusions with sufficient credibility for programmatic decision-making. Finally, when an intervention is offered in different ways across different sites (for example if communities select from a “package” of interventions) or changes significantly over time (for instance, when implementation details change significantly during the “start-up” phase of an activity), information from an impact evaluation will be less likely to apply to other settings or be useful in decisions about scale up.

DESIGN

This section outlines types of IE designs to increase understanding of what these approaches entail. Agency staff are encouraged to seek outside assistance from experts with evaluation methods training.

Although there are many variations, impact evaluations are divided into two categories: quasi-experimental and experimental. Both categories of impact evaluations rely on the same basic concept - using the counterfactual to estimate the changes caused by the intervention. The counterfactual answers the question, “What would have happened to intervention participants if they had not participated in the intervention?” The comparison of the counterfactual to the observed change in the group receiving USAID assistance is the true measurement of an intervention’s effects.

Impact evaluations compare outcomes for groups that do and do not receive the intervention to answer questions about the counterfactual situation. While ‘before-after’ measurements of a single group using a baseline allow the measurement of a single group both with and without participation, this design does not control for all the other confounding factors that might influence the participating group during implementation. When well-constructed, comparison groups provide a clear picture of the effects of interventions on the target group by differentiating project effects from the effects of multiple other factors in



the environment which affect both the target and comparison groups. This means that in situations where economic or other factors that affect both groups are making everyone better off, it will still be possible to see the additional or incremental improvement caused by the intervention, as Figure 1 illustrates.

When a comparison group is generated using a random process, the evaluation is considered an *experimental evaluation* and the comparison group is referred to as a *control group*. When a comparison group is generated using other, non-random methods, the evaluation is considered a *quasi-experimental evaluation*.

QUASI-EXPERIMENTAL EVALUATIONS

To estimate intervention effects, quasi-experimental designs estimate the counterfactual by conducting measurements of a non-randomly selected comparison group. In many cases, intervention participants are selected based on certain characteristics, whether it is level of need, location, social or political factors, or some other factor. While evaluators can often identify and match many of these variables (or account for them in a regression analysis), it is impossible to match all factors that might create differences between the treatment and comparison groups, particularly characteristics which are more difficult to measure or are unobservable, such as motivation or social cohesion. For example, if an intervention is targeted at communities which are likely to succeed, then the target group might be expected to improve relative to a comparison group that was not chosen based on the same factors. On the other hand, if an intervention is targeted at the “neediest” potential beneficiaries, then the changes that the intervention expects to achieve may occur at a slower rate than with other, better-off individuals. Failing to account for this in the selection of the comparison group would lead to a biased estimate of intervention impact. *Selection bias* is the difference between the comparison group and the treatment group caused by the inherent differences between the two groups, and the uncertainty or error this generates in the measurement of intervention effects. All quasi-experimental evaluation designs must account for the extent to which they have minimized or measured selection bias.

Common quasi-experimental designs include:

- **Non-Equivalent Group Design.** In this design, a comparison group is hand-picked to match the treatment group as closely as possible. Since hand-picking the comparison group cannot completely match all characteristics with the treatment group, the groups are considered to be ‘non-equivalent’.
- **Matching:** The most common means for selecting a comparison group is matching, wherein the evaluator picks a group of similar units based on observable characteristics that are thought to influence the outcome. For example, the evaluation of an agriculture intervention aimed at increasing crop yield might seek to compare participating communities against other communities with similar weather patterns, soil types, and traditional crops, as communities sharing these critical characteristics would be most likely to behave similarly to the treatment group in the absence of the intervention. A type of matching design occurs when a comparison group is selected based on shared observable characteristics with the treatment group. However, rather than choosing matches based on a small number of variables, *propensity score matching* uses a statistical process to combine information from all data collected on the target population to create the most accurate matches possible based on observable characteristics. Neither type of matching can account for unobservable characteristics such as motivation.
- **Regression Discontinuity.** Interventions often have eligibility criteria based on a cut-off score or value of a targeting variable. Examples include interventions accepting only households with income below \$2,000 USD, organizations or individuals or organizations just above and just below the cut-off value would demonstrate only marginal or incremental differences in the absence of USAID assistance, as families earning \$2,001 USD compared to \$1,999 USD are unlikely to be significantly different except in terms of

eligibility for the intervention. Because of this, the group just above the cut-off serves as a comparison group for those just below (or vice versa) in a regression discontinuity design.

In all of the above cases, the evaluation team should compare the treatment and comparison groups at **baseline** to make sure that the groups are in fact comparable. If there are significant differences at baseline in variables that may influence the outcome (for instance, the treatment group consists of wealthier communities) then the evaluation's ability to attribute later differences between the treatment and comparison groups to the intervention being evaluated will be less credible. If the evaluation is commissioned after the intervention begins, but baseline data is available, it is possible to conduct a *retrospective quasi-experimental design*.

EXAMPLE OF A QUASI-EXPERIMENTAL EVALUATION

USAID commissioned an impact evaluation of the Colombia Strategic Development Initiative, which provides U.S. assistance to the Colombian government's program to expand state presence in vulnerable areas and "consolidate" the rule of law. There are two separate mechanisms for this evaluation: A consortium of academics based at Princeton University and funded by the Department of Defense collaborated with USAID/Colombia's M&E program and gave technical advice to the firm that was contracted to conduct the evaluation.

The evaluation team used propensity score matching to identify municipalities that were similar to those selected for the program. They estimated propensity to receive treatment based on the historical presence of armed groups, market integration (or lack of), trends in contestation, presence of illicit crops, and population importance. They also measured trends in key outcome variables (from 2002 to 2010) to ensure that the treatment and comparison communities were in fact comparable. Data collection was conducted at the household (19,000), community, project, and municipal levels. The evaluation team had developed survey questions that had never been used before, in particular those that addressed sensitive issues such as participation or contact with armed guerrilla groups, so they piloted the questionnaire in one municipality before applying it to the entire evaluation sample.

The impact evaluation will allow both the Government of Colombia and USAID to learn which programs work where, and why. Given the substantive importance of the issue, as well as the resources invested in the programs by the USG globally, this is crucial. The fieldwork for the baseline was finished on June 2013 and two more waves of data collection are expected.

- **Interrupted Time Series.** In some situations, a comparison group is not possible, often because the intervention affects everyone at once, as is typically the case with policy change. In these cases, data on the outcome of interest is recorded at numerous intervals before and after the intervention takes place. The data form a time-series or trend, which the evaluator analyzes for significant changes around the time of the intervention. Large spikes or drops immediately after the intervention signal changes caused by the intervention. This method can be strengthened by the use of a comparison group to rule out potentially confounding factors, reducing uncertainty in evaluation conclusions. Interrupted time series are most effective when data is collected regularly both before and after the intervention, leading to a long time series, and when the analysis can account for alternative causes.

EXPERIMENTAL EVALUATION

In an experimental evaluation, the treatment and comparison groups are selected from the target population by a random process. Because the selection of treatment and control groups involves a random process, experimental evaluations are often called randomized evaluations or randomized controlled trials (RCTs).

Random selection from a target population into treatment and control groups is the most effective tool for eliminating selection bias because it removes the possibility of any individual characteristic influencing selection. Because units are not assigned to treatment or control groups based on specific characteristics, but rather are divided randomly, all characteristics that might lead to selection bias, such as motivation, poverty level, or proximity, will be roughly equally divided between the treatment and control groups. If an evaluator uses random assignment to determine treatment and control groups, she might, by chance, get 2 or 3 very motivated communities in a row assigned to the treatment group, but if the intervention is working in more than a handful of communities, the number of motivated communities will likely balance out between treatment and control groups.

Because random selection completely eliminates selection bias, experimental evaluations are often easier to analyze and provide more credible evidence than quasi-experimental designs. Random assignment can be done with any type of unit, whether the unit is the individual, groups of individuals (e.g. communities or districts), organizations, or facilities (e.g. health center or school) and usually follows one of the following designs:

WHAT UNIT TO RANDOMIZE?

A good rule of thumb is to randomize at the level in which the intervention takes place. For example, in an evaluation of a teacher training intervention, it would be impractical to ask the teacher to apply her new skills with some students and not others, and even if she could, selected students could influence their classmates anyway (see “spillover” below). Furthermore, it might not be politically or logistically feasible to train some teachers and not others within an individual school. It is more realistic to assign some schools to the treatment group and others to the control group, as long as the sample of schools is sufficiently large to detect statistically significant results. This type of decision is usually made in consultation with the evaluator, the implementing partner, and relevant USAID staff.

- Simple Random Assignment.** When the number of intervention participants has been decided and additional eligible individuals are identified, simple random assignment through a coin flip or lottery can be used to select the treatment group and control groups. Interventions often encounter or can generate ‘excess demand’ naturally, for example in training interventions, participation in study tours, or where resources limit the number of partner organizations, and simple random assignment can be an easy and fair way to determine participation while maximizing the potential for credible evaluation conclusions. For example, in a recently released USAID-funded impact evaluation conducted by the National Democratic Institute of a governance project in Cambodia, each field officer had to choose two communities that he or she felt that the project should work with. The evaluation team then flipped a coin for each pair, generating one treatment and one control community for each officer.

- Phased-In Selection.** Even if an intervention plans to treat all eligible beneficiaries, there may be logistical reasons that prevent implementation from beginning everywhere at the same time. This type of schedule creates a natural opportunity for using an experimental design. Consider an intervention where delivery of a conditional cash transfer was scheduled to operate in 100 communities during year one, another 100 the second year and a final 100 in the intervention’s third year. The year of participation can be randomly assigned. Communities selected to participate in Year 1 would be designated as the first treatment group (T1). For that year all the other communities, which would participate in years two and three, form the

initial control group. In the second year, the next 100 communities would become the second treatment group (T2), while the final 100 communities would continue to serve as the control group until the third year. This design is also known as *pipeline* or *stepped wedge design*.

- **Randomized Promotion (Encouragement Design).** In cases where randomized assignment is difficult, evaluators can randomize promotion of a particular intervention. For instance, a microfinance institution might be unwilling to turn potential clients away just because they are assigned to a control group, preferring to serve anyone who seeks to open a savings account. Evaluations of savings interventions instead randomly select some people within a community to receive a special invitation or incentive to open an account. If there is a substantial difference in uptake between those who receive an invitation to join and those who do not, then evaluators can compare the “invitation” and “no invitation” groups using an instrumental variable analysis (see above).

EXAMPLE OF A USAID EXPERIMENTAL EVALUATION

The evaluation of the A Ganar program is currently examining the impact of a sports-based youth workforce development program in Guatemala and Honduras on outcomes such as employment, school enrollment, and prevalence of risky behavior. The evaluation team piloted the study logistics and data collection with a small group (174 survey respondents in Honduras and Guatemala), which allowed them to: 1) refine the baseline survey and interview protocols; 2) determine the randomization strategy; and 3) work out the division of responsibilities with implementers. Local partners were willing to recruit a larger number of potential beneficiaries and allow the evaluation team to randomly allocate spots, but were not willing to exclude youth who had worked to recruit their peers into the program. Therefore they were allowed to select up to three youth to participate in the program – that is, not subject to randomized selection, and therefore included in the intervention but not in the evaluation sample. The rest were randomly assigned to treatment and control groups, and a subset of these was invited to participate in the baseline.

The full roll-out of the program will have a sample size of 1300 youth in Honduras, divided into one treatment and one control group (to test the effect of the program overall). In Guatemala, 1500 applicants will be divided into two treatment groups, one receiving a sports-based program and one receiving an equivalent program, and one control group. The design in Guatemala will allow the evaluation team to isolate the effect of the use of sports. Aside from the baseline survey, they will conduct two additional data collection events (immediate and long-term follow up) as well as qualitative longitudinal case studies to supplement quantitative findings.

- **Blocked (or Stratified) Assignment.** When it is known in advance that the units to which an intervention could be delivered differ in one or more ways that might influence the outcome, e.g., age, size of the community in which they are located, ethnicity, etc., evaluators may wish to take extra steps to ensure that such conditions are evenly distributed between an evaluation’s treatment and control groups. In a simple block (stratified) design, an evaluation might separate men and women, and then use randomized assignment within each block to construct the evaluation’s treatment and control groups, thus ensuring a specified number or percentage of men and women in each group.
- **Multiple Treatments.** It is often the case that multiple approaches will be proposed or implemented for the achievement of a given result. If an evaluation is interested in testing the relative effectiveness of 3 different strategies or approaches, eligible units can be randomly divided into 3 groups. An HIV prevention service, for example, could provide just prevention education to one group and prevention

education and peer support to another. Each group participates in one approach, and the results can be compared to determine which approach was most effective. Variations on this design can include additional groups to test combined or holistic approaches and a control group to test the overall effectiveness of each approach. The multiple treatment groups can be generated using any of the methods outlined above.

ANALYSIS

In an impact evaluation, quantitative analysis can be as simple as comparing outcome means between treatment and comparison or control groups. When baseline measures exist, evaluators typically measure changes between baseline outcome measures and final outcome measures and compare these changes between treatment and control or comparison groups. This method allows them to take into account differences between the two groups that are constant over time and is known as a **difference-in-difference** analysis. Other analysis tools, such as **multivariate regressions**, or *analyses of covariance (ANCOVA)*, are more complex. Agency SOWs should require that evaluators report the results of analyses conducted using various tools and to use results from qualitative data collection to deepen explanations of findings.

KEY CONSIDERATIONS

EFFECT SIZE

In planning for an impact evaluation, it is important to clarify how large or small an *effect size* – that is, the magnitude of difference between the treatment and control group - the evaluator will be expected to measure. In theory, with unlimited evaluation funding and sample sizes, an impact evaluation could find that participants in the treatment group had a 0.001% higher income, but from a practical perspective, it is not worth determining whether an intervention has such a tiny impact. Considerations of effect size usually take into account what other interventions have accomplished given a certain level of funding and what has typically been achieved in a particular sector.

COST

Impact evaluations will almost always cost more than performance evaluations that do not require comparison groups. However, the additional cost can sometimes be quite low depending on the type and availability of data to be collected. Moreover, findings from impact evaluations may lead to future cost-savings, outweighing initial costs, through improved programming and more efficient use of resources. Nevertheless, USAID managers must anticipate these additional costs, including the additional staff resources implied by the level of attention to detail required, when considering and budgeting an impact evaluation. The largest cost of an impact evaluation is usually data collection, which in turn depends on the sample size (see below). PPL will provide additional guidance on budgeting for impact evaluations.

ETHICS

The use of comparison groups is often criticized for denying services to potential beneficiaries. This is less of a concern if the intervention has not been tested before, as there is also an ethical argument for demonstrating that an intervention does not have negative effects before implementing it at a widespread level. In addition, interventions can often take advantage of existing operational restrictions. For instance, most interventions have finite resources and must select a limited number of participants or geographic areas among those who would be eligible. In other cases, there is enough funding to work in an entire country but the implementer may not have the capacity to operate in all areas at once, which presents an opportunity to use a phased-in design. Random selection of participants or communities is often viewed, even by those beneficiaries who are not selected, as being the fairest and most transparent method for determining participation.

A second ethical question emerges when an intervention seeks to target participants that are thought to be

most in need. In some cases, impact evaluations require a relaxing of targeting requirements in order to identify enough similar units to constitute a comparison group, meaning that perhaps some of those identified as the ‘neediest’ might be assigned to the comparison group. However, it is often the case that the criteria used to target are not definitively known and rarely with the degree of precision required to confidently rank-order potential participants. Alternatively, situations where the cutoff point for participation is such that those just below and just above are very similar to each other present an appropriate opportunity to use a regression discontinuity design.

Some countries require in-country ethical clearance for research. See *Protection of Human Subjects in Research Supported by USAID - A Mandatory Reference for ADS Chapter 200* for more information on protection of human subjects required by USAID. In cases where an evaluation firm hires an academic, they are usually required to secure clearance from their university’s Human Subjects review board and provide evidence of having completed a human subjects training course.

SAMPLE SIZE

During the analysis phase, impact evaluations use statistical tests to determine whether any observed differences between treatment and comparison groups represent actual (statistically significant) differences or whether the difference could have occurred due to chance alone. The ability to make this distinction depends principally on the size of the change and the total number of units in the treatment and comparison groups, or sample size. That is, there is always a chance that a group of communities that is randomly allocated to the treatment group may be more or less motivated (or more urban, or have another characteristic that influences the outcome) than the control group. With larger samples, this likelihood is reduced. The more units, or higher the sample size, the easier it is to attribute change to the intervention rather than to random variations. During the design phase, impact evaluations calculate the number of units (or sample size) required to confidently identify changes of the size anticipated by the intervention. An adequate sample size helps prevent declaring a successful intervention ineffectual (false negative) or declaring an ineffectual intervention successful (false positive). Sample size calculations should be done before each evaluation, in consultation with an expert, and take into account expected effect size and existing variability in a population. As a rule of thumb, impact evaluations are rarely undertaken with less than 100 (total) units of analysis.

SPILLOVER

Interventions are often designed to incorporate ‘multiplier effects’ whereby effects in one community naturally spread to others nearby. These effects help to broaden the impact of an intervention (and are desirable if the impact is positive), but they can result in bias in impact evaluation conclusions when the effects on the treatment group spillover to the comparison group. When comparison groups also benefit from an intervention, for example, this can lead to an underestimation of impact since they appear better off than they would have been in the absence of the intervention. In some cases, spillovers can be mapped and measured, but most often, they must be controlled in advance by selecting treatment and control groups or units that are unlikely to significantly interact with one another. For example, it is usually more appropriate to divide classrooms or schools into treatment and control groups rather than individual students.

A special case of spillover occurs in substitution bias wherein governments or other organizations target only the comparison group to provide services similar to those provided to the treatment group(s). This is best avoided by ensuring coordination between USAID projects and other development actors in the region.

DISSEMINATION

Evaluations are only useful to the extent that results are available to interested stakeholders and decision makers. The ADS requires that evaluation results be posted to the DEC within 90 days of completion. Impact

evaluation contracts could also specify additional ways of disseminating results including: publications in academic journals, two-to-four page “policy briefs” with key findings, as well as conferences, workshops, and videos or other media. In some cases “pilot” projects funded by USAID may be scaled up by host country governments or other partners. In those cases, involving implementers and government stakeholders in the evaluation early on can ensure that they are invested in results and will increase their willingness to scale up successful projects.

ADDITIONAL RESOURCES

The following resources provide more information on impact evaluations.

Millennium Challenge Corporation (MCC)

- Evaluations Resource Page: <http://www.mcc.gov/pages/results/evaluations>

World Bank:

- World Bank Evaluation resources: <http://go.worldbank.org/X5X013RJZ0>
- Impact Evaluation in Practice: <http://elibrary.worldbank.org/content/book/9780821385418>
- Handbook for Impact Evaluation: <http://go.worldbank.org/9H20R7VMP0>
- The Strategic Trust Fund for Impact Evaluation: <http://go.worldbank.org/Q2XYY39FW0>
- The Development Impact Evaluation Initiative: <http://go.worldbank.org/IFIW42VYV0>

Abdul Latif Jameel Poverty Action Lab (JPAL)

- Methodology Resources: <http://povertyactionlab.org/methodology>
- 'Evaluating Social Programs' Course: <http://www.povertyactionlab.org/course/> (An online version is available for free on I-tunes: <https://itunes.apple.com/us/course/abdul-latif-jameel-poverty/id495065985>)

InterAction

- Impact Evaluation Guidance Note and Seminar Series: <http://www.interaction.org/impact-evaluation-notes>

International Initiative for Impact Evaluation: <http://www.3ieimpact.org/en/>

- Theory-Based Impact Evaluation: Principles and Practice: http://www.3ieimpact.org/media/filer/2012/05/07/Working_Paper_3.pdf

Center for Global Development:

- 'Evaluation Gap Working Group': http://www.cgdev.org/section/initiatives/_active/evalgap
- When Does Rigorous Impact Evaluation Make a Difference? The Case of the Millennium Villages: http://www.cgdev.org/sites/default/files/1424496_file_Clemens_Demombynes_Evaluation_FINAL.pdf

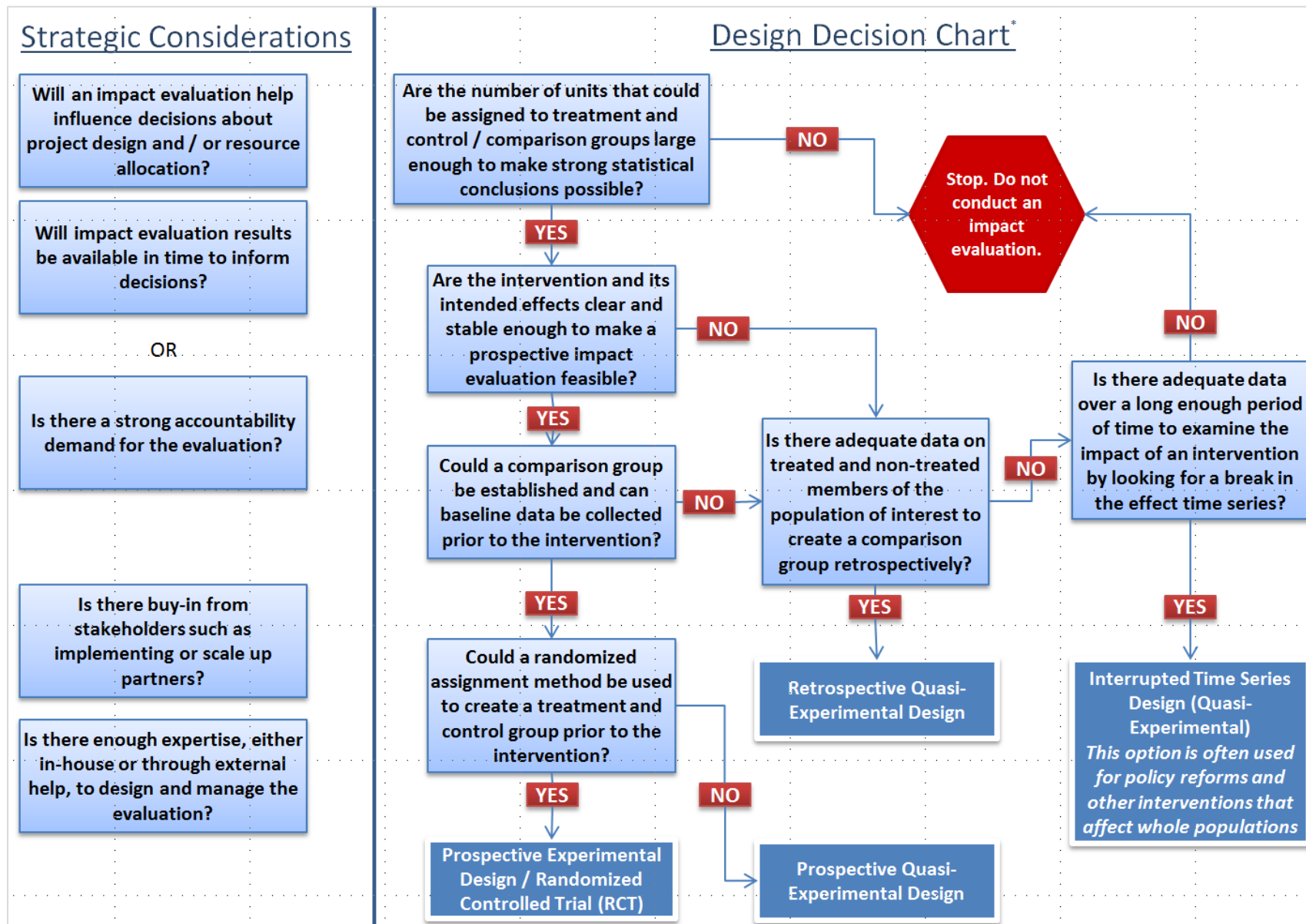
USAID:

- Evaluation for Evaluation Specialists (EES) Course – <http://university.usaid.gov/>
- Value chain wiki: <http://microlinks.kdid.org/good-practice-center/value-chain-wiki/impact-assessment>
- E3 Trade Facilitation Monitoring, Evaluation and Learning Toolkit: <http://usaidsite.carana.com/content/evaluation-pathway-4-rigorous-impact-evaluations>
- Feed the Future M&E Guidance Series Volume 4 – Impact Evaluation http://www.feedthefuture.gov/sites/default/files/resource/files/Volume4_FTFImpact.pdf

Additional Information:

- Sample Size and Power Calculations: <http://www.statsoft.com/textbook/stpowan.html>
- <http://www.mdrc.org/publication/core-analytics-randomized-experiments-social-research>

ANNEX I: DECISION AIDS



*Adapted from the Project Starter toolkit developed by Carana Corporation for the Office of Trade and Regulatory Reform (E3/TRR)

EVALUATION RESOURCE

EVALUATION IN THE COUNTRY DEVELOPMENT COOPERATION STRATEGY

Missions are required to include the following evaluation components in their Country Development Cooperation Strategy (CDCS) (ADS 201.3.3.4):

1. Identification of high-priority evaluation questions for each Development Objective (DO) that can address: (1) the development hypotheses and key assumptions underlying the DO; (2) estimated DO impact; (3) policy approach in a specific sector; and/or; (4) the efficiency of the USAID implementation approach (with attention to costs). High-priority evaluation questions may apply to performance evaluations or impact evaluations.
2. Identification of at least one opportunity for impact evaluation of a project or project component within each DO. (Note: This is not required for an abbreviated CDCS.)
3. Incorporation by missions of USAID's Gender Equality/Female Empowerment Policy by asking relevant evaluation questions about whether reducing gaps between males and females contributes to project outcomes.

It is not expected that every high-priority evaluation question or impact evaluation opportunity identified in the CDCS will be evaluated. Rather, the CDCS process provides a chance for mission leadership and Technical Officers to consider evaluation questions and which performance and impact evaluation opportunities could be operationalized, if feasible, during project design stages.

Tips for Writing Good Evaluation Questions for a CDCS

1. Frame your high-priority evaluation question as a specific *question*. In some CDCS documents, the evaluation “question” is written as, “Evaluate the effectiveness of the programs in achieving the DO.” This is neither a question nor specific enough to make it a high priority. Write your evaluation questions clearly and specifically. Moreover, it is best at this stage to keep your high-priority evaluation questions to one question per evaluation. You should have multiple evaluation questions in your CDCS, but they should not all be questions that relate to a single planned evaluation.

2. Make it useful. The answer to the evaluation question that you propose should be useful for decision-making. The purpose of developing evaluation questions at the CDCS stage of the program cycle is not only to make sure that missions do evaluations. Rather, the purpose is to encourage missions to consider during the strategy process what information they will want or need from an evaluation at a later date. So consider what the most useful thing you need to know may be. The evaluation questions should not try to answer everything—just the most important thing that can be addressed with an evaluation. Also, a question is not very useful if the answer is expected to come directly from top-line monitoring data. Ask questions that might require more digging through monitoring data or that will require additional data collection.

EVALUATION RESOURCE

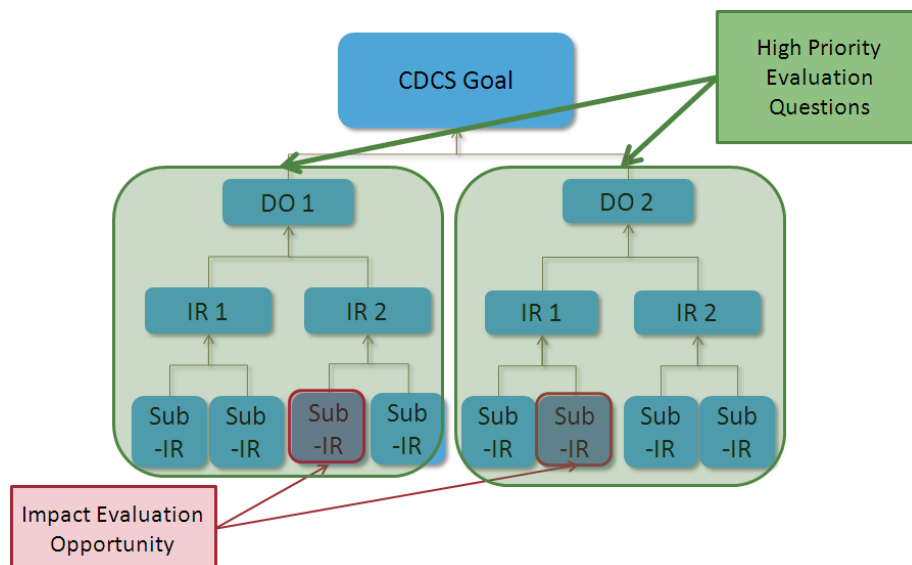
3. Know what evaluations are required. Keep in mind which of your projects (which typically will be at the intermediate result level) under each DO are likely to be required to be evaluated:

- Large projects: Those at or above average dollar value for projects within each DO managed by an Operating Unit (performance evaluation); or
- Pilot projects/activities of any size: Those that demonstrate new approaches that are anticipated to be expanded in scale or scope if the approach is proven successful (impact evaluation, if feasible).

4. Think also about potential evaluations that are not required. Once you start implementing the CDCS, you are likely to focus on meeting your evaluation requirements and conducting evaluations of individual projects and activities. During strategy development, it is a good time to think about other kinds of evaluations you could do—such as evaluations of higher-order results (e.g., DOs), evaluations of critical assumptions, or evaluations that cut across projects—while you have these issues fresh in your mind. Consider evaluation questions that address:

- The development hypotheses and key assumptions;
- Estimated impact of interventions;
- Policy approach in a specific sector; and/or
- The efficiency of the USAID implementation approach.

5. Impact evaluations are generally best suited for project components or interventions—not an entire DO. Impact evaluations are best for focusing on project components or interventions that are being piloted with an aim to scale up. Do not focus impact evaluations on a program that is supposed to have nationwide impact, since it is not easy to measure such impact when you do not have a nationwide counterfactual. It is also best not to focus an impact evaluation on large, complicated projects with many components and activities, since you will not know exactly which part of the intervention caused the impact. Finally, the intervention you evaluate with an impact evaluation should be fairly stable, not constantly adapting. The less stable the intervention, the less useful the impact evaluation, since you will not know exactly what caused the impact (if you can even determine impact).





Performance Management Plan (PMP) Toolkit

Module 2.9: Develop/Refine PMP Evaluation Plans

OVERVIEW

The PMP Evaluation Plan helps the Mission to identify and track evaluations over the course of the R/CDCS timeframe. This module describes a Mission PMP Evaluation Plan and how and why it supports performance monitoring. The module also describes the roles and responsibilities associated with the development and refinement of the PMP Evaluation Plan and what tools can assist with Evaluation Plan development.

TOOLS

- [Evaluation Plan Summary and Schedule](#)
- [Worksheet: Which Evaluations are Required?](#)

PMP EVALUATION PLANS - OVERVIEW

One of the PMP requirements is an Evaluation Plan that identifies and tracks evaluations across the Mission over the entire R/CDCS timeframe. The PMP Evaluation Plan should include (at a minimum):

- The projects/activities/programs to be evaluated
- Evaluation type (performance or impact evaluation)
- Possible evaluation questions
- Estimated budgets
- Planned start dates and estimated completion dates of evaluations (usually presented in a Gantt chart)
- Whether the evaluation is required

Roles and Responsibilities

The project teams will prepare project M&E plans as a part of the Project Design process, and the Mission's PMPOCs will ensure the project plans meet requirements, are consistent with Mission R/CDCS, and are reflected in the PMP's multi-year Evaluation Plan. Project teams should work with the Mission's Program Office to ensure that the Mission-wide PMP is regularly updated from new project M&E plans.

The Evaluation Plan may include additional information useful for planning and tracking evaluations, such as:

- Evaluation titles and key questions
- POCs for the evaluations
- Start and end dates of projects/activities
- Reason for evaluation
- Whether the evaluation will be externally led or internally led

See [Sample Multi-Year Mission-Wide Evaluation Summary and Schedule](#) (Annex 18).

DEVELOPING THE PMP EVALUATION PLAN

The following tasks represent the key steps in developing the PMP Evaluation Plan, broken up into those that occur while the PMP is still in the process of being developed and those that come after PMP approval.

During the initial PMP development phase, it is necessary to do the following:

1. **Review the illustrative evaluation questions and impact evaluation opportunities** listed in the R/CDCS. During the R/CDCS process, each Mission is required to provide illustrative high-priority evaluation questions for each Development Objective (DO) and identify an impact evaluation opportunity for each DO. These should be reviewed during PMP development. If these questions and opportunities remain relevant, then they should be included in the Evaluation Plan with the additional details required in the Evaluation Plan.
2. **Review currently planned or ongoing evaluations.** Evaluations that were planned prior to the R/CDCS approval that are still planned to continue or are currently ongoing should be included in the Evaluation Plan.
3. **Determine required evaluations.** Certain projects are required to be evaluated over the life of the R/CDCS per ADS 203. These include:

- *Large projects.* Each USAID Mission is required to conduct at least one evaluation of each large project it implements. For these purposes, a “large project” is one that equals or exceeds in dollar value the mean (average) project size for each Development Objective (DO) for the USAID Mission. All field Operating Units should calculate the average project size at the DO level using the definition for project provided in ADS 201.

The goal of this approach is to ensure that major projects in each DO undergo evaluation, even when a DO is a relatively small share of an OU’s budget. Missions can use several means of calculating a large project. The main principle is that Missions conduct an appropriate analysis to determine the mean project size and document their analysis. See the [Which Evaluations Are Required? Worksheet](#) (Annex 19) to help calculate and identify “large” projects.

- *Innovative activities.* Additionally, any activity within a project involving untested hypotheses or demonstrating new approaches (e.g., designated as “pilot” or “proof of concept”) that are anticipated to expand in scale or scope through USG foreign assistance or other funding sources will, if feasible, undergo an impact evaluation. If it is not possible to effectively undertake an impact evaluation, USAID Missions may undertake a performance evaluation, provided that the final evaluation report includes a concise but detailed statement about why an impact evaluation was not conducted.
4. **Determine non-required evaluations selected for management purposes.** USAID Missions are encouraged to identify opportunities for evaluations at the program or sector level. This is particularly valuable in a period preceding the development of a new strategy. USAID Missions are also encouraged to evaluate additional projects for learning or management purposes at any point in implementation. Evaluations should be timed so that their findings can inform decisions such as exercising option years, designing a follow-on program, creating a country or sector strategic plan, or making a policy decision.

Following initial PMP approval, it is necessary to do the following:

1. **Update and revise the Evaluation Plan** as new projects and activities are designed and as decisions are made regarding the details of a planned evaluation.
2. **Include additional evaluations that were not planned.** In the course of implementing a project, the following situations could serve as triggers for an otherwise unplanned evaluation:
 - A key management decision is required, but there is inadequate information to make it;
 - Performance information indicates an unexpected result (positive or negative) that should be explained, such as unanticipated results affecting either men or women (refer to gender analysis conducted per ADS 201);
 - Customer, partner, or other informed feedback, such as a contractor performance evaluation required by the Federal Acquisition Regulation (48 CFR Subpart 42.15) and USAID Acquisition Regulation (48 CFR Subpart 742.15) (ADS 302.3.8.7), suggests that there are implementation problems, unmet needs, or unintended consequences or impacts;
 - Issues of sustainability, cost-effectiveness, or relevance arise;
 - The validity of Results Framework hypotheses or critical assumptions is questioned—for example, due to unanticipated changes in the host country environment; or
 - Periodic Portfolio Reviews have identified key questions that need to be answered or require consensus.
3. **On an annual basis, update the evaluation registry section of the Performance Plan and Report** with information about evaluations completed in the past year and ongoing and planned evaluations based on the PMP Evaluation Plan.

Good Practice

In developing the Evaluation Plan, Missions should revisit the PIRS to ensure that any performance indicators needed for a planned evaluation (in addition to those indicators already identified for performance monitoring) are collected at baseline and on an ongoing basis. In developing the Evaluation Plan, Missions should ensure that baseline data collection is done prior to project or activity implementation. Although it is always good practice to collect data on target and comparison groups (i.e. a group not part of the project), for impact evaluations baseline data must be collected for treatment and control or comparison groups. (See ADS 203.3.1.1).

DEVELOPING PROJECT MONITORING AND EVALUATION PLANS

Missions must develop a project M&E plan during the Project Design phase and include it as an annex to their Project Appraisal Document (PAD) (see ADS 201.3.9.4). The project M&E plan serves to measure progress towards planned results and identify the cause of any delays or impediments during implementation. The M&E Plan for the project:

- Provides a framework for monitoring and evaluation that pulls together performance information from all activities contributing to a project;
- Identifies what questions will be addressed through evaluation, sketches out evaluation methods or approaches, and plans any data collection in addition to that identified for monitoring; and
- Constitutes one component of a broader Mission learning plan that guides Missions in strengthening the evidentiary base of their portfolios, speeds learning, and adapts project implementation to achieve high-quality development results as quickly and sustainably as possible

The evaluation portion of the project M&E Plan should include the following:

1. Description of what type of evaluation, if any, is required under ADS 203. If an evaluation of the project is not required under ADS 203 (i.e., if the project is not large or innovative), the DO team or Mission leadership could still decide to plan for an evaluation for other management or learning purposes.
2. A limited number of key evaluation questions that are explicitly linked to specific future decisions made by USAID or other key stakeholders or essential elements of learning.
3. Additional information about the evaluation, such as whether it is a performance evaluation or an impact evaluation. The Evaluation Plan should identify when the evaluation will take place during the project and provide a timeline for specific actions needed to draft the evaluation scope of work, procure an external evaluation team, and finalize the evaluation in time to inform specific decisions.

For impact evaluations, project design and evaluation design must be developed together so that baseline data can be collected on both the treatment and control groups. Parallel contracts are one option to consider as they can be procured to bring on an evaluation team at the same time as the Project Design team.

4. The estimated budget that will be set aside from the project budget and used for the evaluation.

The Project M&E Plan is included as an Annex to the Project Appraisal Document (PAD). After the PAD is approved, the PMP Evaluation Plan should be updated to incorporate any planned evaluations over the life of the project.

SUMMARY

By now you should have an understanding of:

- The importance of the PMP Evaluation Plan in managing evaluations across the life of the R/CDCS
- How to develop the PMP Evaluation Plan
- How to update the PMP Evaluation Plan following the development and approval of Project Monitoring and Evaluation Plans

REFERENCES

[ADS 201](#)

[ADS 203](#)

Importance of Planning for Evaluation during Project Design

Planning ahead for evaluations during Project Design better ensures that evaluations are relevant, timely, and useful. If an impact evaluation is planned, its design should be summarized in the Project Appraisal Document. Impact evaluations require that project implementation incorporate specific design requirements and data collection needs for effectively estimating project impact, including designating a 'target' group from the 'control' group throughout the life of the project.

Evaluation also strengthens the analytical quality of the Project Design process and potentially affects project implementation by:

- Clarifying project logic and development hypotheses;
- Identifying knowledge gaps and implicit assumptions;
- Defining key evaluation questions that will guide identification of performance indicators and data collection; and
- Contributing to plans to ensure learning during implementation.

EVALUATION RESOURCE

WHICH EVALUATIONS ARE REQUIRED?

PMP Toolkit Excerpt Annex 19: Part 2, Module 9

Evaluations are required for large and pilot projects¹ of each Development Objective. There is no single required way to calculate “large” projects, but this chart may be a useful place to start:

Calculating Large Projects

	Total Program Budget for R/CDCS period		# Projects for		Large projects = > this #	List Large Projects
DO 1		÷		=		
DO 2		÷		=		
DO 3		÷		=		

The guidance recommends that “OUs should calculate the average project size at the Development Objective level”. Following the R/CDCS, the mission should have a sense of how activities and projects align to each of the DOs, or if they don’t align to DOs and are being phased out. Once you know how the projects align to each DO and you have an estimate of how much each project will cost, then you should be able to calculate large projects.

a. Question: What about projects funded before the new R/CDCS that will continue, but do not “fit” under the new R/CDCS? Should they be evaluated?

- i. Example: Georgia has a number of ongoing projects that were funded by a one-time \$1 billion dollar pool of supplemental funds for reconstruction prior to the R/CDCS. They do not “fit” under the new R/CDCS.
- ii. Answer: On-going projects that don’t fit under the new DOs do not need to be evaluated under the new R/CDCS.

b. Question: Evaluate projects or mechanisms?

- iii. Answer: If 3 mechanisms comprise a single project (less than one project) and that project is a “large project” within its DO, then you are required to do one evaluation of that project. Such an evaluation may address the project as a whole, just one of the mechanisms, or even a component of one of the mechanisms.

¹ A set of planned and then executed interventions identified through a design process, which are together intended to achieve a defined development result, generally by solving an associated problem or challenge. The term project does not refer only or primarily to an implementing mechanism, such as a contract or grant. (USAID Evaluation Policy, p. 4).

EVALUATION PLAN SCHEDULE AND SUMMARY TEMPLATE

PMP Toolkit Excerpt: Annex 18: Part 2 Module 9

This template includes a sample multi-year evaluation plan for a PMP and a table that can be utilized to prepare an evaluation summary.

Sample Multi-Year Evaluation Plan for a PMP

Evaluation Plan Schedule

Evaluation	FY13				FY14				FY15				FY16				FY17			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Health Performance Evaluation																				
Economic Growth Performance Evaluation																				
Education Performance Evaluation																				
Local Governance Impact Evaluation																				
Agriculture Performance Evaluation																				

Key

1. Design and SOW Start		
2. Final SOW		
3. Awarded by		
4. Field Work		
5. Final Report Completed		

EVALUATION PLAN TEMPLATE

Evaluation Plan Summary

Evaluation Title/Questions	POCs	Project/ activity/ program to be Evaluated	P/A/P Start/ End Dates	Required (and reason required) or Optional	Evaluation Type (performance or impact), and Projected Use	Internal or external	Estimated Evaluation budget	Evaluation Start/ End Dates
Example								
Family Planning Project evaluation 1. To what extent did the MFP project increase capacity of local family planning centers? 2. Did use of modern family planning methods increase in target areas?	Speedy Analyst	Increased use of modern family planning methods Project	Jan. 2011/ Dec. 2013	Required – large project	Performance; to decide whether to exercise option years or re- compete	External	\$180,000	Dec. 2012/ Dec. 2013

Instructions: see notes on the next page.

EVALUATION RESOURCE

Notes on Evaluation Plan Summary Fields

In all cases, if the information for a particular field is not yet known, enter TBD, but update field as relevant decisions are made.

Field	Instructions
Evaluation Title/Questions	Include the planned Evaluation title and any key questions that have been identified thus far. These questions may come from the R/CDCS, a Project Design Document, or other evaluation planning if such planning is already underway. Only include the 1-5 key questions, not detailed sub-questions.
POC	Enter the point of contact(s) for the evaluation with responsibility for ensuring the evaluation is completed as planned. Ideally this will include one point of contact in the program office and one point of contact in the technical office.
Project/ activity/ program to Be Evaluated	Evaluations may focus on individual activities, projects, programs (an entire DO, for instance), or even cross-cutting issues. Enter here what is to be evaluated. If multiple projects, activities, or programs are to be included in the evaluation, include the name of each one that will be included.
P/A/P Start/ End Dates	Include the start and end dates of the projects, activities, and programs that are to be evaluated. If multiple projects, activities, and programs are included in the evaluation, include all start and end dates.
Required (and reason required) or Optional	Evaluations may be required because a project has been determined to be a large project or because it is a pilot or innovative project . A large project is one that equals or exceeds in dollar value the mean (average) project size for each Development Objective (DO) for the USAID Mission/Office. A pilot or innovative project is one that includes any activity within the project involving untested hypotheses or demonstrating new approaches that are anticipated to be expanded in scale or scope through USG foreign assistance or other funding sources. If an evaluation is required, note here whether it is because of the large project requirement or the innovative intervention requirement. If the evaluation is not required, but a commitment has been made to do the evaluation, then note here that it is an optional evaluation.
Evaluation Type (performance or impact), and Projected Use	Note here what type of evaluation is planned. There are two types of evaluations. Impact evaluations measure the change in a development outcome that is attributable to a defined intervention. Impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change. Performance evaluations often incorporate before-after comparisons, but generally lack a rigorously defined counterfactual. Performance evaluations focus on descriptive and normative questions, such as, what a particular project or program has achieved; How it is being implemented; How it is perceived and valued; Whether expected results are occurring; and other questions pertinent to program design, management, and operational decision-making.
Internal or external	Note here whether the evaluation is external or internal (and the type of internal evaluation). An external evaluation is one in which (at minimum) the lead evaluator is an independent expert outside of USAID, with no fiduciary relationship with the implementing partner. In most cases these will be managed by the Program Office. USAID Mission/Office management may make exceptions under unusual circumstances to management by the Program Office, but the exception should be documented in an addendum to this evaluation plan and included in the PMP. An internal evaluation is one that does not meet the standards of an external evaluation. These are generally of two types. An <i>implementer internal evaluation</i> is led by an individual with a fiduciary relationship to the implementing partner, such as an evaluation led by implementer staff or under a subcontract of the implementer. A <i>USAID internal evaluation</i> is one that is led by USAID staff.
Estimated Evaluation budget	Enter the estimated budget for the evaluation.
Evaluation Start/End Dates	Enter the estimated start date for the evaluation (i.e., when the evaluation will be awarded) and the estimated end date of the evaluation. Note that numerous steps must take place prior to the estimated start date, such as development of the Statement of Work.

PROGRAM CYCLE RESOURCE

EVALUATION IN PROJECT MONITORING AND EVALUATION PLANS

Missions must develop a project Monitoring and Evaluation (M&E) plan during the Project Design phase and include it as an annex to their Project Appraisal Document (PAD) (see ADS 201.3.9.4). The project M&E plan serves to measure progress towards planned results and identify the cause of any delays or impediments during implementation. The M&E Plan for the project:

- Provides a framework for monitoring and evaluation that pulls together performance information from all activities contributing to a project;
- Clearly describes how the project will collect needed data from project inception (baseline data) and periodically over the life of the project.
- Identifies what questions will be addressed through evaluation, sketches out evaluation methods or approaches, and plans any data collection in addition to that identified for monitoring; and
- Constitutes one component of a broader Mission learning plan that guides Missions in strengthening the evidentiary base of their portfolios, speeds learning, and adapts project implementation to achieve high-quality development results as quickly and sustainably as possible.

Importance of Planning for Evaluation during Project Design

Planning ahead for evaluations during Project Design better ensures that evaluations are relevant, timely, and useful. If an impact evaluation is planned, its design should be summarized in the Project Appraisal Document. Impact evaluations require that project implementation incorporate specific design requirements and data collection needs for effectively estimating project impact, including designating a 'target' group from the 'control' group throughout the life of the project.

Evaluation also strengthens the analytical quality of the Project Design process and potentially affects project implementation by:

- Clarifying project logic and development hypotheses;
- Identifying knowledge gaps and implicit assumptions;
- Defining key evaluation questions that will guide identification of performance indicators and data collection; and
- Contributing to plans to ensure learning during implementation.

The evaluation portion of the project M&E Plan should include the following:

Description of what type of evaluation, if any, is required under ADS 203. If an evaluation of the project is not required under ADS 203 (i.e., if the project is not large or includes a pilot that will be scaled-up), the DO team or Mission leadership could still decide to plan for an evaluation for other management or learning purposes.

1. A limited number of key evaluation questions that are explicitly linked to specific future decisions made by USAID or other key stakeholders or essential elements of learning.

PROGRAM CYCLE RESOURCE

2. Additional information about the evaluation, such as whether it is a performance evaluation or an impact evaluation. The Evaluation Plan should identify when the evaluation will take place during the project and provide a timeline for specific actions needed to draft the evaluation scope of work, procure an external evaluation team, and finalize the evaluation in time to inform specific decisions.
3. For impact evaluations, project design and evaluation design must be developed together so that baseline data can be collected on both the treatment and control groups. Parallel contracts are one option to consider as they can be procured to bring on an evaluation team at the same time as the Project Design team.
4. The estimated budget that will be set aside from the project budget and used for the evaluation.

The Project M&E Plan is included as an Annex to the Project Appraisal Document (PAD). After the PAD is approved, the PMP Evaluation Plan should be updated to incorporate any planned evaluations over the life of the project.

PROGRAM CYCLE RESOURCE

ACTIVITY MONITORING AND EVALUATION PLANS

At the activity/implementing mechanism level, implementers are expected to submit an activity M&E plan to USAID CORs/AORs within the first 90 days of an award (generally at the same time as an approved work plan) and before major activity implementation actions begin.

These activity M&E plans describe the performance monitoring indicators and associated performance indicator reference sheets (PIRS) that the implementer will collect and report, along with any internal evaluations that the implementer plans to conduct. Other recommended components of an Activity M&E Plan include a logic model or framework, implementing team M&E roles and responsibilities, M&E task calendar, and data collection instruments to be used as part of the implementer's internal monitoring and evaluation process.

Before Award

Consistent with ADS 300.3.5, all solicitations should include language designating indicators derived from the Project M&E Plan which the implementing partners will be required to collect and report.

Per ADS 203, during acquisition and assistance solicitation, CORs/AORs must work with COs/AOs to ensure that solicitations include instructions to offerors/applicants to include the costs of data collection, analysis, and reporting as a separate line item in their budgets to ensure that adequate resources are available. In the design of the activity, USAID may also consider including a requirement or suggestion for an M&E specialist to be included on the implementing partner's team. If this is included in the technical requirements then it is important to work with the CO/AO to ensure that this is clear in the instructions for the cost proposal and is linked to the requirement for budgeting for M&E. A dedicated line item for M&E should also be included in the award budget so that adequate resources are allocated and it is more easily tracked.

There is not a one-size-fits-all approach to how much M&E should be budgeted at the award/mechanism level. Some awards/mechanisms could require 10%, or maybe even more, to undertake monitoring and evaluation (e.g. impact evaluations of activities/IMs or interventions in remote geographic areas); others could require 1% or less of the total budget.

After Award

Following award, project managers must work with assigned COR/AOR for the activity to review that the activity M&E plan is complete and consistent with other mission M&E planning documents. In particular, the performance indicators (and their related PIRS, baselines, targets, disaggregates) in the activity M&E plan should be consistent with and meet the data collection needs of the Project M&E plan, the mission's Performance Management Plan (PMP), and the Performance Plan and Report. COR/AORs must also ensure that implementing partners are aware of any planned external evaluations that will focus on the implementing partner's activities and that the activity M&E plan responds to the data needs of the planned evaluation.

PROGRAM CYCLE RESOURCE

CONSIDERATIONS FOR THE M&E OF G2G ASSISTANCE

ADS 220 M&E Requirements (in conjunction with ADS 201 and ADS 203)

- ADS 220 outlines considerations for the monitoring and evaluation of **risk** AND the monitoring and evaluation of **performance**.
 - The “monitoring” of G2G projects involves a range of mission staff to develop a comprehensive monitoring approach (e.g. technical and program office staff for performance monitoring; controllers for monitoring and oversight of risk mitigation measures and projects that build institutional capacity for PFM).
- Per ADS 220.3.b(5), the Project Appraisal Document (PAD) for G2G projects should:
 - Include an M&E plan, per ADS 203, that encompasses [performance] monitoring of all project activities; and
 - Specifically address the coordination, oversight, monitoring, and evaluation of any risk mitigation measures. The monitoring plan should:
 - Include provisions to ensure partner government follow-up on any risk mitigation measures
 - If appropriate, incorporate the results of the PFMRAF (stages 1 and 2) and any technical assistance to address diagnosed weaknesses.
- Per ADS 220.3.4.1, the Approval for Use of Partner Government Systems (AUPGS) must:
 - Outline the management, monitoring, and reporting roles and responsibilities over the project’s G2G activities (including adherence to the risk mitigation plan); and
 - Document that the PFM systems that will be used for project implementation will be subject to evaluation of **the outcomes and effectiveness of the project’s G2G activities and of related capacity building support** to implementing partner government entities.
- Per ADS 220.3.4.3, USAID and its partner government counterparts must agree on a reporting plan that includes periodic progress reports from the responsible government counterpart, including:
 - Reporting on the performance indicators identified in the M&E plan that measure progress towards goals and objectives of the USAID-funded project;
 - Periodic implementation progress meetings; and
 - Subsequent, jointly-agreed upon plans of action to address implementation problems.

Considerations during Project Design that Inform M&E

- **Theory of Change.** What is the rationale for using G2G assistance? What is the role of the government entity vis-à-vis other local actors in achieving desired development outcomes?
- **Systems Lens.** What resources, (formal and informal) roles, relationships, and rules guide the system or network in which the partner government entity is embedded? (see ADS 220.3(a))
- **Capacity Development Objectives.** What capacities (e.g. individual, institutional, system) are needed to achieve and sustain which results? (Also consider: incentives for capacity development, different actors’ perspectives of “capacity,” time horizons, sequencing of capacity development efforts, emergent capacity needs).

PROGRAM CYCLE RESOURCE

- **Multi-actor Monitoring of Performance.** What are the roles and responsibilities of the potential actors, in addition to USAID, which will be engaged in monitoring? For example:
 - partner government M&E or audit offices;
 - partner government IPs (e.g. sub-national government entities or NGOs);
 - USAID-contracted third party technical assistance providers;
 - USAID M&E platforms; and
 - third party stakeholders in oversight, quality assurance, or accountability roles, such as CSOs/CBOs engaged in citizen monitoring, local engineering or construction firms, etc.
- **Sustainability.** Which institutional capacities, along with which actors, interrelationships, and incentives, are needed to achieve and sustain development outcomes?

Considerations for Identifying Performance Metrics and Evaluation Questions

- Selection of **performance metrics** requires thinking through the logic chain associated with desired capacity development and development outcomes.
 - Identification of Needed Capacities to Improve Institutional Performance → Strengthened Institutional Capacity → Improved Performance and More Effective Delivery of Services (within a system) → Sustainability of outcomes (within a system).
 - PFMRAF or similar assessments (e.g. PEFA) may inform metrics or be used as a baseline to measure certain aspects of progress. *Update: a Local Solutions Task Team is looking at government performance measurement, including PFM indicators that track with other donors.*
 - As generally we can only measure our contribution, as compared to attribution, to observed outcomes, we may need to think creatively about measurement approaches (e.g. complexity-aware monitoring).
- Consider **evaluation questions** that include the **effectiveness and sustainability** of the use of partner government systems in meeting assistance objectives and the **effectiveness of related capacity building support** to partner government entities.

Opportunities to Engage Partner Government Entities on M&E

- | | |
|---|--|
| <ul style="list-style-type: none">• Defining mutually agreed upon objectives• Problem analysis and solution identification• Identifying performance indicators (particularly those aligned with data already collected through government systems)• Clarifying/negotiating roles and responsibilities <u>upfront</u>, including with regard to performance monitoring (ADS 203 applies), data collection, and reporting• Clarifying expectations with regard to site visits, DQAs, reporting, and other oversight and accountability measures• Reviewing assumptions and risks and how these will be monitored | <ul style="list-style-type: none">• Determining how the project will collect needed data from project inception (baseline data) and periodically over the life of the project• Data collection in ongoing performance monitoring• Analysis and interpretation of monitoring data• Identifying priority evaluation questions• Engagement in the dissemination and utilization of findings from evaluations and assessments• Identifying opportunities for joint implementation reviews and collaborative reflection and assessment of progress |
|---|--|

For more in-depth discussion, please see the discussion paper <i>Considerations for the Monitoring and Evaluation of G2G Activities</i> , which is available on ProgramNet and Learning Lab.
--

EVALUATION RESOURCE

POST-EVALUATION ACTION PLANS: GUIDANCE AND TEMPLATES

Introduction

The USAID Evaluation Policy and Automated Directives System (ADS 203) underscore the significance of evaluation as a performance management tool that can provide evidence-informed decision-making. Specifically, ADS 203.3.1.9, **Responding to Evaluation Findings**, states:

“USAID Missions/Offices should address findings and recommendations of evaluations that relate to their specific activities and Development Objectives (DOs). To help ensure that institutional learning takes place, and evaluation findings can be used to improve development outcomes, Missions should take the following basic steps upon completion of the evaluation:

1. Meet with the evaluation team to debrief and discuss results or findings and provide feedback on any factual errors;
2. Review the key findings, conclusions, and recommendations systematically;
3. Determine whether the team accepts/supports each finding, conclusion, or recommendation;
4. Identify any management or program actions needed and assign responsibility and timelines for completion of each set of actions;
5. Determine whether any revision is necessary in the joint country assistance strategy or USAID Country Development Cooperation Strategy (CDCS), results framework, or project, using all available information; and
6. Share and openly discuss evaluation findings, conclusions, and recommendations with relevant customers, partners, other donors, and stakeholders, unless there are unusual and compelling reasons not to do so. In many cases, the USAID Mission/Office should arrange the translation of the executive summary into the local written language.”

This guidance describes the responsibilities and steps in preparing post-evaluation Action Plans, covering steps 2-4 in responding to evaluation report findings, conclusions, and recommendations as described above. This guidance, and the attached templates, are designed for USAID management (e.g., Contracting Officer's Representatives [CORs], Activity Managers, Technical Officers, etc.) to support a common process across USAID for responding to evaluations. This guidance does not take the place of a Mission Order on Evaluation, and USAID staff members are encouraged to consult their Mission Order to determine if there are specific procedures for their particular mission.

EVALUATION RESOURCE

Prepare the management response

Upon completion of an evaluation report, the Program Office will lead relevant Mission staff through the process of responding to evaluation findings, conclusions, and recommendations.¹ The Program Office and the relevant Technical Office(s) should review the evaluation report and consider how to respond to it. The particular individuals involved in such a review may vary, but should at least involve the COR(s)/Agreement Officer's Representative(s) (AOR) of the activity or activities being evaluated and the COR of the evaluation. Additional participation by the relevant Technical Office directors, the Program Office director, and their staff can help ensure sufficient internal support for the management response. For high-profile evaluations, mission leadership may also need to be involved in the management response. In addition, USAID staff and external stakeholders—particularly the implementing partners who were subject to the evaluation—should be consulted for their perspectives.

A sample Management Response template is included at the end of this document. The Management Response template focuses on recommendations from evaluations, but missions may choose to widen the focus to include report conclusions or narrow it to include only major recommendations. The template indicates whether the mission (or other operating unit) accepts, partially accepts, or rejects the recommendations stated in the evaluation report. When rejecting or partially accepting a recommendation, it provides space to document reasons as to why a particular recommendation is not accepted and to show amendments made to a partially accepted recommendation.

There are often very good reasons to reject evaluation recommendations, and missions/offices should not hesitate to initiate a robust discussion of the merits of individual evaluation recommendations. For example: external evaluators may be unaware of particular opportunities and constraints of USAID internal processes in making their recommendations; the evaluation conclusion that the recommendations seek to address may be poorly supported by the evaluation findings; or a recommendation may involve questionable assumptions about the transferability of good practices in distant contexts to the mission's own operating environment. Evidence from evaluations should ideally inform decision-making, but evidence alone does not dictate decisions.

In cases where mission management disagrees with evaluation conclusions and recommendations and wishes to publicly state those differences, the mission should draft an evaluation Statement of Difference to be appended to the evaluation report, as described [here](#).

Create a post-evaluation Action Plan

Based upon the decisions made in the management response, the Program Office and relevant Technical Office(s) should identify any management or program actions needed and assign responsibility and the timelines for completion of each set of actions completed. These management or program actions may be based on accepted recommendations explicitly suggested in the evaluation report, or they may be additional actions deemed appropriate by management based on evidence, findings, or conclusions in the evaluation report. If the actions

¹ The Standardized Mission Order on Performance Monitoring states that this process will be led by the Program Office. Missions should consult with their own Mission Orders regarding specific assignment of roles and responsibilities.

EVALUATION RESOURCE

necessitate revisions to the joint country assistance strategy, USAID CDCS, results framework, or project design, then these should be noted in the Plan. A sample post-evaluation Action Plan template is attached to help facilitate this process.

The post-evaluation Action Plan contains the following key elements:

- **Management and program actions needed:** Planned actions intended to achieve expected results that address needs identified in the evaluation or by appropriate stakeholders based on the evaluation evidence.
- **Reason for action:** Intended to document why the action has been deemed necessary. For instance, it may be based on a recommendation from the evaluation report that was accepted or partially accepted during the management response.
- **Individual responsible for completing action:** Specifies who will take the lead in ensuring the action is addressed in the manner laid out in the Action Plan.
- **Budget (if applicable):** Funds required to implement the planned action. Note that many recommendations will not have budgetary implications.
- **Date:** Indicates the date by which the action is expected to be completed.
- **Status:** A classification of the status of each action: (1) implementation not yet begun; (2) implementation underway; or (3) action completed.

Review the Action Plan

The Program Office and relevant Technical Office(s) should review the completed post-evaluation Action Plan to ensure that all relevant parties agree to the assigned responsibilities in the Plan. The Action Plan should then be saved somewhere that is accessible to USAID staff members and used as a regular reference point as recommendations are implemented.

Follow-up

The post-evaluation Action Plan should periodically be reviewed to determine if progress has been made on each of the actions in the Plan. Updates to the status can be described in the last column of the Plan. At the times specified in the deadlines established in the Action Plan, or at regular periodic intervals, the relevant Technical Office staff should meet with the Program Office to assess progress on each action in the Plan. If appropriate, joint discussions with relevant stakeholders should be organized to report on progress or to convene to identify solutions as roadblocks occur. During Portfolio Reviews, the status of Action Plans for evaluation findings, and their use in respective decisions, will be discussed and documented (as noted in the Standardized Mission Order on Performance Monitoring).

Post-evaluation Management Response Template

Evaluation Title:

Evaluation Completion Date:

Date of Management Response:

Evaluation Recommendation	Management Response : <i>Accept / Partially Accept / Reject</i>	If not accepted, give reasons for rejection or, if partially accepted, describe any amendments
1		
2		
3		
4		
5		

EVALUATION RESOURCE

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						

EVALUATION MANAGEMENT PROCESS CHECKLIST

Planning	
Task/Milestone	Completed Date
Evaluability Assessment conducted (optional)	
Decision made to undertake an evaluation	
Evaluation Manager/Contracting Officer's Representative (COR) named	
Evaluation parameters defined: <ul style="list-style-type: none"> • What type (performance or impact) • Who will evaluate (external or internal evaluators) • When it will be completed 	
Mission-wide Evaluation Plan updated	
Initial evaluation communication and dissemination plan developed	
Stakeholder input obtained (as appropriate)	
Evaluation purpose and questions determined	
Evaluation Statement of Work drafted	
Evaluation Independent Government Cost Estimate (IGCE) drafted	
Peer review of Evaluation Statement of Work	
Final Evaluation Statement of Work approved	
Mechanism selected (for external evaluations)	
Solicitation issued (for external evaluations)	
Managing	
Task/Milestone	Completed Date
Technical evaluation of proposals (review)	
Evaluation contract awarded (external) or evaluation team selected (internal)	
Disclosure of Conflict of Interest forms received	
Post-award orientation ("Kick-off") meeting	
Evaluation workplan submitted	
Evaluation background review/desk review/inception report submitted (if requested)	
Evaluation design submitted	
Evaluation design shared with country-level stakeholders and implementing partners for comment	
Evaluation design approved	

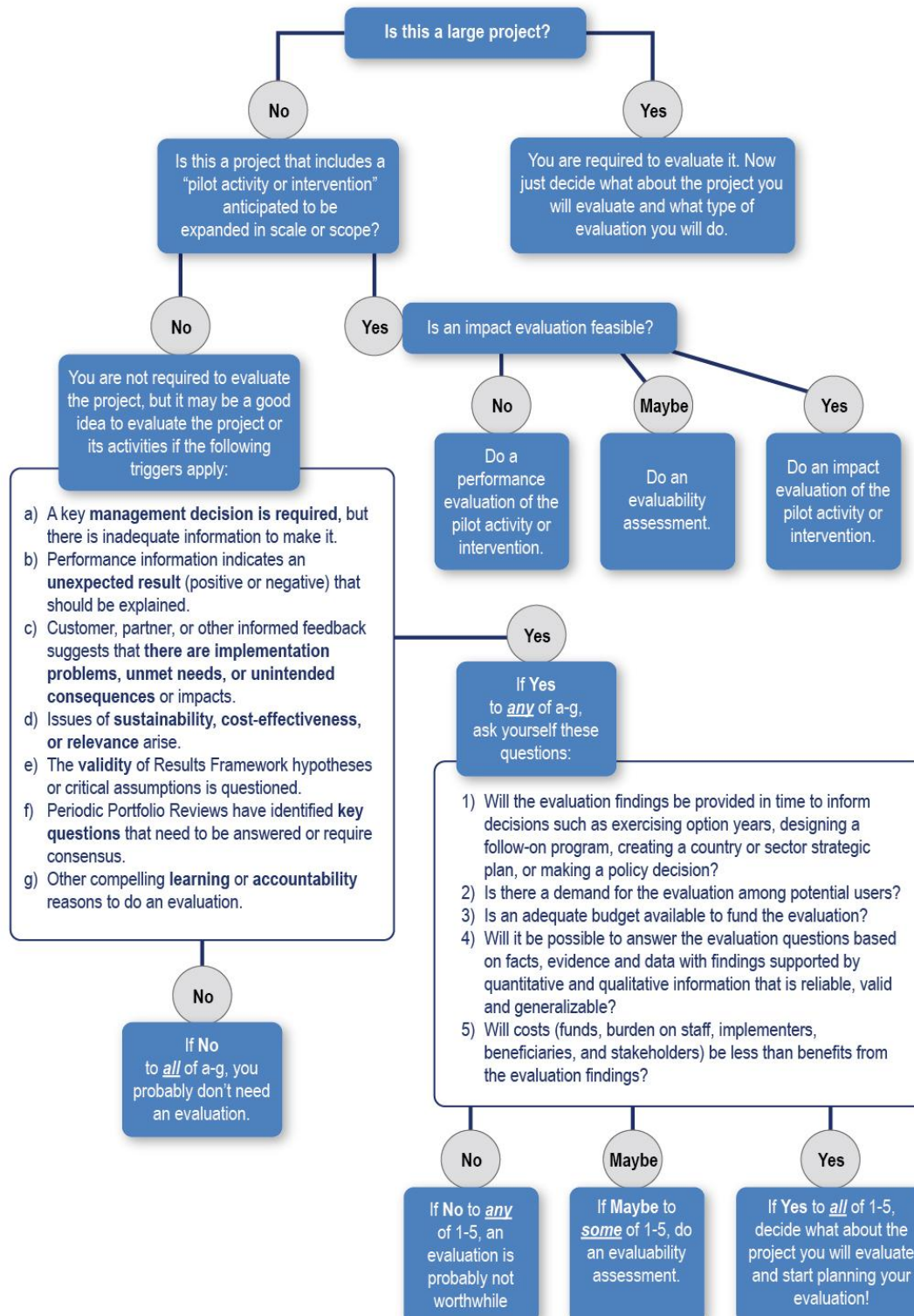
EVALUATION MANAGEMENT PROCESS CHECKLIST

Managing	
Task/Milestone	Completed Date
In-brief for Mission and/or Evaluation Stakeholders (if requested)	
Mid-evaluation briefing/periodic briefing (if requested)	
Out-briefing (if requested)	
Submission of draft report	
Peer review of draft report	
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	

Sharing, Reporting, Using, and Learning	
Task/Milestone	Completed Date
Evaluation Report dissemination plan updated	
Evaluation Report disseminated	
Evaluation Report submitted to USAID Development Experience Clearinghouse	
Evaluation data submitted to USAID Development Data Library (if applicable)	
Evaluation contractor performance assessed in Contractor Performance Assessment Reporting System (CPARS) (if applicable)	
Post-evaluation Mission 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)	

DECIDING TO EVALUATION FLOWCHART

This flowchart is intended to provide summary guidance to help a USAID Mission or Operating Unit decide if they should commission an evaluation of a USAID project or its component activities. For more detailed guidance, please see the USAID Evaluation Toolkit.



EVALUATION RESOURCE

USAID REQUIRED AND NON-REQUIRED EVALUATIONS: QUESTIONS AND ANSWERS

I. When is an evaluation required?

Large projects and pilot projects/activities that are to be scaled up are required to be evaluated with an external evaluation.

1. Large Projects

According to the Automated Directives System (ADS) 203.3.1.3:

"Each USAID Mission/Office is required to conduct at least one evaluation of each **large project** it implements. For these purposes, a "large project" is one that equals or exceeds in dollar value the mean (average) project size for each Development Objective (DO) for the USAID Mission/Operating Unit (OU). All field Operating Units should calculate the average project size at the Development Objective level. Use the definition of "project" provided in ADS 200. The goal of this approach is to ensure that major projects in each DO undergo evaluation, even when a DO is a relatively small share of an OU's budget. Missions can use several means of calculating a large project. The main principle is that Missions conduct an appropriate analysis to determine the mean project size and document their analysis."

How does one determine what is a "large project" requiring evaluation?

"Project" is defined not as an implementing mechanism but according to the current ADS definition: "A set of executed interventions, over an established timeline and budget, intended to achieve a discrete development result (i.e., the project purpose) through resolving an associated problem. It is explicitly linked to the Country Development Cooperation Strategies [CDCS] Results Framework."

All field OUs should calculate the average project size at the DO level. However, there is no required way of calculating "large projects." Field OUs should use available data, reasonable estimates, and defensible decisions in calculating what a large project is for their mission. The Bureau for Policy, Planning and Learning Office of Learning, Evaluation, and Research (PPL/LER) recommends that the mission divide the total estimated program expenditures over the entire life of a DO (i.e., over the CDCS timeframe) by the number of projects under that DO (current and planned) to calculate the mean. Projects with a budget or estimated budget during the life of the DO that is above the mean are required to be evaluated.

In cases where there are factors that make it difficult to calculate mean project size—for example, when many projects are co-funded with other U.S. Government partners—OUs should consult with PPL/LER to determine an appropriate means of calculation.

EVALUATION RESOURCE

Do you have an example of calculating a large project?

Imagine a mission with a CDCS that includes three DOs. One of the DOs focuses on health issues and includes three intermediate results, each of which is aligned to a project. The first project has a total planned budget of \$20 million, the second project has a budget of \$10 million, and the third project has a budget of \$6 million. Total planned budget of these three combined projects is \$36 million, and the average budget is \$12 million. Since only the first project has a planned budget that exceeds the average for this DO, only the first project with a total planned budget of \$20 million is required to be evaluated. A similar calculation should be conducted for the Mission's other DOs.

To meet the large project evaluation requirement, must a Mission/OU evaluate the whole project?

No. To meet the requirement of evaluating large projects, an OU must conduct at least one evaluation of the large project over the life of the large project, but the evaluation does not need to evaluate that project as a whole or even evaluate a majority of the activities/implementing mechanisms that constitute the large project. OUs should focus on answering those questions about their large projects or its constituent activities that are the most critical for filling in knowledge gaps and informing decision-making. An evaluation of the large project as a whole, an evaluation of one or more of the activities/implementing mechanisms under a large project, or even an evaluation of a single task under a large project would meet the large project requirement.

2. Pilot Projects/Activities

According to ADS 203.3.1.3:

“Any activity within a project involving untested hypotheses or demonstrating new approaches that are anticipated to be expanded in scale or scope through [U.S. Government] foreign assistance or other funding sources will, if feasible, undergo an impact evaluation. If it is not possible to effectively undertake an impact evaluation, USAID Missions/OUs may undertake a performance evaluation, provided that the final evaluation report includes a concise but detailed statement about why an impact evaluation was not conducted.”

How does one determine what is a “pilot project/activity” requiring evaluation?

DO teams should identify pilot project/activities during the project design stage. Not every untested hypothesis or new approach is required to be evaluated. Evaluation is only required where there is well-supported expectation or clear plan to expand or scale up the new approach if the findings of the evaluation support the expansion.

To meet the pilot project evaluation requirement, must an OU evaluate the whole project?

No. You do not need to do an impact evaluation of an entire project. Note that the terminology “pilot activity” is generally more appropriate in the USAID context than “pilot project,” since it is often not feasible or advisable to do an impact evaluation of the project as a whole. Instead, focus on answering the questions that would be most useful in informing future scale up. One option is to evaluate alternative ways to deliver an intervention.

EVALUATION RESOURCE

II. Are missions required to conduct evaluations based on the high-priority evaluation questions and impact evaluation opportunities described in their CDCS?

No. Missions must identify high-priority evaluation questions and opportunities for an impact evaluation for each DO in their CDCS. While these high-priority evaluation questions and impact evaluation opportunities should inform a mission's evaluation plan, missions are not required to complete evaluations based on them. USAID policy recognizes that mission information and evidence needs may change over the life of the CDCS, and investigations into the feasibility of an evaluation may prompt changes to evaluation plans. However, if a mission does not choose to complete evaluations based on these high-priority evaluation questions and impact evaluation opportunities, it is recommended that missions document these decisions.

III. When should one consider a non-required evaluation?

As stated in ADS 203.3.1.3:

"USAID Missions/Offices may evaluate additional projects for learning or management purposes at any point in implementation. Evaluations should be timed so that their findings can inform decisions such as exercising option years, designing a follow-on program, creating a country or sector strategic plan, or making a policy decision. In the course of implementing a DO, the following situations could serve as triggers for an evaluation:

- A key management decision is required, but there is inadequate information to make it;
- Performance information indicates an unexpected result (positive or negative) that should be explained, such as unanticipated results affecting either men or women (refer to gender analysis conducted per ADS 201);
- Customer, partner, or other informed feedback, such as a contractor performance evaluation required by the Federal Acquisition Regulation (48 CFR Subpart 42.15) and USAID Acquisition Regulation (48 CFR Subpart 742.15)(ADS 302.3.8.7), suggests that there are implementation problems, unmet needs, or unintended consequences or impacts;
- Issues of sustainability, cost-effectiveness, or relevance arise;
- The validity of Results Framework hypotheses or critical assumptions is questioned, for example, due to unanticipated changes in the host country environment; or
- Periodic Portfolio Reviews have identified key questions that need to be answered or require consensus."

Missions are encouraged to make strategic choices and prioritize evaluations based on management and learning purposes. Missions are also encouraged to consider more evaluations at the project level or the DO level—such as those that address issues across mechanisms within a project or within a DO—rather than just focusing on a single mechanism. Non-required evaluations may include, for instance, DO-level evaluations or evaluations based on the high-priority evaluation questions identified in the CDCS.

EVALUATION RESOURCE

IV. Which policies apply to required and non-required evaluations?

Must all evaluations (required and non-required) be external?

In short, required evaluations must be external. Non-required evaluations may be external or internal.

1. Required Evaluations

According to ADS 203.3.1.1:

“Required evaluations (i.e., for large or pilot projects, ADS 203.3.1.3) at USAID must be led by an external team leader.”

And from ADS 203.3.1.7:

“For required evaluations (i.e., large or innovative), the evaluation team leader must be an independent expert from outside USAID with no fiduciary relationship with the implementing partner.”

USAID staff may participate in required external evaluations but may not lead them. From ADS 203.3.1.7:

“In cases where USAID Mission/Office management determines that appropriate expertise exists within the Agency, and that engaging USAID staff in an evaluation will facilitate institutional learning, an evaluation team may be predominantly composed of USAID staff. However, an outside expert with appropriate skills and experience will be recruited to lead the team, mitigating the potential for conflict of interest. The outside expert may come from another [U.S. Government] agency uninvolved in project implementation or be engaged through a contractual mechanism.”

2. Non-required Evaluations

Non-required evaluations may be external or internal. For instance, non-required evaluations can be led by a contracted external team leader, led by a USAID staff member, or led by the implementing partner.

Must all evaluations (required and non-required) be managed by the Program Office?

Most, if not all, *external* evaluations should be managed by the Program Office of an OU. Therefore, most *required* evaluations should be managed by the Program Office of an OU. In unusual circumstances, exceptions may be made as decided by Mission management and documented in the Mission's Performance Management Plan. Non-required *internal* evaluations may be managed by the technical offices of an OU, particularly if the internal evaluation is conducted or commissioned by the implementing partner who is being evaluated.

Do other evaluation procedures apply to both required and non-required evaluations?

Yes. Other evaluation procedures and standards for conducting a USAID evaluation, such as the requirements for peer reviews, requirements for documenting and sharing evaluations, the evaluation quality criteria, and branding and graphic standards, apply to both required and non-required evaluations.

EVALUATION RESOURCE

CHOOSING BETWEEN A USAID EXTERNAL OR INTERNAL EVALUATION

The purpose of this guidance document is to provide USAID staff in the Missions and at Headquarters with guidance on choosing whether to conduct an internal or external evaluation. While required evaluations must be external, USAID Missions/Operating Units (OUs) are encouraged to conduct non-required evaluations as needed for management purposes or organizational learning. Non-required evaluations may be *external* or *internal* evaluations (Automated Directives System [ADS] 203.3.1.7).

What are external evaluations?

An external evaluation is one in which USAID—not the implementing partner (IP)—has commissioned a third party to implement an evaluation. As per USAID policy, to be counted as an external evaluation, the team lead must be an independent expert from outside the Agency who has no fiduciary relationship with the IP of the activity or project being evaluated.

External evaluations may include a USAID staff member, but the team leader must be from outside the Agency. An evaluation with a team lead from USAID/Washington is not an external evaluation. An evaluation contracted through a subcontract of the IP is not an external evaluation.

Required evaluations necessitate an external evaluation. They must be led by an external team leader, managed in most cases by Program Office staff (ADS 203.3.1.3). USAID Missions/Offices are expected to devote approximately three percent of total program funding (OU total budgets), on average, to external evaluations (ADS 203.3.1.4).

What are internal evaluations?

Internal evaluations are those that are either: (1) commissioned by USAID in which the evaluation team leader is USAID staff (a USAID internal evaluation); or (2) conducted or commissioned by an IP concerning their own project (an IP internal evaluation).

Concerning the latter, funding may be dedicated within a project design for IPs to engage in evaluative work for their own institutional learning or accountability purposes (ADS 203.3.1.7). In such cases, the IP should discuss plans for an internal evaluation with their Agreement Officer's Representative/Contracting Officer's Representative (AOR/COR) and include it in their Activity Monitoring and Evaluation plan and annual work plan. In addition, the IP should provide the AOR/COR with the evaluation statement of work for review, CVs of the consultants hired or assigned (if any) to conduct the evaluation, and the final evaluation report for feedback and approval.

EVALUATION RESOURCE

What purpose do external evaluations serve?

An external evaluation serves the purpose of validating the standard of *independence*, which aims to mitigate bias and ensure credibility and integrity to the evaluation process. The importance of an evaluator's independence from program management often provides greater credibility of the evaluation findings and report. An external evaluation reduces real or perceived conflict of interest—a situation in which a party has interests that could improperly influence that party's performance of official duties or responsibilities, contractual obligations, or compliance with applicable laws and regulations. A real or perceived conflict of interest of an evaluator translates to a lack of impartiality, objectivity, and integrity and has the potential to jeopardize the credibility and validity of the findings.

In addition to required evaluations, external evaluations at USAID are also recommended for non-required evaluations when the evaluation is expected to primarily serve an accountability purpose for external audiences. In addition, external evaluations at USAID may serve to supplement USAID expertise and provide a valuable outside perspective.

What purpose do internal evaluations serve?

The purposes for planning and implementing internal evaluations include: (1) to benefit from insider expertise and knowledge of program or Agency operations; (2) to better ensure that learning from an evaluation is captured internally, utilized, and institutionalized in the OU or Agency; (3) to develop the capacity of USAID Mission staff in the process of planning and implementing high-quality evaluations; and (4) to more quickly (than possible through the procurement process) answer a specific development question or collect urgently needed data on a project's performance.

When to plan and conduct internal evaluations

There are many reasons that Missions and IPs would choose to conduct an internal evaluation for a non-required evaluation:

- When there is an evaluation trigger as detailed in ADS 203.3.1.3 (such as performance information indicating an unexpected result or a key management decision is required), but the independence of an external evaluation is not needed or desired;
- When an IP wishes to commission an evaluation using an external team or their own staff for learning purposes;
- When there is inadequate funding to commission an external evaluation, but internal capacity and availability exists;
- When there are inadequate mechanisms available to procure an external evaluation on time, but internal capacity and availability exists;
- To provide capacity building, learning, and practical experiences to USAID staff in conducting evaluations so that they will become better commissioners, implementers, and users of evaluation findings; and
- To promote an Agency-wide culture of learning.

When *not* to conduct an internal evaluation

If the project is required to be evaluated, based on it being a large or pilot project, an internal evaluation should not be used. It is also not recommended that high-profile or highly scrutinized projects be selected for an internal evaluation, regardless of whether it is required or not.

EVALUATION RESOURCE

Similarly, if an evaluation is primarily focused on accountability issues for an external audience (rather than internal learning), an internal evaluation is not suggested. It is recommended that in these cases, external evaluators be used, with attention paid to selecting the highest-qualified, experienced, and independent evaluators to ensure quality and to mitigate bias or perceptions of conflict of interest.

Reporting and dissemination of internal evaluations

In cases where project funding from USAID supports an evaluation conducted or commissioned by an IP, the findings from that evaluation must be shared in written form with the responsible AOR/COR or technical officer within three months of the evaluation's conclusion (ADS 203.3.1.7).

Other evaluation procedures and standards—such as the requirements for peer reviews, requirements for documenting and sharing evaluations, evaluation quality criteria, and branding and graphic standards—also apply to internal evaluations. For instance, final evaluation reports for internal evaluations should utilize the USAID Evaluation Report template and must be uploaded onto the Development Experience Clearinghouse within 90 days of approval. Internal evaluations are reported annually in the Evaluation Registry of the OU Performance Plan Report along with external evaluations.

EVALUATION RESOURCE

CONDUCTING AN EVALUABILITY ASSESSMENT FOR USAID EVALUATIONS

Introduction

When asked about barriers to evaluation quality and use, USAID staff members often cite instances when the evaluation took place at the wrong time, focused on the wrong questions, or failed to engage stakeholders from the beginning. Similarly, external evaluation teams can find themselves challenged to collect and analyze data while, at the same time, they must also reconstruct a project's theory of change and adapt to the changing needs of project stakeholders.

To address these barriers, in recent years, USAID has emphasized better evaluation planning, which is reflected in higher standards for evaluation statements of work, as well as renewed CDCS and project design processes that incorporate evaluation planning at the onset. One tool to further strengthen evaluation planning is to take a strategic pause before conducting an evaluation and conduct an evaluability assessment.

What is Evaluability Assessment?

Evaluability assessment is a method for determining:

- The extent to which a project or activity is ready for an evaluation
- The changes that are needed to increase readiness
- The type of evaluation approach most suitable to assess the project or activity's performance and/or impact

While most staff at USAID consider these concepts at least partially when writing an evaluation SOW, an evaluability assessment offers a systematic process for assessing readiness. It can also generate recommendations for necessary changes to the project or activity be implemented before the evaluation takes place.

Evaluability assessment can take many forms depending on the specific context. USAID staff and partners can conduct the assessment internally, or engage an outside consultant or consultant team. If engaging consultants, it is important to keep in mind that their role is to facilitate the process rather than to produce a deliverable independently. In either case, USAID staff should expect to dedicate time and effort to the activity.

Why conduct an Evaluability Assessment?

Even when valid information needs exist and leadership supports conducting an evaluation for learning and accountability, there may be still be outstanding questions regarding whether conducting a planned evaluation is appropriate and feasible. For instance:

EVALUATION RESOURCE

- 1) Will the evaluation findings be provided in time to inform decisions such as exercising option years, designing a follow-on program, creating a country or sector strategic plan, or making a policy decision?
- 2) Is there a demand for the evaluation among potential users
- 3) Is an adequate budget available to fund the evaluation?
- 4) Will it be possible to answer the proposed evaluation questions based on facts, evidence, and data with findings supported by quantitative and qualitative information that is reliable, valid, and generalizable?
- 5) Will benefits of the evaluation exceed the costs of the evaluation (including both monetary costs and burden on staff, implementers, beneficiaries, and stakeholders).

An evaluability assessment can help answer these questions. Evaluability assessment does not replace the need for strong project design, active management, and internal consideration information needs. However, it does present an opportunity for a strategic pause to maximize the usefulness of a proposed evaluation. Reasons to conduct evaluability assessments include:

- 1) Prevent waste of resources on premature or misfocused evaluation
- 2) Guide decisions on the worth and usefulness of evaluation
- 3) Determine if and how project or activity design, implementation, and monitoring should be adjusted to support a more useful evaluation
- 4) Help design an appropriate evaluation
- 5) Build consensus among decision makers

How do I conduct an evaluability assessment?

Step One: Clarify the purpose of the evaluability assessment

Identify the purpose of the evaluability assessment and what it will focus on. Consider:

- Will the assessment examine an entire project or activity, or a subset of geographic regions or sites?
- What specific components of the project or activity and intended results are of greatest interest?

If you are unsure about either of these questions, you can use the evaluability assessment to clarify them. If you are considering collecting data on comparison groups as part of the evaluation, the evaluability assessment could also explore what comparison groups/sites to use.

Step Two: Involve stakeholders and intended users

In addition to USAID staff, a wide variety of stakeholders often have a role in contributing to and using evaluation findings, including implementing partners, host country governments, and other donors. Engaging stakeholders in the evaluability assessment helps to understand the possible demand for the evaluation and generate buy-in for the evaluation.

First, determine their needs and expectations related to the evaluability assessment. This will include 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 RESOURCE

evaluation itself. Use this stage to understand what type of feedback each stakeholder group would find most helpful, and how they would like to receive this feedback.

Based on this information, determine how often and when each individual/group should be involved. For example, you may plan to conduct a stakeholders' workshop or convene a small advisory committee to oversee key decisions for the evaluation, such as what evaluation questions to include in the SOW.

Step Three: Document the project or activity design and clarify intent

Collect information from a variety of sources in order to develop a strong understanding of the project or activity:

- Goals and objectives
- Theory of change
- Resources and organizational inputs
- Target population
- Activities
- Outputs
- Desired short-term and long-term outcomes
- Context

Much of this information is documented in the mission's CDCS, the Project Appraisal Document (PAD), RFPs/RFAs, proposals, implementer workplans, and M&E plans. However, other data collection approaches could include conducting a literature review related to the project theory of change, identifying other evaluations on similar topics or with similar target populations, and brainstorming with project staff and stakeholders to fill gaps in the project or activity design.

Step Four: Document project or activity implementation, data quality, and evaluation capacity

If the project or activity to be evaluated has already started implementation, the evaluability assessment should determine the current status of implementation and explore *how* the project or activity is being implemented to better understand if it is ready for evaluation and, if so, what type of evaluation design would be appropriate. Information collected in this step will help you assess options for more rigorous evaluation designs.

This is the typically the most time-intensive stage of the evaluability assessment. You may use a variety of methods such as document review (including a review of the most recent workplans and quarterly and annual performance reports), key informant interviews, focus groups, and observation. Available performance indicator data should be reviewed to determine if data is being collected and targets are being achieved. During a full evaluation, the details of project or activity implementation will be explored more completely; during the evaluability assessment, you are seeking information on key aspects of "implementation reality" compared to project or activity design that would affect the evaluation, such as: start-up timing and broader roll-out activities; any shifts in the overall project focus and planned activities; notable changes in local political, economic, or security situation; new development partners and new stakeholders; etc.

EVALUATION RESOURCE

In this step, you should also identify and assess the quality of data systems. This includes partner data management as well as information from host country information systems, third-party sources, and any other sources you may use in the evaluation. Recent data quality assessments of relevant performance indicators should be reviewed. Document the strengths and limitations of the data systems to monitor implementation of activities, measure outputs, and measure outcomes. This analysis will help you understand what data is available and to what extent it can be used to answer the questions in your evaluation.

Step Five: Analyze readiness for evaluation and make recommendations

Based on the information collected during the evaluability assessment, analyze the project or activity's readiness for evaluation in the areas listed below.

- **Project or activity Design:** Is it plausible that the project or activity would achieve its results? Consider:
 - Past research/theory
 - Logic of the development hypothesis
 - Level of resources provided
 - Level, consistency, fidelity of implementation of activities
 - Reach of the program
 - Timeframe for implementation
- **Data Availability:** Is there sufficient information available in order to answer key evaluation questions, or can this information be collected? Consider:
 - The outcomes and results of greatest interest to stakeholders
 - Partner monitoring system and capacity to provide data for evaluation
 - Quality of data on key areas of intervention
 - Baseline data availability
 - Costs of new data collection and analysis (financial and time)
- **Context and Environment:** Is the operating environment conducive to conducting the evaluation? Consider:
 - USAID and partner commitment to learning from and using evaluation findings
 - Timing due to political events, seasonality, and other factors.
 - Resources to complete the evaluation (financial, staff, equipment, etc.)

Using this analysis, revisit the initial evaluation plans. If the evaluability assessment suggests that the project or activity is ready for evaluation as planned, the information from the evaluability assessment should be used to refine the evaluation statement of work and evaluation design.

If the evaluability assessment suggests that the project or activity is not ready to be evaluated as planned, you may (1) consider alternative evaluation approaches or (2) develop recommendations based on the evaluability assessment to increase project or activity readiness. For example, the evaluability assessment may determine it is too early to see significant quantitative results of project end-outcomes suggesting that it is not appropriate to

EVALUATION RESOURCE

focus an evaluation on the achievement of those outcomes. However, it may be appropriate to launch an evaluation to explore intermediate results through quantitative and qualitative methods. Or, the evaluability assessment may determine that you cannot feasibly measure all desired outcomes in the evaluation, but the evaluation could focus on a subset that stakeholders are most interested in.

Recommendations to increase project or activity readiness may include:

- Strengthen the project or activity design. Revise the project logic to establish a theory of change that is clear, measurable, realistic, and agreed-upon.
- Improve implementation: Modify project activities to have a greater chance of reaching the stated objectives.
- Improve project or activity management: Apply corrective management actions to improve fidelity to project design and smooth operations
- Improve measurement: Develop or improve data systems for project and activity monitoring.
- Strengthen commitment to evaluation: Generate common understanding of the evaluation's purpose and use among stakeholders.

Acknowledgments and References

Materials for this discussion note were drawn from Dr. Deborah Rog's course on Evaluability Assessment at [The Evaluator's Institute](http://theevaluatorinstitute.org), as well as:

Design, Monitoring and Evaluation for Peace. Evaluability Assessments in Peacebuilding Programs. January 15, 2015. <http://dmeforpeace.org/learn/evaluability-assessments-peacebuilding-programming>

DFID. Planning Evaluability Assessments: A synthesis of the literature with recommendations. Working Paper 40: October 2013. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/248656/wp40-planning-eval-assessments.pdf

Leviton, L.C, Kettel-Khan, L, Rog, D.J., Dawkins, N., and Cotton, D. (2010). Evaluability assessment. Annual Review of Public Health, 31, 215-234.

Thurston, W. E. & Potvin, L. (2003). Evaluability assessment: A tool for incorporating evaluation in social change programmes. Evaluation, 9(4), 453-469.

Trevisan, M. S., & Huang, Y. (2003). Evaluability assessment: A primer. Practical Assessment, Research and Evaluation, 8(20),

PROGRAM CYCLE RESOURCE

EVALUATION STAKEHOLDER ANALYSIS

What: Evaluation Stakeholder Analysis (SA) is a technique used to identify and assess the importance of key people, groups, or institutions in the evaluation. It is a systematic way to establish stakeholder interests and needs and generates information that is critical to planning and implementing your evaluation.

Why: Stakeholder analysis helps you anticipate the influence different stakeholders may have on mobilizing support or resistance to the evaluation; as well as the use of evaluation findings and implementation of recommendations. It also helps you provide a foundation and strategy for participation in the evaluation - which often leads to more useful, cost-effective evaluations!

How: Conducting an evaluation stakeholder analysis is a tiered process. A general guideline is below:

1. Identify the main users of the evaluation. USAID is often one of the main users, but there are others. They may include partner country institutions, implementing partner organizations, or other donors. Ask what these users really need to know about the program and how they envision using evaluation findings. This step will help to identify what information you need from the evaluation and how to get it.
2. Identify clients and other key stakeholders. This group is likely larger than the group you identified in Step 1. A broader group of stakeholders are those who stand to be interested in and/or impacted by the evaluation, both positively and negatively. Make sure to include marginalized groups, if applicable. This group of stakeholders has important interests, but will have less voice in the evaluation.
3. Provide early opportunity for each of these groups to raise issues for the evaluation to address.
4. Assess stakeholder interests, influence, importance, and potential impact on the evaluation. Use the attached Evaluation Stakeholder Analysis Worksheet to determine the extent to which each individual or group plays a role, or could play a role, in the evaluation's planning, implementation, and use. In doing so, consider the relationships between stakeholders, agreeing or conflicting interests, and both short and long-term implications of the evaluation.
5. Develop strategies to appropriately involve stakeholders in the evaluation, as suitable for the context and needs of the evaluation. Even if specific groups or individuals are not explicitly involved, it is still important that their interests and needs have been taken into account.
6. Consider offering periodic updates or briefings to stakeholders during the process of the evaluation. This may help to improve buy-in and eventual use of evaluation findings.
7. Ensure that your evaluation report indicates the nature and extent of stakeholder involvement.

PROGRAM CYCLE RESOURCE

Evaluation Stakeholder Analysis

Evaluation stakeholders	Make Policy	Make operational decisions	Make inputs to evaluation (describe)	React to findings	For interest only	Proponent of the evaluation (describe why)	Opponent of the evaluation (describe why)

EVALUATION RESOURCE

DEVELOPING AN EVALUATION DISSEMINATION PLAN

Transparency is a key practice of evaluation at USAID. As noted in USAID's [Evaluation Policy](#):

“[F]indings from evaluations will be shared as widely as possible, with a commitment to full and active disclosure,” and “The presumption of openness in the conduct of USAID evaluations will be manifested at two stages: (1) when an evaluation design is agreed upon; and (2) when the evaluation report is completed, that report will be disseminated.”

The Policy further stipulates that USAID evaluations of all types will include *a dissemination plan*.

A dissemination plan is not difficult to prepare and helps ensure that evaluations are useful and shared effectively. This guidance document and the associated template are intended to assist USAID Missions and Bureaus in developing a dissemination plan that satisfies the Policy and USAID's interest in transparency, accountability, and learning. USAID Missions and Bureaus may use the dissemination plan template as a guide to formulating and drafting dissemination plans.

Timing: 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 statement of work (SOW). This way, dissemination activities can be written into the evaluation team's SOW and appropriately resourced and budgeted. As work progresses, the dissemination plan may be updated or revised as needed based on new information or learning.

Components: The first step in developing an evaluation dissemination plan is identifying stakeholders. With input from the program office, technical specialists, and implementing partners, identify the groups and/or individuals who are likely to be affected by or interested in the evaluation results. A smaller group of stakeholders (perhaps including the host government counterpart, key beneficiary groups, or local research organizations) will be consulted during the design phase. Decide on who will be included in this group during the stakeholder identification stage. Additional tools are available for stakeholder analysis.

For each identified stakeholder, think through USAID's goal in disseminating the evaluation results. Are we aiming to change policy? Influence the design of other projects or activities? Contribute to the technical knowledge base? Satisfy accountability concerns? Attract a new partnership? Prevent repetition of mistakes?

Articulating a goal for dissemination assists with the next piece of the strategy—identifying the appropriate communication tools. These may simply include reports, briefs, blog posts, press releases, graphics, and e-mails to listservs. They may also include more involved efforts including presentations, meetings, facilitated workshops or discussions, videos, and journal

EVALUATION RESOURCE

articles. Again, for the identified stakeholders, consider how they receive information and if there are existing forums or channels through which to reach them. Factors to consider include:

- Literacy level and native language
- Access to the Internet, radio, or other sources of media
- Existing websites, journals, or email listservs where dissemination would be appropriate
- Planned events or decision-making points
- Resources available for design and editing

The evaluation manager should identify a responsible party and timeline for each dissemination activity. In some cases, the development and dissemination of knowledge products from an evaluation would be included in the SOW for the evaluator(s). In others, USAID or another stakeholder should take the lead. For the timeline, consider external factors—such as the political cycle or project design decision points—that may influence your communications objectives, and build in time for copyediting, design, and translation, if necessary.

Finally, an evaluation dissemination plan may include products that are designed to monitor and document the impact of the evaluation and the associated knowledge products. These may include, for example, a follow-up survey or feedback forms at events.

EVALUATION RESOURCE

TIPS FOR DEVELOPING GOOD EVALUATION QUESTIONS (FOR PERFORMANCE EVALUATIONS)

I. Performance evaluations typically focus on *descriptive* and *normative* questions

- What a particular project or program has achieved (either at an intermediate point in execution or at the conclusion of an implementation period);
- How it is being implemented;
- How it is perceived and valued;
- Whether expected results are occurring;
- Other questions that are pertinent to program design, management, and operational decision-making.

Some helpful definitions:

Descriptive Questions: Seek to determine what is or was

Normative Questions: Compare what is with what should be

Cause and Effect Questions¹: Determine causal connection between the intervention and outcomes or make other causal inferences

II. Principles of a good evaluation question

Principle 1: A good evaluation question should be both a question and evaluative.

- *Tip #1: A question for a sector assessment or a needs assessment **is not** an evaluation question.*

Sector assessments or needs assessments tend to look at an entire sector or population to identify problems or needs and potential interventions to address them. Evaluations focus on what USAID's projects or programs did in that sector. You may choose to combine an evaluation with a sector assessment, but the statement of work (SOW) should be clear regarding which questions are evaluation questions and which are sector assessment questions.

- *Tip #2: A request for a recommendation **is not** an evaluation question.*

Recommendations might be important, but asking what you should do in the future (sometimes called a prospective evaluation question) should be based on what you can observe from the past. Your questions should ask the evaluators to understand something about what happened in the past of the project being evaluated. You can always request a recommendation from your evaluators regarding future options, but such requests should be distinguished from your evaluation questions and, ideally, linked to the evaluation question that will inform the recommendation.

¹ It should be noted that this type of question is not mentioned in the definition of performance evaluations at USAID, and we often think of impact evaluations as addressing cause and effect questions. However, in certain circumstances, performance evaluations can address (and in practice, typically do address) cause and effect questions; however, they don't use the experimental logic of an impact evaluation, but rather, use other evidence-based methods and arguments for causal inference, such as process tracing or contribution analysis.

EVALUATION RESOURCE

Principle 2: A good evaluation question should be limited (in scope)

- *Tip #3: No more than five evaluation questions per SOW.*

Even fewer than five questions is a good idea. Moreover, five questions doesn't mean including five main questions followed by fifty sub-questions. Keep the sub-questions limited in number and germane to the main question. Nor does this mean writing one question that is actually several questions wrapped into one. Focus on what is most important to know.

- *Tip #4: Evaluation questions need not address every aspect of the activity/project/program; instead, address specific issues where you need further information.*

It's not just the number of questions you ask, but the *scope* of the questions you ask. It is impossible for an evaluator to tell you everything about your project in one evaluation. USAID projects and activities often have multiple objectives, a variety of tasks, different sets of beneficiaries, etc. Focus on the most critical issues where you need evidence-based confirmation or additional information to help you to make decisions about current or future programming. The more you limit the scope of your questions, the more likely they will be understood by your evaluator and the more likely your evaluators will be able to answer them in-depth.

Principle 3: A good evaluation question should be clear

- *Tip #5: Each word in the evaluation question should be clearly defined. Be especially careful about important (but ambiguous) terms, such as "effective," "sustainable," "efficient," "relevant," "objective," and "success."*

For instance, if you ask if a project is "effective," does that mean you are asking if monitoring targets were reached or that the intervention had a causal effect on beneficiary standards of living, or simply that stakeholders perceived it to be successful? You have to define terms. Otherwise, the evaluator will define the term for you in ways that are neither transparent nor match your understanding of the terms.

- *Tip #6: Include additional narrative along with the evaluation question to provide context and/or define your terms.*

Evaluation questions do not need to be simply listed in the SOW, one question after the other. Provide some additional narrative to the evaluation questions—descriptions as to why you are asking the question or what you mean by some terms. Make sure the narrative provides explanation relevant to the question and does not list additional questions to answer.

Principle 4: A good evaluation question should be researchable

- *Tip #7: To be researchable, there must be a way to generate objective evidence to answer the evaluation question with social science methods that will likely be applied based on the methodology section of the SOW and evaluation resources.*

Even if a question is limited and clear, it might not be answerable with the resources you have available. Consider the methods that are likely to be applied to answer the question and whether empirical data could be collected by those methods that would objectively answer the question. Don't ask an impact evaluation question about measuring the project's effect on beneficiary income when you are only planning to do stakeholder interviews. Don't ask about stakeholder perceptions of the project if you are not planning to do stakeholder interviews.

EVALUATION RESOURCE

- *Tip #8: If you ask a normative question, it is only researchable if clear, measurable standards or criteria can be identified.*

If you ask normative questions, you are asking the evaluator to make a judgment based on the evidence collected. In order to make a judgment that is objective and verifiable, agreed-upon standards or criteria as the basis for the evaluator judgment are required. You may choose to state these in the SOW or request them from your evaluators. If you are going to ask if a project is, for example, effective, efficient, or sustainable, you need to have specific criteria for effectiveness, efficiency, or sustainability for this particular evaluation. Those criteria should be explicit. Don't start the data collection until the evaluators and the evaluation manager are both clear on the criteria that will be used to judge the evidence in answering a normative question.

Principle 5: A good evaluation question should be useful

- *Tip #9: Link your evaluation questions to the evaluation purpose (but don't make your purpose another evaluation question).*

The evaluation purpose should provide the learning or accountability framework for developing specific evaluation questions. It tells the evaluators why you are doing the evaluation and what audience the evaluation serves. But the purpose section of the evaluation SOW is not the place to state evaluation questions or to guide the evaluators as to what you expect them to do. Keep the purpose section of the evaluation SOW short. Avoid a scenario where the evaluation team is trying to address conflicting guidance in the purpose section of the report and the evaluation questions section of the report.

- *Tip #10: Involve stakeholders in developing questions.*

Consider the intended audiences of the evaluation and what will be useful for these audiences. Consider the interests of those beyond the primary USAID audience—e.g., implementers, government officials, funders, and beneficiaries. They may have good ideas about what the evaluation should address, and involving them early can increase the likelihood that the evaluation will be used.

III. Examples of going from *bad* to *better* evaluation questions:

Example 1: Questions from an evaluation of an HIV service provider support project

From:

"To what extent is the project **relevant?**"

To:

- Is the training and technical support to HIV service providers being delivered as intended according to project design?
- Does the training and technical support to HIV service providers meet the needs and priorities of project stakeholders?
- What are the financial and organizational characteristics, organizational mission, and coverage area of HIV service providers who have received the project training and technical assistance?
- Have the appropriate (as defined in project design documents) HIV service providers received the project training and technical support?

EVALUATION RESOURCE

Example 2: Questions from an evaluation of a youth employment project

From:

“To what extent is the project **effective** in meeting its objectives?”

To:

- To what extent did the intended outcome of increasing youth employment in targeted regions occur over the course of the project? Did employment outcomes differ by region or gender?
- Did the project meet its targets in training youth in employable skills? Why or why not?
- Are key stakeholders satisfied with the performance of the implementer in training youth in appropriate skills? Why or why not?

Example 3: Questions from an evaluation of a municipal capacity development project

From:

“To what extent is the activity **efficient**?”

To:

- How much did it cost to provide each municipality with budgeting software and software training? How does the cost per municipality compare to other similar projects?
- Did the project provide the software and training to the appropriate number of municipalities and individuals on time per the workplan?
- How quickly did the project respond to requests from municipalities with software installation and training? Were municipal stakeholders satisfied with the speed of the response?



HOW-TO NOTE

Evaluation Statements of Work

Monitoring and Evaluation Series

Evaluation statements of work should clearly communicate why the evaluation is needed, how it will be used, and what evaluation questions will provide answers managers need.

How-To Notes

are published by the Bureau for Policy, Planning and Learning and provide guidelines and practical advice to USAID staff and partners related to the Program Cycle. This How-To Note supplements USAID ADS Chapter 203.

INTRODUCTION

This How-To Note addresses key issues for USAID staff who are developing a Statement of Work for an externally contracted evaluation. Following these practices will help to establish clear expectations and requirements for the evaluation team. These practices also serve as a guide for reviewing the quality of evaluation statements of work for internal peer review processes. While the information in this Note is applicable to both performance and impact evaluation, the complexity of an impact evaluation and multi-stage evaluations will be addressed with additional guidance for SOW development.

BACKGROUND

An Evaluation Statement of Work contains the information that those who conduct the evaluation need to know:

- purpose of the evaluation and how it will be used;
- background and history of the activities, projects, or programs being evaluated;
- questions that must be answered, and how they might be answered;
- expected deliverables;
- expertise needed to do the job; and
- time frame and budget available to support the task.

KEY ISSUES IN PREPARING THE SOW

INTERRELATED ELEMENTS

Drafters need to ensure that evaluation questions are consistent with the evaluation purpose, that the evaluation methods are appropriate for answering the evaluation questions, and the evaluation team members have the requisite skills to employ the proposed evaluation methods.

STRIKING A BALANCE

A SOW must balance the number and complexity of the evaluation questions with the time allotted to conduct the evaluation and the availability of funding. Finding the appropriate balance often requires an iterative process in which the drafter revisits, and sometimes adjusts, each of the elements of the SOW.

FLEXIBILITY

There will always be unanticipated problems and opportunities that emerge during an evaluation. It is helpful to build in flexibility to the SOW, particularly in the methodology section, to accommodate ideas from the evaluation team and necessary changes during the evaluation process.

ADEQUATE TIME

The drafters of the SOW are, in essence, the architects of the evaluation. It is important to commit adequate time and energy to the task, including time to gather and analyze information, build productive relationships with stakeholders, and incorporate stakeholder feedback into the SOW. It is recommended that the drafters draw on an evaluation expert when drafting an SOW.

STAKEHOLDER INVOLVEMENT

Ensuring SOWs are of the highest quality and adhere to the standards of the USAID Evaluation

Policy requires collaboration between the program office and technical offices of a mission. Each mission's Evaluation Mission Order should specify roles and responsibilities for the preparation of an evaluation SOW. Typically, but not always, evaluation questions originate from the technical offices, while program offices (as the eventual managers of most evaluations) finalize the SOW.

Program offices take the lead in ensuring that final SOWs for external evaluations adhere to the standards in the Evaluation Policy and organize in-house peer reviews to assess quality of evaluation SOWs, engaging regional and technical bureaus as needed. Technical offices should participate in the peer reviews.

USAID encourages participation by national counterparts and country-level stakeholders in the development of evaluation SOWs. Stakeholders may encompass a wide array of people and institutions, including policy makers, program managers, implementing partners, other relevant US government agencies, host country organizations, and beneficiaries. Involvement by stakeholders can both improve the list of questions to be answered as well as increase acceptance of the evaluation purpose and process, leading to increased utilization of the evaluation findings, conclusions, and recommendations.

THE SOW IN THE EVALUATION PROCESS

Initial planning for an evaluation should long precede the drafting of the SOW. Preparation of the SOW itself should begin at least six months prior to the planned award date to allow time for various actions needed to complete a quality SOW and award the contract. Steps in procuring an external evaluation:

1. Reviewing and preparing background material about the activity/project/program to be evaluated;
2. Determining the appropriate evaluation questions, suggested methods, and evaluator qualifications;
3. Drafting the SOW;
4. Preparing a budget and independent government estimate;
5. Choosing a mechanism;
6. Conducting an in-house peer review of the SOW;
7. Sharing the SOW with relevant stakeholders;
8. Revising the SOW based on the peer review and stakeholder feedback;
9. Submitting to OAA for approval and RFP preparation;
10. Proposal preparation and submission by external evaluators; and
11. Selection of the evaluation team and award.

THE ELEMENTS OF A WELL WRITTEN EVALUATION SOW

DESCRIBE WHAT IS TO BE EVALUATED

Evaluations can focus on activities, projects, or programs being implemented within a single country or across multiple countries. In some instances the focus of an evaluation may be a single innovative intervention within a project. An SOW introduces the activities, projects, or programs that USAID wishes to evaluate, stating the title(s), start and end dates, funding levels, funding sources (e.g., mission, regional office, or Washington accounts), implementing partners, and sectors or topics.

PROVIDE BACKGROUND AND PERFORMANCE INFORMATION

This section of the SOW should give a description of the context, history, and current status of the activities, projects, or programs, and other information to help the evaluation team understand background and performance. State in detail the specific problem or opportunity the activity, project, or program was designed to address. State the development hypothesis(es) and clearly describe the theory of change that underlies the project or program's design. Include the CDCS results framework and project design logical framework. If the evaluated project or program operates in particular geographic areas of a country and/or with particular target groups, these should be identified as well. Maps are highly recommended.

Specify what project documents will be available to evaluators. In particular, identify the existence and availability of relevant performance information sources, such as performance monitoring indicators and/or previous evaluation reports. Including a summary of the types of data available, the timeframe, and an indication of their quality and reliability will help the evaluation team to build on what is already available.

STATE THE PURPOSE, AUDIENCE, AND USE OF THE EVALUATION

A good SOW states why an evaluation is being conducted (the purpose), who will use the results of the evaluation, and how they will do so. In

SOW ELEMENTS AT A GLANCE

1. Description of Program/Project/Activity to be Evaluated
2. Background
3. Purpose
4. Questions
5. Methods
6. Deliverables and Timeline
7. Team Composition
8. Scheduling and Logistics
9. Budget

general, evaluations at USAID have two primary purposes: accountability to stakeholders and learning to improve effectiveness. In this section of the SOW, though, drafters should provide a more specific purpose that explicitly links the evaluation to future decisions to be made by USAID leadership, partner governments, and/or other key stakeholders. The clearer the purpose, the more likely it is that the evaluation will produce credible and useful findings, conclusions, and recommendations that can be used to achieve greater effectiveness and results.

The purpose of the evaluation should be consistent with, but not replicate the evaluation questions. The purpose should also be consistent with the timeframe of the evaluation in relation to the project or program's life cycle. For instance, an evaluation whose main purpose is to inform a follow-on project should ensure that the evaluation will be conducted after the project has generated enough evidence to inform the follow-on project, but prior to the main design decisions for the follow-on.

IDENTIFY THE EVALUATION QUESTIONS

The core element of an evaluation SOW is the list of questions posed for the evaluation. These questions should be aligned with the evaluation's purpose and expected use. Each question should be answerable using the best methods appropriate

to the questions to generate the highest quality and most credible evidence possible.

One of the most common problems with evaluation SOWs is that they contain a long list of poorly defined or “difficult to answer” questions given the time, budget, and resources provided. While a participatory process ensures wide ranging input into the initial list of questions, it is equally important to reduce this list to a limited, manageable number of key questions, generally between three and five questions. Keep only those questions of essential importance to the evaluation purpose where USAID is willing to provide the management commitment, time, and budget resources. Not every aspect of a project or program needs to be, or should be, the focus of the evaluation.

While keeping the number of evaluation questions limited is important, the content of those questions is equally if not more important. Questions should be precise in what is being asked. Vague terms which can be defined or applied in a variety of ways, such as “relevance” and “effectiveness,” should be clearly defined in this section or in the methodology section of the SOW. Questions should also be researchable, that is, they should have an answer that can be obtained through the use of social science methods and tools rather than evaluator specific judgments.

In addition to specifying which questions the team should address, a well-written SOW indicates the priority assigned to each evaluation question. An SOW can accomplish this by arranging questions in order of importance (and stating that it has done so) or it can estimate the likely level of effort expected to be invested in answering each question.

Finally, the evaluation questions section of the SOW should identify all evaluation questions for which gender-disaggregated data are expected and questions for which an examination of gender specific effects of the activity, project, or program are expected.

IDENTIFY METHODOLOGICAL APPROACH

This section of the SOW should clarify any expectations the drafter of the SOW may have with respect to the evaluation’s design and methodology. As noted in the Evaluation Policy, the methodology should “generate the highest quality and most credible evidence that corresponds to the evaluation questions being asked.” Drafters should also consider the purpose and resources when determining the level of rigor required.

USAID evaluation SOWs vary considerably in the degree to which they prescribe an evaluation design and methodology. At minimum, the SOW should state whether the evaluation will be a performance evaluation or an impact evaluation as defined in the Evaluation Policy. Preferably, it should include some suggestions about the design and methods to be used or the overall methodological approach, while also soliciting the evaluator’s input on what might be most appropriate. The details of illustrative methods can be worked out workplaning with the evaluation team.

Regardless of the specificity of the suggested methodological approach or design, it is helpful for the SOW to link the suggested methods to the specific questions that each data collection method will be used to answer. For instance, if a beneficiary survey and focus group are two of the methods suggested in this section, it should be clear which questions will be answered using either, both, or neither of these methods. Even the most basic methodological suggestions can communicate to evaluators what USAID is expecting regarding the type and strength of evidence for answering each evaluation question. The suggested evaluation methods should be consistent with the type of evaluation question asked and will ideally include more rigorous methods than simple key informant interviews.

In addition to the data collection methodology, a well-written SOW communicates any expectations regarding how evaluation data will be analyzed.

If the evaluation questions ask the evaluators to make judgments from the evidence to be gathered about the evaluated activity, project, or program's overall "effectiveness," "relevance," "efficiency," etc., then this section should suggest criteria for making such judgments or request appropriate criteria from the evaluators. The analysis methods section should also note where analysis of gender, age, or other relevant aspects of beneficiaries are needed.

SPECIFY DELIVERABLES AND TIMELINE

The SOW must specify the products, the time frame, and the content of each deliverable that evaluators are required to complete in the evaluation contract. At minimum, required deliverables should include an evaluation design, draft evaluation report, final evaluation report, and evaluation data.

As noted in the Evaluation Policy, the written design should include identification of key questions, methods, main features of data collection instruments, and data analysis plans. This design will be shared with country-level stakeholders as well as with implementing partners before being finalized. Requiring a draft evaluation report will enable the mission to provide feedback following the peer review process, prior to the submission of the final report. The SOW should define specific expectations for the final evaluation report. For information regarding what should be included in a final evaluation report see the How-To Note "Preparing Evaluation Reports."

The SOW should request all evaluation data to be provided at the end of the evaluation. Moreover, all quantitative data collected by the evaluation team should be provided in an electronic file in an easy to read format; organized and fully documented for use by those not familiar with the project or evaluation.

Any number of additional deliverables may also be requested. These may include: an evaluation work plan, an in-brief with USAID or other stakeholders, an initial document review, progress reports,

photographs of activity sites, an out-brief with USAID or other stakeholders, etc.

A good SOW also specifies the timeline for submission of deliverables, languages of the final report and/or executive summary, maximum or expected number of pages, how the report should be submitted (electronic and/or hard copy), and the number of hard copies requested.

Formatting and branding requirements should also be specified (see How-To Note on Evaluation Reports).

Finally, the SOW should note that the evaluation reports will be reviewed against the Evaluation Policy's "Criteria to Ensure the Quality of the Evaluation Report" as described in Appendix I of the USAID Evaluation Policy.

CLARIFY THE COMPOSITION OF THE EVALUATION TEAM

A good evaluation SOW describes the intended size of an evaluation team, the roles and responsibilities of team members, and the specific qualifications that the team members possess. These skills may include evaluation or methodological expertise, regional or country experience, language skills, management skills, experience working with USAID, technical subject matter expertise, etc. As noted in the Evaluation Policy, teams should include "appropriate methodological subject matter expertise to conduct an excellent evaluation." Team leaders should be "an outside expert with appropriate skills and experience." At least one team member should be an evaluation specialist and all team members should be familiar with the USAID Evaluation Policy.

USAID encourages evaluation specialists from partner countries to lead or participate in evaluation teams. Where appropriate, USAID staff and/or implementing partners may also participate in the evaluation team and the SOW should describe the intended roles of any participating staff. This section should also note that all team members will be required to provide a signed

statement attesting that they have no conflict of interest, or describing an existing conflict of interest.

ADDRESS SCHEDULING, LOGISTICS, AND OTHER SUPPORT

An SOW provides information to potential evaluators on any scheduling constraints that could affect the evaluation. It states the expected period of performance, identifying any specific dates that need to be incorporated in the evaluation plan. Good scheduling and effective local support contributes greatly to the efficiency of the evaluation team. For evaluations involving complex designs and/or survey research data collection methods, the schedule must allow enough time, for example, to develop sample frames, prepare and pretest survey instruments, training interviewers, and analyze data. In some cases, an advance trip to the field by the team leader and/or methodology expert may be justified where extensive pretesting and revision of instruments is required or when preparing for an evaluation in difficult or complex operational environments.

An SOW also outlines the specific kinds of support USAID will provide, along with any additional logistical roles or responsibilities that it expects the team to fulfill. If the SOW requires the team to make site visits to distant or difficult locations, such planning must be incorporated into the SOW.

Budget considerations have to be part of the decision making process for developing the SOW from the beginning. The proposed evaluation questions, methods, timeframe, and expertise required must be balanced against each other and the budget limitations. The calculation of the independent government estimate of the evaluation budget is beyond the scope of this How-To Note. A key aspect of the estimated budget, though, is the level of effort required by the evaluators. A good SOW should include illustrative information about the level of effort expected, preferably in the form of a matrix that displays team member days allotted by evaluation task on a notional basis.

ADDITIONAL RESOURCES

The following resources provide more information. Some other resources exist but are out-of-date with current USAID guidance. Where information differs, the USAID Evaluation Policy and the USAID ADS (Automated Directives System) take precedence over that in other resources.

- USAID ADS Chapter 203: Assessing and Learning : <http://transition.usaid.gov/policy/ads/200/203.pdf>
- USAID Evaluation Policy, January 2011. (http://pdf.usaid.gov/pdf_docs/PDACQ800.pdf).
- Evaluation Statements of Work: Good Practice Examples, August 2011: (<http://transition.usaid.gov/policy/evalweb/documents/EvaluationSOW-GoodPracticeExamples.pdf>).
- How-to Note: Preparing Evaluation Reports, July 2012: <http://transition.usaid.gov/evaluation/HowtoNote-PreparingEvaluationReports.pdf>



STATEMENT OF WORK

[Insert type of evaluation, e.g. Performance or Impact evaluation]

OF

[Insert activity/project/program name]

I. PURPOSE OF THE EVALUATION

Instructions: Insert why the evaluation is being conducted (the purpose), who will use the results of the evaluation, and how they will use it. Explicitly link the evaluation to future decisions to be made by USAID leadership, partner governments, and/or other key stakeholders. The clearer the purpose, the more likely the evaluation will produce credible and useful findings, conclusions, and recommendations. The purpose of the evaluation should be consistent with, but not replicate, the evaluation questions (Section IV).

Note: The Evaluation Purpose will often be picked up by the Contracting Officer and added to the contract that is executed. It is included first in this template for that reason. It can also come after the Background Section.

II. SUMMARY INFORMATION

Instructions: Utilize this section to describe the activity/project/program being evaluated. There are two suggested formats.

Option 1: For activities, projects or programs with one implementing partner

Activity/Project Name	[Insert name of activity/project/program being evaluated]
Implementer	[Insert prime implementing partner]
Cooperative Agreement/Contract #	[Insert Agreement or Contract #]
Total Estimated Ceiling of the Evaluated Project/Activity(TEC)	[Insert total estimated cost]
Life of Project/Activity	[Insert start month/year and end month/year, e.g., April 2011–May 2014]
Active Geographic Regions	[Insert geographic regions, particularly if there are specific geographic areas you would like to focus on]
Development Objective(s) (DOs)	[Insert number and name of the DO that this evaluation relates to]
USAID Office	[Insert the name of the office in the Mission or Washington OU]

Option 2: For projects/programs with multiple implementing partners, including for a sector or thematic evaluation

Activity Name	USAID Office	Implementer	Cooperative Agreement/ Contract #	TEC	Life of Project / Activity	Active Geographic Regions	Mission DO	Required? Public or internal?

STATEMENT OF WORK TEMPLATE

III. BACKGROUND

Instructions: Provide a detailed description of the **context, history, goals and objectives, current status of the activity/project/program**, and other relevant information to help the evaluation team understand the design and implementation plan. Complete the sections noted below. Sections can be consolidated.

A. Description of the Problem, Development Hypothesis(es), and Theory of Change

Instructions: Include details on:

- The specific problem or opportunity the activity/project/program to be evaluated was designed to address;
- The development hypothesis(es) often expressed as an if/then statement¹;
- The theory of change that underlies the design (including a list of the **intended results** and **critical assumptions**);

B. Results Frameworks

Instructions: Include here or as an annex the graphic of the **Mission's Results Framework** and the **Project's Logical Framework** (if applicable) highlighting the elements to be evaluated. If the evaluation is at the Activity level then include the **Activity's Logical Framework** (and linkages to the project-level).

C. Summary Activity/Project/Program to be evaluated

Instructions: Summarize the **primary interventions or tasks** implemented by the activity/project/program. Also include a summary of any substantive changes (modifications) in the evaluated activity/project/program and when they were effective. Describe the specific geographic areas in which the activity/project/program operates and/or targeted groups, as applicable. Attach maps if available.

D. Summary of the Activity/Project M&E Plan

Instructions: Specify what relevant documents will be available to the evaluators. In particular, identify the existence and availability of **relevant performance information sources**, such as performance monitoring indicators and/or previous evaluation reports. In addition, identify any other documents or sources of information from outside of USAID that would be useful to the evaluation team (e.g., government or international data). If this section is long it may also be included in an annex]

¹ If the design document does not contain an implicit development hypothesis, consult with the DO Team to develop the development hypothesis.

IV. EVALUATION QUESTIONS

Instructions: Include **3–5 specific questions** focused on key program areas and/or performance and **directly linked to the purpose of the evaluation and its expected use**. Sub-questions may be included to elaborate on the main question, but not to add new areas of inquiry.

NOTE: Not every aspect of an activity, project, or program needs to be, or should be, the focus of the evaluation. Rather, the evaluation should examine specific aspects of the activity, project, or program where there are questions unanswered by performance monitoring or other data.

Guidelines:

1. **Questions should be precise.** Vague terms that can be defined or applied in a variety of ways (such as “relevance,” “effectiveness,” etc.) should be defined clearly for the evaluator. If any specific terminology or standards are included in the evaluation questions indicate the source or definitions.
2. **Questions should be researchable.** Questions should have an answer that can be obtained through the use of social science methods and tools (qualitative and quantitative) rather than relying on the evaluators’ judgments.
3. **Questions should integrate gender.** Questions should identify when sex-disaggregated data are expected. Where appropriate, the evaluation questions can include a separate question aimed at evaluating the gender-specific effects of the activity or project. [See the [How-To Note on Engendering Evaluation](#)]
4. **Questions should be presented in order of priority**, or the priority of questions should otherwise be identified.
5. **A request for recommendations is not an evaluation question.** If you want the evaluators to provide recommendations, describe what aspects of the activity, project, or program you want recommendations to address in a separate paragraph or following the questions.

V. EVALUATION DESIGN AND METHODOLOGY

Instructions: This section may include suggestions or illustrative descriptions about the methodological approaches. If the evaluation design team has depth of experience in methodologies and methods then this section may be quite detailed and include methodological suggestions. Otherwise, it may request the evaluators’ expertise and input in the proposal and during the evaluation design phase. At a minimum this section should confirm that it is a performance evaluation.

Guidelines: When drafting this section consider and then include narrative that describes clearly:

1. The suggested or expected data collection methodology and the corresponding data sources that will generate the highest-quality and most credible evidence that corresponds to the evaluation purpose and questions.
2. How suggested methods are linked to at least one evaluation question.
3. Any expectations regarding sites to be visited or groups to be interviewed.
4. Any expectations regarding how the evaluation data collected should be analyzed (e.g., comparison of particular groups or precision of response criteria, such as “margin of error must be less than +/- 10 percent”).
5. If performance monitoring data are to be used. If so, include information about how they have been verified. Or if the data have not been verified, that it is the expectation that the proposed design should include this requirement.
6. When analysis of disaggregated data are needed (e.g., sex, age, or other relevant aspects of beneficiaries).
7. Any known limitations to the data to be collected.

STATEMENT OF WORK TEMPLATE

The following simple design matrix can be included as a summary of evaluation design and methods, and to supplement the narrative section above, but should not replace the narrative.

Questions	Suggested Data Sources (*)	Suggested Data Collection Methods	Data Analysis Methods
1. [Insert Evaluation question]	[Documents (including. performance monitoring data, previous evaluations, etc.), national statistics, project staff, stakeholders, expert knowledge, beneficiaries...]	[Key informant interviews, questionnaires or surveys, focus group discussions, direct observation, desk review...]	[To be determined by evaluation team] [Requested level of disaggregation—gender, ethnicity, location (district, province), etc....]
2. [Insert Evaluation question]	ditto	ditto	ditto
3. [Insert Evaluation question]	ditto	ditto	ditto

Notes: (*) It is acceptable to include data sources that do not need to be collected but may be analyzed by the evaluation team. In planning for and preparing the Evaluation SOW it is a good practice to examine available data sources especially performance monitoring data.

VI. DELIVERABLES AND REPORTING REQUIREMENTS

Instructions: List specific deliverables, reporting requirements, audiences, and timeframes that the evaluation team should know. The only required deliverables are the evaluation design, draft report, and final report, but additional deliverables may be beneficial. Sample text is provided below to be adapted as relevant and useful to your Operating Unit.

Please consider the time and location of when the evaluation team can reasonably complete the deliverable. For example, preparation of the draft report requires analysis of the data collected; therefore, the exit-briefing for an international team will likely not be able to include requirements for presentation of recommendations.]

- Evaluation Work plan:** [SUGGESTED] Within [# weeks] of the award of the contract, a draft work plan for the evaluation shall be completed by the lead evaluator and presented to the Agreement Officer's Representative/Contracting Officer's Representative (AOR/COR). The work plan will include: (1) the anticipated schedule and logistical arrangements; and (2) a list of the members of the evaluation team, delineated by roles and responsibilities. [The work plan may include the Evaluation Design (a requirement of all evaluations). However, it is not always feasible to complete an evaluation design immediately upon award. Therefore, it is advised to separate the deliverable that kicks-off the evaluation from the design. It can take weeks to develop a good design and prepare data collection instruments that are participatory, utilization-focused, and incorporate all of the existing data.]
- Evaluation Design:** [REQUIRED] Within [# weeks] of approval of the work plan, the evaluation team must submit to the Agreement Officer's Representative/Contracting Officer's Representative (AOR/COR) an evaluation design (which will become an annex to the Evaluation report). The evaluation design will include: (1) a detailed evaluation design matrix that links the Evaluation Questions in the SOW to data sources, methods, and the data analysis plan; (2) draft questionnaires and other data collection instruments or their main features; (3) the list of potential interviewees and sites to be visited and proposed selection criteria and/or sampling plan (must

STATEMENT OF WORK TEMPLATE

include calculations and a justification of sample size, plans as to how the sampling frame will be developed, and the sampling methodology); (4) known limitations to the evaluation design; and (5) a dissemination plan. *[If applicable add a requirement to include a conflict of interest mitigation plan based on the Disclosure of Conflict of Interests submitted with the awardee's proposal].*

[RECOMMENDED language to include to #2] USAID offices and relevant stakeholders are asked to take up to *[# business days]* to review and consolidate comments through the AOR/COR. Once the evaluation team receives the consolidated comments on the initial evaluation design and work plan, they are expected to return with a revised evaluation design and work plan within *[# days]*. *[It is best practice to have the design reviewed and accepted by USAID before the evaluation team begins data collection or at a minimum within a period of time when it is still possible to change data collection strategies]*

3. **In-briefing:** *[OPTIONAL]* Within *[# days]* of arrival in *[specify location]*, the evaluation team will have an in-briefing with the *[insert offices/audience]* for introductions and to discuss the team's understanding of the assignment, initial assumptions, evaluation questions, methodology, and work plan, and/or to adjust the Statement of Work (SOW, if necessary).
4. **Mid-term Briefing and Interim Meetings:** *[OPTIONAL]* The evaluation team is expected to hold a mid-term briefing with *[specify USAID offices and/or staff]* on the status of the evaluation, including potential challenges and emerging opportunities. The team will also provide the evaluation COR/manager with periodic briefings and feedback on the team's findings, as agreed upon during the in-briefing. If desired or necessary, weekly briefings by phone can be arranged.
5. **Final Exit Briefing:** *[OPTIONAL]* The evaluation team is expected to hold a final exit briefing prior to leaving the country to discuss the status of data collection and preliminary findings. This presentation will be scheduled as agreed upon during the in-briefing. *[Specify guidelines of the presentation, e.g., who should be included, such as implementing partner staff or other stakeholders; preferred medium (joint or separate briefings); and expected maximum length]*
6. **Final Presentation:** *[OPTIONAL]* The evaluation team is expected to hold a final presentation in person/by virtual conferencing software to discuss the summary of findings and recommendations to USAID. This presentation will be scheduled as agreed upon during the in-briefing. *[Specify guidelines of the presentation, e.g., who should be included, such as implementing partner staff or other stakeholders; preferred medium (joint or separate briefings); expected maximum length; and timing (before or after the final report)].*
7. **Draft Evaluation Report:** *[REQUIRED]* The draft evaluation report should be consistent with the guidance provided in Section IX: **Final Report Format**. The report will address each of the questions identified in the SOW and any other issues the team considers to have a bearing on the objectives of the evaluation. Any such issues can be included in the report only after consultation with USAID. The submission date for the draft evaluation report will be determined in the evaluation work plan. Once the initial draft evaluation report is submitted, *[insert office/s]* will have *[number]* business days in which to review and comment on the initial draft, after which point the AOR/COR will submit the consolidated comments to the evaluation team. The evaluation team will then be asked to submit a revised final draft report *[number]* business days hence, and again the *[insert office/s]* will review and send comments on this final draft report within *[number]* business days of its submission. *[A good practice is for the evaluation team to share an early draft or detailed outline that includes main findings and bullets before finalizing the draft evaluation report]*
8. **Final Evaluation Report:** *[REQUIRED]* The evaluation team will be asked to take no more than *[number]* business days to respond/incorporate the final comments from the *[insert office/s]*. The

STATEMENT OF WORK TEMPLATE

evaluation team leader will then submit the final report to the AOR/COR. All project data and records will be submitted in full and should be in electronic form in easily readable format, organized and documented for use by those not fully familiar with the project or evaluation, and owned by USAID.

VII. EVALUATION TEAM COMPOSITION

Instructions: Before the standard language offered below, describe the **intended size** of an evaluation team, the appropriate expertise related to evaluation approaches (or **methodologies**), **methods, and subject matter expertise** required of the team or specific team members. Other skills that maybe included as requirements include language, geographic experience, among others.

Notes: A typical team should include one team leader who will serve as the primary coordinator with USAID. At least one team member should be an evaluation specialist. The **recruitment of local evaluators is highly encouraged**. Requested qualifications and/or skills may relate to: (1) evaluation design, methods, management, and implementation; (2) specific relevant technical subject matter expertise, (c) experience in USAID's cross-cutting program priorities, such as, gender equality and women's empowerment (d) regional or country experience; (e) local language skills.

All team members will be required to provide a signed statement attesting to a lack of conflict of interest or describing any existing conflict of interest.

The evaluation team shall demonstrate familiarity with USAID's [Evaluation Policy](#) and guidance included in the USAID Automated Directive System (ADS) in Chapter 200.

Optional: The [\[insert name\]](#) will participate on the evaluation team in [\[describe role\]](#). [See Guidance for USAID Staff Participation on External Evaluations for language.](#)

Optional: The COR of the Evaluation may observe [\[insert all or some\]](#) of the data collection efforts.

VIII. EVALUATION SCHEDULE

Instructions: Provide an estimated **timeframe (in days)** for the evaluation (period of performance) to be conducted as well as an anticipated start date. Period of performance should include the time it takes for USAID to review the draft and final evaluation reports and for all work to be completed for the evaluation. Likewise it is very important that the schedule include time for review throughout the process with key stakeholders and USAID staff. Consider including a timeline table (GANTT chart) or indicative schedule in narrative form.

Guidance: The sample table outlines these main phases of a performance evaluation. The guiding questions are:

1. What is the period of time (duration) you expect the evaluation team to take to review activity/project documents and become familiar with the program (prior to travel)?
2. How long will it take to get the necessary clearances for travel and to complete any protocols to visit communities and prepare for data collection?
3. How many sites/regions will the team be expected to visit, and what is a realistic timeframe for such requirements? Will the team be split up into smaller units during data collection to speed up the time required to collect the data?
4. What is the period of time (duration) it take to collect data?
5. What is the period of time (duration) allocated to analyze the data following data collection?
6. What is the period of time (duration) to prepare briefings and reports? If data visualization and graphical requirements are included state these.

STATEMENT OF WORK TEMPLATE

Sample Format: Illustrative Schedule

Timing (Anticipated Months or Duration)	Proposed Activities	Important Considerations/Constraints
	Preparation of the work plan and evaluation design	
	USAID review of the work plan and evaluation design	<i>Take into account availability in the Mission or Washington OU</i>
	Travel [optional: evaluation design] and preparations for data collection	<i>Take into account visa requirements (if an expatriate team is being mobilized)</i>
	In-Briefing	
	Data Collection	<i>Take into account the number of sites, methods, sectors, etc.</i>
	Data Analysis	<i>Take into account the number of sites, methods, sectors, etc.</i>
	Report writing	<i>Take into account the number of sites, methods, sectors, etc.</i>
	USAID review of Draft Report	<i>Take into account availability in the Mission or Washington OU</i>
	Incorporate USAID comments and prepare Final Report	

Instructions: The section should also include illustrative information about the level of effort (work days) to complete the evaluation. However, it is not required that specific and detailed level of effort be provided by team member. Requirements associated with the level of specificity for the level of effort are determined by the contracting mechanisms.

Level of effort calculations by team member are generally required to prepare an accurate Independent Government Cost Estimate (IGCE). See the Guidance Note on IGCE for Evaluations for a detailed explanation for estimating level of effort. Some key factors for determining the level of effort (number of work days to complete a task) include:

1. Planning/Evaluation Design: How many documents are there to review and how methods of data collection are anticipated? Time is required to review the documentation, prepare a work plan, and design instruments. Each method of data collection will require its own instrument.
2. Preparations for Data Collection: Is there an expatriate team? How long does travel take? How much travel is required outside of the capital city?
3. Data Collection: How many different geographic locations will be are required? How many people will travel to each location? How many days per person are required by method for data collection?
4. Analysis: How many different types of data sets are going to be generated? Are there quantitative data? If so, allocate time for data entry and cleaning.
5. Reporting and Briefing: How many different deliverables are required? Allocate time by deliverable and by person (not all team members will spend the same amount of time).

STATEMENT OF WORK TEMPLATE

The sample table shells are illustrative for a **simple evaluation** with four team members.

Sample Table: *Estimated LOE in days by activity for a team of four*

Task	LOE for Expat Team Lead	LOE for Expat [subject matter] Specialist	LOE for Local [subject matter] Specialist	LOE for Local [subject matter] Specialist	Total LOE in days
<i>Document review/desk review/work planning (evaluation design remote or in-country)</i>					
<i>Preparations for travel and organizing data collection (contracting translators, vehicles, etc.).</i>					
<i>In-brief, Evaluation Design (including meetings with USAID)</i>					
<i>Preparations for data collection (scheduling)</i>					
<i>Data collection days by method by site</i>					
<i>Data analysis</i>					
<i>Briefing</i>					
<i>Draft final report and debrief to USAID [include time for translation if necessary]</i>					
<i>Final report</i>					
Totals					

Sample Table: *Estimated LOE in days by position for a team of four*

Position	Preparation	Travel to/from Country	In-Country Data Collection	Finalization of Report	Total LOE in days
<i>Expat Team Leader</i>					
<i>Expat Specialist</i>					
<i>Local Specialist</i>					
<i>Local Specialist</i>					
Totals					

IX. FINAL REPORT FORMAT

The evaluation final report should include an executive summary; introduction; background of the local context and the projects being evaluated; the main evaluation questions; the methodology or methodologies; the limitations to the evaluation; findings, conclusions, and recommendations; and lessons learned (if applicable) as described [here](#). The report should be formatted according to the evaluation report [template](#).

The executive summary should be 3–5 pages in length and summarize the purpose, background of the project being evaluated, main evaluation questions, methods, findings, conclusions, and recommendations and lessons learned (if applicable).

STATEMENT OF WORK TEMPLATE

The evaluation methodology shall be explained in the report in detail. Limitations to the evaluation shall be disclosed in the report, with particular attention to the limitations associated with the evaluation methodology (e.g., selection bias, recall bias, unobservable differences between comparator groups, etc.)

The annexes to the report shall include:

- The Evaluation SOW;
- Any statements of difference regarding significant unresolved differences of opinion by funders, implementers, and/or members of the evaluation team;
- All tools used in conducting the evaluation, such as questionnaires, checklists, and discussion guides;
- Sources of information, properly identified and listed; and
- [Disclosure of conflict of interest forms](#) for all evaluation team members, either attesting to a lack of conflicts of interest or describing existing conflicts of.

In accordance with [AIDAR 752.7005](#), the contractor will make the final evaluation reports publicly available through the Development Experience Clearinghouse within 30 calendar days of final approval of the formatted report.

X. CRITERIA TO ENSURE THE QUALITY OF THE EVALUATION REPORT

Per the USAID Evaluation Policy and USAID ADS 203, draft and final evaluation reports will be evaluated against the following criteria to ensure the quality of the evaluation report.²

- The evaluation report should represent a thoughtful, well-researched, and well-organized effort to objectively evaluate what worked in the project, what did not, and why.
- Evaluation reports shall address all evaluation questions included in the SOW.
- The evaluation report should include the SOW as an annex. All modifications to the SOW—whether in technical requirements, evaluation questions, evaluation team composition, methodology, or timeline—need to be agreed upon in writing by the AOR/COR.
- The evaluation methodology shall be explained in detail. All tools used in conducting the evaluation—such as questionnaires, checklists, and discussion guides—will be included in an annex in the final report.
- Evaluation findings will assess outcomes and impact on males and females.
- Limitations to the evaluation shall be disclosed in the report, with particular attention to the limitations associated with the evaluation methodology (selection bias, recall bias, unobservable differences between comparator groups, etc.).
- Evaluation findings should be presented as analyzed facts, evidence, and data and not based on anecdotes, hearsay, or the compilation of people's opinions. Findings should be specific, concise, and supported by strong quantitative or qualitative evidence.
- Sources of information need to be properly identified and listed in an annex.
- Recommendations need to be supported by a specific set of findings.
- Recommendations should be action-oriented, practical, and specific, with defined responsibility for the action.

² See Appendix I of the Evaluation Policy and the Evaluation Report Review Checklist from the Evaluation Toolkit for additional guidance.

STATEMENT OF WORK TEMPLATE

OTHER REQUIREMENTS

[This section may include other requirements].

All quantitative data collected by the evaluation team must be provided in machine-readable, non-proprietary formats as required by USAID's Open Data policy (see ADS 579). The data should be organized and fully documented for use by those not fully familiar with the project or the evaluation. USAID will retain ownership of the survey and all datasets developed.

All modifications to the required elements of the SOW of the contract/agreement, whether **Select those that are applicable and included:** in technical requirements, evaluation questions, evaluation team composition, methodology, or timeline, need to be agreed upon in writing by the COR. Any revisions should be updated in the SOW that is included as an annex to the Evaluation Report.

X. LIST OF ANNEXES

<p><i>Instructions:</i> Include annexes to the SOW that will help the evaluation team design an effective proposal. This includes primary USAID guidance documents, publically available reports and data on the activity/project/program to be evaluated, and prior evaluation, etc.</p>
--

EVALUATION RESOURCE

EVALUATION STATEMENT OF WORK PEER REVIEW PROCESS

The Evaluation Statement of Work (Evaluation SOW) peer review is a required process to be completed after a SOW is drafted but before solicitation. This helps to ensure the quality of the Evaluation SOW.

Policy

Each operating unit's Program Office is required to organize peer reviews of evaluation statements of work as noted in the Evaluation Policy and Automated Directives System (ADS) 203.3.1.2. In doing so, they will work together with the technical office and may choose to work with individuals from regional and other Washington bureaus to assist in the peer review process. The Standardized Mission Order on Evaluation and this guidance document provides further details on the peer review process, but each mission may have customized aspects of the peer review.

Why conduct a peer review?

There are numerous reasons to conduct a peer review beyond it being a requirement of USAID policy. These include:

- Ensuring that the required elements of an Evaluation SOW are included.
- Improving the overall quality of the Evaluation SOW before it is used to direct the work of the evaluation team. The peer review process can involve experts in evaluation and the technical subject area of the evaluation to improve the SOW quality.
- Increasing the independence and objectivity of the Evaluation SOW. By bringing in additional staff members from other parts of the mission or Agency, the Evaluation SOW can benefit from perspectives that are not as close to the activity, project, or program being evaluated, thereby promoting a more neutral and unbiased perspective.
- Ensuring buy-in from internal stakeholders regarding the key features of the evaluation (purpose, questions, methods, timing, etc.). Evaluations will only be a worthwhile endeavor if they can produce credible evidence that can be and is used. By bringing in the primary audiences who may use the evaluation results, the drafters can help ensure that the SOW meets these audiences' needs.

Before the peer review

Development Objective (DO) teams will typically initiate and lead the writing of Evaluation SOWs in cooperation with the Program Office. The Program Office, Technical Office, and Contract Office should start talking as early as possible and expect to meet many times during the development of the SOW. The formal peer review should not be the first time the Program

EVALUATION RESOURCE

Office and Technical Office discuss the evaluation¹. Consider who needs to be involved in early stages and ensure that they are engaged. Consult the How-To Note on Evaluation Statements of Work for further details on drafting an Evaluation SOW.

The peer review

Once the Evaluation SOW is drafted and ready to be shared beyond those most directly involved in the drafting, the mission should organize the peer review. While there is no standard way of conducting a peer review, there are some standard practices outlined in the standardized Mission Order on Evaluation. Mission staff should consult their own Mission Order on Evaluation for peer review practices specific to their mission. Here are some of the key issues to be decided when conducting a peer review:

When will the peer review take place?

When planning the peer review, missions should aim for conducting it after a full draft of the Evaluation SOW has been completed. The draft should clearly represent the intention of the drafting team, but there is still time to make substantial changes based on the comments of the peer review process. Peer reviews should not be given an incomplete or half-formed Evaluation SOW; if a particular individual's input is needed to complete an Evaluation SOW draft, then he or she should be individually consulted prior to the peer review. Nor should peer reviews receive a fully completed SOW where only cosmetic changes are to be considered.

When scheduling the peer review, the Program Office and DO team drafters should ensure that there will be sufficient time after the peer review to incorporate comments and revise the SOW prior to preparing the solicitation.

How many and which individuals will be involved in the peer review?

As noted in the Standardized Mission Order on Evaluation, the mission's Evaluation point of contact (Evaluation POC) in the Program Office (or his or her designee) will lead a peer review of the SOW. The Evaluation POC or the designee should take lead responsibility for ensuring that the SOW meets the procedural standards and requirements of ADS 203 so that other peer reviewers can focus on substantive content.

The peer review should include no fewer than two individuals in addition to the Evaluation POC (or the designee). Emphasis should be placed on finding at least one peer reviewer with evaluation methods expertise. Peer reviewers may include individuals from the DO team and Program Office as well as USAID/Washington regional and technical bureaus, the Bureau for Policy, Planning and Learning Office of Learning, Evaluation, and Research (PPL/LER), external subject matter and evaluation experts, and local partners. It is best practice, however, for no more than half of the peer reviewers to be from the Technical Office that oversees the activity, project or program being evaluated. USAID/Washington regional bureaus have a particular responsibility to participate in peer reviews when requested by missions. Mission staff should consult the [Monitoring & Evaluation's POC List](#)² for Washington Bureau contact information.

¹ An SOW is also not the last time to discuss how the evaluation will be conducted. The SOW is not an evaluation design. Further refinement of SOW elements will occur after the evaluator is selected. Evaluation designs must be shared with relevant external stakeholders.

² The link to this document will only work when logged into the USAID information system.

EVALUATION RESOURCE

How much time will be provided for the peer review?

The Mission Order on Evaluation in each mission should specify the length of time peer reviewers will have to review the SOW. Good practice is to allow for 5 -10 business days for comments. Some USAID/Washington offices may have their own standards for how long they typically take to review an Evaluation SOW, so check with them if considering including USAID/Washington staff member in your peer review.

How will comments be received for the peer review?

There are a variety of ways of structuring the peer review process. Missions may choose to have a peer review meeting where individuals can discuss their comments on the draft SOW, request written comments on the draft SOW, or both. Many missions choose to ask the peer reviewers to fill out standard review sheets or checklists (see Evaluation Toolkit: Evaluation SOW Checklist and Review Template) while others prefer reviewers to send comments in an e-mail or in the document itself. The Evaluation POC should work with the evaluation drafters to determine what method would be most useful for receiving comments. Regardless of the method chosen, the Evaluation POC (or designee) should provide clear instructions to the peer reviewers regarding the means for providing comments on the draft SOW.

After the peer review

The DO team that initiated the draft SOW will typically make any revisions to the Evaluation SOW based on the peer review process in cooperation with the Program Office before final clearance by the Program Office. The leader of the peer review should be sure to follow up with the peer reviewers and provide the final evaluation SOW to them so that they are able to view the results of their participation.

For more information

For more information on the peer review process or on what to look for in an Evaluation SOW, check out the following:

- [Webinar: Good Practices for Peer Reviews of Evaluation SOWs](#). This webinar featured PPL, regional, and technical bureau representatives who have participated and led evaluation SOW peer reviews. They discussed good practices in conducting peer reviews and how to get the most out of the process, resulting in a strong SOW with buy-in from key stakeholders.
- [How-to Note: Evaluation Statements of Work](#). This Note addresses key issues for USAID staff who are developing a SOW for an externally contracted evaluation. It also serves as a guide for reviewing the quality of Evaluation SOWs for internal peer review processes.
- [Webinar: Developing Good Evaluation Questions](#). This webinar discusses how to develop evaluation questions that will help ensure high-quality performance evaluations from external evaluators. Critical tips will be presented, along with a variety of negative and positive examples from real USAID Evaluation SOWs.



HOW-TO NOTE

Engendering Evaluation at USAID

Monitoring and Evaluation Series

This Note describes key steps and good practices in engendering evaluation.

How-To Notes

are published by the Bureau for Policy, Planning and Learning and provide guidelines and practical advice to USAID staff and partners related to the Program Cycle. This How-To Note supplements ADS 203, ADS 205, the Evaluation Policy, and the Gender Equality and Female Empowerment Policy.

INTRODUCTION

The [USAID Evaluation Policy](#) (2011) and the related standalone [Automated Directives System \(ADS\) 203](#) affirm the importance of conducting and learning from rigorous evaluations as an integral part of the USAID program cycle. The release of [USAID's Gender Equality and Female Empowerment \(GE/FE\) Policy](#) in 2012 and the related standalone [ADS 205](#) heightened attention to gender integration across all phases of the program cycle, including the development of Country Development Cooperation Strategies (CDCSs), the project design process, drafting solicitations, and the monitoring and evaluation of program results.

Previous How-to Notes have addressed how to report on gender in operational plans (OPs) and Performance Plans and Reports (PPRs) and how to integrate attention to gender in project design and the resulting Project Appraisal Document (PAD). The purpose of this How-to Note is to describe key steps and good practices in engendering evaluations with the goal of assisting USAID staff to:

1. Design, manage, and participate in evaluations that reflect attention to gender issues;
2. Examine the extent to which USAID programs address gender issues and/or produce results that benefit people of both sexes; and
3. Assess whether addressing key gender gaps has resulted in better development outcomes.

USAID REQUIREMENTS RELATED TO GENDER IN EVALUATION

ADS 203, ADS 205, the Evaluation Policy, and the GE/FE Policy require rigorous monitoring and evaluation (M&E) that takes gender into account. In general, Bureaus and Missions should strive to ensure that evaluation designs, methodologies, data collection and analyses adequately capture the situations and experiences of both males and females. ADS 205 details several specific requirements, including that operating units (OUs) collect appropriate sex-disaggregated data, develop indicators designed to track changes in key gender gaps, and ask clear questions to uncover intended and unintended positive and negative changes for women or men using appropriate qualitative and quantitative methodologies.

OUs should also consider whether key evaluation questions examine the extent to which closing gender gaps has improved project outcomes and whether the project has transformed gender norms and reduced gender gaps for men and women across diverse sub-groups (e.g., different ages, people with disabilities, etc.), where applicable.

GENDER-SENSITIVE OR “ENGENDERED” EVALUATION

Meeting the requirements of ADS 205 is part of ensuring that an evaluation is gender-sensitive or “engendered.” Engendering an evaluation means that all stages of the evaluation reflect: (1) an awareness that the degree and meaning of program participation, program results, and potential sustainability are shaped by gender; (2) a recognition that explicit attention to gender issues must be integrated into the evaluation if gender equality objectives are to be addressed; and (3) a commitment to examining the extent to which gender equality was achieved as a result of the program or project that was implemented.¹ A fully gender-sensitive approach would include these elements in the Evaluation Statement of Work (SOW); the evaluation design, methodological approach, and data collection methods; and throughout data analysis and reporting. Without engendered evaluation, USAID will be unable to examine the extent to which its programming achieves positive results and improves quality of life for women as well as men; reduces gender gaps and empowers women and girls; and contributes to the high-level outcomes articulated in the GE/FE Policy.

ENGENDERING AN OPERATING UNIT’S EVALUATION PLANNING

Successful integration of gender into evaluations starts early. Although it is never too late to consider gender issues in an evaluation, better integration is likely to occur if the intersection of gender and evaluation is considered well before an individual evaluation is planned. From strategies to project design and implementation, there are numerous opportunities in the program cycle to consider integrating gender in order to produce more successful engendered evaluations. Some particularly opportune stages include:

The Country Development Cooperation Strategy (CDCS). Gender analysis is a mandatory analysis for a CDCS and should occur prior to development of the results framework. Final CDCS documents are required to identify high-priority evaluation questions. The mandatory gender analysis is a valuable source for considering where in the results framework an OU might want to focus evaluation questions that address gender issues, or when, over the life of the CDCS, it makes sense to ask gender-sensitive evaluation questions

KEY DEFINITIONS

Gender equality concerns fundamental social transformation—working with men and boys, women and girls, to bring about changes in attitudes, behaviors, roles and responsibilities at home, in the workplace, and in the community. Genuine equality means expanding freedoms and improving overall quality of life so that equality is achieved without sacrificing gains for males or females.

Female empowerment is achieved when women and girls acquire the power to act freely, exercise their rights, and fulfill their potential as full and equal members of society. While empowerment often comes from within, cultures, societies, and institutions create conditions that facilitate or undermine the possibilities for empowerment.

Gender integration involves identifying and then addressing gender inequalities during strategy and project design, implementation, and M&E. Since the roles and power relations between men and women affect how an activity is implemented, it is essential that project managers address these issues on an ongoing basis.

Gender-sensitive indicators point out to what extent and in what ways development programs and projects achieved results related to gender equality and whether/how reducing gaps between males/females and empowering women leads to better project/development outcomes.

¹ USAID (2014). Gender-Sensitive Evaluation: Best and Promising Practices in Engendering Evaluation. Prepared by Sharon Brisolara for JBS International, Inc., under contract AID-OAA-BC-10-00005 to the Bureau for Europe and Eurasia.

that can help the Mission understand to what extent and how gender equality goals are being met.

The Mission Performance Management Plan. The Mission evaluation plan is a required element of a Performance Management Plan (PMP) that is to be completed within six months after the development of a CDCS and continuously updated over the life of the CDCS. Although it only includes summary information about planned evaluations, viewing the entire set of evaluations together can help reveal where it would be helpful to further integrate gender concerns.

The Portfolio Review. Missions should consider during Portfolio Review where gaps in knowledge about the effects of USAID programming on gender issues might be addressed with an evaluation along with what has been learned about gender gaps from evaluations since the last Portfolio Review.

Project Design and Implementation. The GE/FE Policy and ADS 205 require that a gender analysis be carried out as part of the project design process and that the results of the analysis inform the design itself. This is to ensure that the project addresses the needs of both women and men, maximizes the likelihood that members of both sexes will be able to participate, and produces equally positive results for both. With these design considerations in mind, and by incorporating the collection of gender-sensitive indicators at both baseline and end line in the Project M&E Plan, design teams can ensure that gender-related results will be monitored across the life of the project and that evaluations can be designed to effectively address gender gaps and female empowerment. When Project M&E Plans and Activity M&E Plans are not designed, from inception, in a gender-sensitive manner, it is unlikely that a post-hoc decision to evaluate gender-related results will be successful.

ENGENDERING THE PLANNING AND DESIGN OF AN EVALUATION

The planning and design of evaluations, from the development of the SOW to the final design submitted by the evaluation design team, should integrate explicit attention to gender issues. Key areas of attention include:

Evaluation Purpose. Evaluation begins with a purpose. The evaluation purpose states why the evaluation is being conducted, who will use the results of the evaluation, and how they will do so. Purpose statements often link the evaluation to future decisions to be made by USAID leadership, partner governments, and/or other key stakeholders. Because USAID evaluations cannot address all aspects of a project from every angle, it is encouraged that they have a specific focus consistent with the resources devoted to the evaluation. It is entirely appropriate for a USAID evaluation to focus exclusively on issues of gender equality in the activity/project/program being evaluated. For example, a portfolio-level evaluation may focus on how gender has been addressed in the design and implementation of portfolio activities, or whether gender outcomes have improved in those areas where a variety of activities have been implemented.

However, in cases where gender is not the exclusive focus of the evaluation, or where the activity/project/program being evaluated does not have a primary focus on gender, addressing gender issues may still be a subsidiary purpose that is reflected throughout the evaluation SOW and evaluation design.

Evaluation Questions. For evaluations of gender equality programs/projects/activities with an exclusive focus on gender issues, evaluation questions that focus on specific gender concerns will naturally follow. For evaluations that do not have a primary purpose of addressing gender, gender should nevertheless be considered in the development of the evaluation questions. ADS 203 notes that an Evaluation SOW should identify all evaluation questions for which sex -disaggregated data are expected as well as identify questions for which an examination of gender-specific or gender-differential effects are expected.

In some cases, such integration will take the form of one or more questions that specifically focus on gender, such as whether an activity effectively included women when reaching out to potential beneficiaries or whether the empowerment of women increased over the course of a project. It is good practice to review the gender analyses that were carried out for the relevant CDCS and PAD when drafting evaluation questions to ensure that they are engendered.

At a minimum, the primary evaluation questions should include sub-questions that ensure that the data reported are sex-disaggregated. For instance, questions from a survey of project beneficiaries about knowledge gained from a recent USAID training should be reported separately for men and women.

Engendering the Methodology and Design. ADS 203 and 205 both note that evaluation methods should use sex-disaggregated data and incorporate attention to gender relations in all relevant areas.

For USAID impact evaluations, attention to gender will typically mean that the experimental or quasi-experimental design should estimate the impact of the USAID interventions on both male and female beneficiaries where appropriate. Doing so will require sufficient sample sizes and consideration as to whether the intervention is expected to have differential impacts on males and females.

For USAID performance evaluations, attention to gender will require the disaggregation of person-level output and outcome data that is presented as evidence in answering evaluation questions. The choice of evaluation designs and methods for performance evaluations will depend on the specific evaluation questions that must be addressed by the evaluation team, but should also take into account how design and method choices will affect the ability of the evaluation team to address gender. Both quantitative and qualitative methods may be appropriate, and often a mix of the two are optimal for engendered evaluations.

Because many development programs are conceived in a gender-neutral manner, project managers may fail to recognize the unintended consequences for women that result from their programs. Similarly, many evaluation methods only examine the expected outcomes of the project on the expected beneficiaries. In developing an engendered evaluation design, evaluators and evaluation managers should consider methods and designs that are capable of identifying both positive and negative unintended consequences for women or girls. For instance, this might include qualitative interviews or focus groups with women who were expected to benefit from the project but did not, or women who were only indirectly involved in the project. Participatory evaluation approaches may be particularly relevant, since deep involvement of local stakeholders, including women, in the design and conduct of an evaluation can help ensure that unintended consequences for women are avoided or addressed, and issues of gender equality are not overlooked.

KEY DEFINITIONS

Impact evaluations measure the change in a development outcome that is attributable to a defined intervention. Impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change.

Performance evaluations represent a broad range of evaluation methods. They often incorporate before/after comparisons but generally lack a rigorously defined counterfactual. Performance evaluations focus on what a particular project or program has achieved how, it was implemented, how it was perceived and valued, whether expected results occurred, and other questions that are pertinent to project design, management, and operational decision-making.

ENGENDERING THE CONDUCT OF THE EVALUATION

Gender expertise on the evaluation team. The extent and nature of the gender expertise needed on an evaluation team will, to some extent, depend on the type of evaluation questions that are being examined. If the evaluation is designed to examine questions that are primarily or wholly focused on gender, then at least one member of the design team should be a gender expert with experience in gender analysis and designing or leading engendered evaluations. It will also be beneficial if this team member or another person has specific knowledge of key gender issues in the sector being examined.

If only a small subset of the evaluation questions address gender issues, it may not be necessary to include a team member with sole responsibility for integrating gender in the evaluation. Nevertheless, one or more team members should have experience in engendered evaluation methods and knowledge of gender issues in the relevant sector. The evaluation team should also include one or more members with local cultural expertise, including an awareness of gender norms, how gender interacts with other identity elements, and which sub-groups of women may be at risk for exclusion from the project or evaluation.

Ideally, evaluation teams should include members of both sexes. Gender-balanced evaluation teams are particularly important in cultural contexts in which constraints prohibit women from talking to unrelated men, or where women may not be comfortable speaking to a man. Including local evaluators with relevant gender and cultural expertise can be particularly valuable in this regard.

Gender-sensitive data collection. Evaluators will need to be attuned and responsive to factors that might influence the likelihood that disproportionate numbers of males and females will participate in data collection for the evaluation, including factors such as where and how they spend their time, how much leisure time they have, whether there are prohibitions on women appearing in certain places or speaking with certain types of people, and whether powerful cultural gatekeepers have control over who participates.

Data collection instruments and protocols should also reflect an understanding of gender roles and constraints in a particular cultural context. For example, questions on a data collection tool may need to use locally recognized symbols or terminology, be sensitive to potentially different meanings that males and females might ascribe to the same terms, acknowledge and collect information about the different roles that men and women play in the sector being examined, or ask sex-specific questions to tap into the unique experiences of men and women. Data collection protocols will also need to reflect local contexts and norms concerning the conditions under which women (or men) feel empowered to speak freely. These considerations could determine, for example, whether it is best to collect data individually or in groups, whether groups should include all people of the same sex or both sexes, or whether groups should also be stratified by age. These considerations could also determine *where* it is best to collect data, since local contexts and norms may influence whether women (or men) feel empowered to speak freely in various locations, such as the home, the street, a village square, or an institutional setting, such as a hospital.

CHALLENGES ENCOUNTERED IN ENGENDERING EVALUATION

There are many and diverse reasons why evaluations may not be fully engendered and/or evaluation results may not fully reveal hidden gendered patterns of participation or results. Many of these reasons can be traced to problems with initial project design, including: (1) the absence of or failure to utilize a quality gender analysis in the relevant sector; (2) a planned timeline insufficient to capture transformative gender results; (3) an engendered project design that was not fully implemented; and (4) lack of attention to contextually relevant gender-sensitive indicators.

Design problems can also lead to segments of the population being absent from the evaluation, especially underrepresented or marginalized groups that include vulnerable sub-groups of women and girls. Gender-blind data collection tools, protocols, and research methods may unintentionally narrow the diversity of perspectives and experiences captured from key stakeholders, especially those who are low in social power. Even evaluations that include a careful and thorough examination of expected gender-related results may suffer from the failure to anticipate and investigate unintended consequences of the program or project, including harmful or negative effects on gender norms, women's experiences, or female empowerment. Engendered evaluation may also be undercut by insufficient knowledge, interest in, or commitment to gender equality goals among USAID technical staff or leadership. Many of these challenges can be managed, at least in part, by including people with relevant gender expertise in all stages of project design and implementation as well as in the design, management, and execution of evaluations.

ADDITIONAL RESOURCES

The following resources can be used to provide additional information. Some other resources exist but are out-of-date with current USAID guidance. Where information differs, the USAID Evaluation Policy and the USAID ADS (Automated Directives System) 200 series take precedence over information in other resources.

- Gender-Sensitive Evaluation: Best and Promising Practices in Engendering Evaluation: http://pdf.usaid.gov/pdf_docs/PA00K43P.pdf
- USAID Evaluation Policy: <http://www.usaid.gov/sites/default/files/documents/1868/USAIDEvaluationPolicy.pdf>
- USAID Gender Equality and Female Empowerment (GE/FE) Policy: http://www.usaid.gov/sites/default/files/documents/1865/GenderEqualityPolicy_0.pdf
- USAID ADS 203: <http://www.usaid.gov/sites/default/files/documents/1870/203.pdf>
- USAID ADS 205: <http://www.usaid.gov/sites/default/files/documents/1870/205.pdf>

EVALUATION RESOURCE

THE EVALUATION DESIGN MATRIX: TEMPLATES

The evaluation design matrix is an essential tool for planning and organizing an evaluation. It is simply a table with one row for each evaluation question and columns that address evaluation design issues, such as data collection methods, data sources, analysis methods, criteria for comparisons, etc. The design matrix links each evaluation question to the means for answering that question. There is no single format for a design matrix, but a number of example templates are provided below.

When to use an Evaluation Design Matrix?

USAID recommends that the Statement of Work (SOW) for an evaluation drafted by USAID include an illustrative design matrix, but it is not required. USAID also recommends that evaluation teams include a design matrix in their proposals and in their evaluation designs prepared after the evaluation contract is awarded. While evaluation design matrices are not required to be in evaluation proposals or evaluation designs under USAID policy, operating units may choose to make them required deliverables from evaluation teams in the SOW.

Why use an Evaluation Design Matrix?

An evaluation design matrix can provide a valuable tool for aligning expectations between USAID and the evaluation team about the evidence that will be generated for answering each evaluation question in an evaluation. In a typical evaluation SOW at USAID, there is one section that lists the “evaluation questions” and one section that describes the “evaluation methodology.” This structure is also sometimes found in evaluation designs provided by evaluation teams. This structure makes intuitive sense and would be adequate if there were only one evaluation question being asked. However, it’s usually the case that multiple questions are asked in USAID evaluations.

A problem thus arises when you have an “evaluation questions” section that lists multiple evaluation questions, but a methodology section that lists a single methodological approach or even a variety of approaches, but does not indicate which methodologies are related to which evaluation questions. In practice, each evaluation question typically requires a specific source of data and/or data collection and analysis methodology for answering that question. The evaluation design matrix explicitly links each question with the data source and methodology. The evaluation design matrix:

- Helps the drafters of the evaluation SOW or evaluation design to carefully consider and systematically organize all of the linkages between questions and methods.
- Helps the drafters of the evaluation SOW or evaluation design to prepare a more accurate budget estimate.
- Clarifies expectations and enhances cooperation between USAID and the evaluation team.

How much detail should go into an evaluation Design Matrix?

At minimum, the evaluation design matrix should identify each question and provide some information regarding how the question will be answered. Some design matrices include many elements that provide detailed systematic design information for each of the evaluation questions. They typically do not include schedules or workplans, although such information may be included in a linked document. Typically, an evaluation design matrix will be less detailed and more illustrative in the evaluation Statement of Work prepared by USAID and more detailed and operational in the approved evaluation design prepared by the evaluation team. There is always the likelihood that evaluation plans change once fieldwork commences, though, and evaluation design matrices provide a useful tool for modifying and updating the evaluation plan.

EVALUATION RESOURCE

Design Matrix Template 1 (Simple version):

Evaluation Name: _____

Evaluation Purpose: _____

Evaluation Question	Suggested methods for answering this question
<i>Question 1</i>	<i>What data sources and data collection and analysis methods will be used to produce the evidence for answering this question?</i>
<i>Question 2</i>	
<i>Question 3</i>	

EVALUATION RESOURCE

Design Matrix Template 2 (GAO version¹):

Evaluation Name: _____

Evaluation Purpose: _____

Researchable Question(s)	Information Required and Source(s)	Scope and Methodology	Limitations	What this Evaluation will Likely allow the evaluator to Say
<p>What questions is the team trying to answer?</p> <p>Identify specific questions that the team must answer.</p> <p>Ensure each major evaluation question is specific, objective, neutral, measurable, and doable.</p> <p>Ensure key terms are defined.</p> <p>Each major evaluation question should be addressed in a separate row.</p>	<p>What information does the team need to address each evaluation question? Where will they get it?</p> <p>Identify documents or types of information that the team must have.</p> <p>Identify plans to address internal controls and compliance.</p> <p>Identify plans to collect documents that establish the "criteria" to be used.</p> <p>Identify plans to follow up on known significant findings that team found in obtaining background information.</p> <p>Identify sources of the required information, such as databases, studies, subject area experts, program officials, models, etc.</p>	<p>How will the team answer each evaluation question?</p> <p>Describe strategies for collecting the required information or data, such as random sampling, case studies, focus groups, questionnaires, benchmarking to best practices, use of existing data bases, etc.</p> <p>Describe the planned scope of each strategy, including the timeframe, locations to visit, and sample sizes.</p> <p>Describe the analytical techniques to be used, such as regression analysis, cost benefit analysis, sensitivity analysis, modeling, descriptive analysis, content analysis, case study summaries, etc.</p>	<p>What are the design's limitations and how will it affect the evaluation?</p> <p>Cite any limitations as a result of the information required or the scope and methodology, such as:</p> <p>—Questionable data quality and/or reliability.</p> <p>—Inability to access certain types of data or obtain data covering a certain time frame.</p> <p>—Security classification or confidentiality restrictions.</p> <p>—Inability to generalize or extrapolate findings to the universe.</p> <p>Be sure to address how these limitations will affect the evaluation.</p>	<p>What are the expected results of the work?</p> <p>Describe what the evaluation team can likely say. Draw on preliminary results for illustrative purposes, if helpful.</p> <p>Ensure that the proposed answer addresses the evaluation question in column one.</p>

EVALUATION RESOURCE

Design Matrix Template 3 (Morra Imas and Rist versionⁱⁱ):

Evaluation Name: _____

Evaluation Purpose: _____

General Evaluation

Approach: _____

1. Question	2. Sub-question	3. Type of Sub-question	4. Measure or Indicator	5. Target or Standard (for Normative Questions only)	6. Baseline Data?

EVALUATION RESOURCE

Evaluation Design Matrix 3 (Columns 7-12)

7. Data Source	8. Design Strategy for Question	9. Sample or Census	10. Data Collection Instrument	11. Data Analysis	12. Comments

ⁱ Adapted from U.S. Government Accountability Office. 1991. Designing Evaluations. Washington, DC PEMD-10.1.4., p.20.

ⁱⁱ Adapted from Morra Imas, Linda G., Ray C Rist. 2009. The Road to Results: Designing and Conducting Effective Development Evaluations. The World Bank, Washington DC., p.243.

EVALUATION RESOURCE

GUIDANCE FOR USAID STAFF PARTICIPATION ON EXTERNAL EVALUATIONS

I. Introduction

The USAID Evaluation Policy places considerable emphasis on external evaluation. It states that “Evaluations of USAID Projects will be undertaken so that they are not subject to the perception or reality of biased measurement or reporting due to conflict of interest or other factors. In most cases, the evaluations will be external...” However, the Evaluation Policy also allows USAID participation on external evaluations. It states that “In cases where a USAID operating unit management determines that appropriate expertise exists within the Agency, and that engaging USAID staff in an evaluation will facilitate institutional learning, an evaluation team may be predominantly composed of USAID staff. However, an outside expert with appropriate skills and experience will be recruited to lead the team, mitigating the potential for conflict of interest.”

This guidance concerns the participation and conduct of USAID staff as team members on *external evaluations*. For the purpose of this guidance, participation refers to actively contributing, in whole or in part, to the evaluation report under the direction of the external evaluation team leader.

For more information regarding participation of USAID staff members on *internal* evaluations for non-required evaluations see the Evaluation Toolkit resource: [Choosing between Internal or External Evaluation](#). For more information regarding monitoring, observing, or providing technical direction to external evaluation teams as an Evaluation COR, see the Evaluation Toolkit resource: [Tips for Managing a USAID Performance Evaluation](#).

II. What are the benefits to USAID staff participating on external evaluation teams?

USAID participation on an external evaluation has the potential to benefit USAID in many ways.

First, participation of USAID staff on evaluations can provide a learning opportunity to the USAID staff member. Participating on an evaluation can help the participant learn more in depth about evaluation practices and procedures, promote a more data-driven approach to project management and decision-making, and raise awareness of the

USAID evaluations are classified by the composition of the evaluation team as either *internal* or *external*. (Automated Directives System [ADS] 203.3.1.1)

External evaluations:

1. Are commissioned by USAID rather than by the implementing partner, and
2. Have a team leader who is an independent expert from outside the Agency and who has no fiduciary relationship with the implementing partner.

Internal evaluations are either:

1. Conducted or commissioned by an implementing partner concerning their own project (an implementer internal evaluation), or
2. Commissioned by USAID in which the evaluation team leader is from inside the Agency (a USAID internal evaluation).

Only external evaluations count toward the “large project” and “pilot project” evaluation requirements. Optional evaluations may be internal.

EVALUATION RESOURCE

practical opportunities and constraints faced by external evaluators. In addition to learning about evaluation, participation can help the USAID staff member gain more insight into activity implementation in the field in general, along with specific information about the particular activity that is being evaluated.

Second, participation of USAID staff on evaluations can help enhance the technical quality of the evaluation through leveraging in-house skills. Many USAID staff bring strong subject matter expertise and familiarity with the technical implementation of USAID programming that can complement and enhance the technical expertise contracted by the external evaluation team.

Third, USAID staff can bring their insider knowledge of USAID to an external evaluation team that may not be familiar with the procedures, norms, culture, and operating context in the Agency or a particular mission. USAID participation on evaluation teams can help, for instance, in making sure evaluation recommendations are relevant and actionable in the USAID context.

Fourth, USAID participation on evaluations can help ensure the learning from the evaluation is institutionalized and the evaluation findings utilized. At the end of the evaluation, the external consultant will move on to the next job, but the USAID staff member who participated in the evaluation can remain available after the formal evaluation period ends, which can be useful for follow-on design work, strategy development, etc. USAID evaluation participants are more likely to internalize findings and incorporate the learning from the evaluation into operating unit (OU) and Agency decision-making.

Finally, USAID participation on evaluation encourages sharing of lessons learned among different USAID OUs and projects. Engaging staff from other missions or USAID/Washington fosters a natural cross-pollination of ideas and learning.

III. What are the challenges for USAID staff participating on evaluations?

There are numerous challenges that USAID staff may need to overcome to participate effectively on a USAID evaluation.

First, USAID evaluations can be time consuming, requiring USAID staff to spend considerable time away from their regular daily activities. If a team member cannot maintain his or her commitments to the evaluation team, it could jeopardize the timeliness and quality of the evaluation report and make completion more difficult for external team members.

Second, if USAID staff members on evaluation teams have different or unclear expectations about their role(s) on the evaluation team, it can lead to conflict with the evaluation team leader and or other members.

Third, as employees of USAID, staff members may bring their biases or preconceived ideas about the project being evaluated, jeopardizing the objectivity of the evaluation report. Even when a USAID staff member is able to be fully objective, his or her presence may still lead to a perceived lack of objectivity by the evaluation team. While having an external team leader is intended to mitigate such biases and perceptions, there is no guarantee that this arrangement will do so.

EVALUATION RESOURCE

Finally, the participation of USAID staff members on evaluation teams has the potential for biasing responses from key informants (particularly implementing partners) when these informants are interviewed by USAID staff or even when a USAID observer is in the interview room.

IV. What recommendations does the Office of Learning, Evaluation, and Research have for USAID participation on external evaluations?

Despite the potential drawbacks, the Bureau for Policy, Planning and Learning Office of Learning, Evaluation, and Research (PPL/LER) generally considers that the overall benefit to USAID participation on evaluations outweighs the costs.

PPL/LER supports the participation of USAID staff members on externally led evaluations and encourages USAID OUs to seek out ways to increase such participation. However, this support does not suggest that the inclusion of USAID staff on evaluations is appropriate in all circumstances. PPL/LER encourages OUs to thoroughly assess the appropriateness of individual USAID staff members to participate in evaluations in each instance it is considered.

To help ensure that USAID participation will benefit the evaluation process, PPL/LER suggests a number of practices to consider when planning for and implementing an external evaluation that includes USAID staff.

A. Participation on evaluation teams

USAID policy does not specify who may or may not participate on an external evaluation as a team member. However, when considering who should participate in an externally led evaluation, PPL/LER recommends:

- ***Agreement Officer's Representatives/Contracting Officer's Representatives (AORs/CORs) of implementing mechanisms being evaluated should not participate as evaluation team members.*** AORs/CORs of implementing mechanisms being evaluated should have a role in the evaluation process, from leading or participating in the drafting of the statement of work (SOW) to reviewing and utilizing the evaluation report. However, their close relationship and responsibilities to the activity being evaluated is likely to harm the perception of objectivity of the evaluation team that is key to an external evaluation if they were to participate as evaluation team members.
- ***CORs of the evaluation contract/task order should not participate as team members.*** CORs of the evaluation contract/task order also have a role in the evaluation process, from leading or participating in the drafting of the statement of work (SOW), supporting and providing technical direction to the evaluation team, monitoring the evaluation team's progress and deliverables, to reviewing and utilizing the evaluation report. However, CORs have obligations to the Contracting Officer and to the external evaluation team leader that conflict with a role of serving under the evaluation team leader on an evaluation.
- ***For other USAID staff members, the Program Office should consider the appropriateness of participation on an evaluation team and serve as the office that approves or does not approve participation.*** As the office that is expected to manage external evaluations in most instances (ADS 203.3.1.4), the Program Office should

EVALUATION RESOURCE

consider the appropriateness of individual USAID staff members' participation on evaluation teams with the above-noted benefits and challenges in mind. USAID staff members from the same Technical Office as the activity or project being evaluated should not automatically be eliminated from participation, but their involvement with the activity/intervention, if any, should be examined. For instance, if the evaluation examines how an implementer has adhered to or deviated from the activity design, a USAID staff member (from the same OU or another OU) who participated in that design may be a welcome addition. However, if the evaluation is also expected to examine whether the activity design was appropriate or relevant given country conditions, then a USAID staff member who participated in the design of the activity is unlikely to be an objective team member. For evaluations that particularly focus on ensuring USAID accountability to external audiences, then no USAID participation may be the most prudent course, while more exploratory evaluations focused on internal adaptation and learning may be particularly appropriate for USAID participation.

- ***USAID participants on evaluation teams should have familiarity with USAID evaluation policies and practices.*** PPL/LER recommends that USAID staff members take a USAID-sponsored evaluation training prior to participating on a USAID evaluation. Also, staff should have read and be familiar with ADS 203 as well as relevant PPL-issued evaluation guidance.
- ***Participation should not be limited solely to senior monitoring & evaluation specialists or senior technical experts.*** There are many reasons for USAID participation on an external evaluation, as well as benefits. It should not be limited to those who are senior experts in evaluation or the technical subject matter of the project or activity being evaluated. While such team members are likely to provide strong contributions to the evaluation, individuals new to evaluation, and/or from a different technical sector, can learn from and also contribute to an evaluation team.

B. Planning for USAID participation on evaluation teams

Prudent planning can help ensure that USAID staff participation will benefit the evaluation. PPL recommends the following.

- ***OUs should decide as early as possible in the evaluation planning process whether a USAID staff member will participate in the evaluation.*** Preferably, such decisions will occur prior to or during the drafting of the SOW, as this will help in preparing the SOW—particularly the section on the team composition and the Independent Government Cost Estimate. In addition to knowing whether or not a USAID staff member or members will participate, it will help to identify:
 - a. The specific individual(s) who will participate;
 - b. The reason for participation (Is it primarily for the individual's learning or for the technical expertise that they can contribute?);
 - c. What expectations they have for contributing to the evaluation (Will they expect to just observe, work on the evaluation design, participate in data collection, etc.?);

EVALUATION RESOURCE

- d. What skills they bring to the evaluation (for instance, country/regional knowledge, language skills, technical subject expertise, evaluation expertise, familiarity with USAID processes, etc.); and
 - e. How much time they have to commit to the evaluation.
- ***OUTS should inform prospective evaluation contractors in the evaluation solicitation that USAID expects to include USAID staff on the evaluation.*** Potential evaluation contractors should be provided with information on the number of participants and other relevant information, including expected roles, skills and experience, and time commitment. If decisions are not made prior to solicitation of the evaluation regarding USAID participation, the evaluation contractor should be informed as early as possible following award and prior to developing a workplan and final design. Following the award, the evaluation team should be given CVs and other additional information about the specific individuals participating on the evaluation team.
- ***The evaluation team leader, the USAID team member(s), and the evaluation COR should meet as soon as possible following award to discuss team roles.*** In addition to providing any further information about USAID staff member(s) to the evaluation team leader, the USAID team member(s), the external team leader, and the COR should meet as soon as possible (in many cases, this may be a virtual meeting) to discuss:
 - a. Expected roles of the evaluation team members;
 - b. The time commitment of the USAID staff member(s);
 - c. Limitations on the participation of USAID staff members in certain aspects of the data collection—such as interviews with particular key stakeholders where USAID presence may bias interview responses;
 - d. Coordination of logistics, for instance, if USAID staff will need to travel in separate vehicles to data collection sites; and
 - e. Coordination on communication within the team and between USAID and the evaluation team.

This meeting should also serve as a time for USAID to reiterate that the external evaluation team leader is recognized by USAID as the leader of the evaluation team and that USAID staff on the evaluation will serve as team *members*, not secondary team leaders. If necessary, the meeting may lead to the development of a Memorandum of Understanding between the Mission and leadership of other OUs if USAID evaluation team members are participating from OUs that are not managing the evaluation.

C. Conducting the evaluation

In the conduct of an external evaluation with USAID participation, PPL recommends:

- ***USAID evaluation team members participating on evaluations to learn the evaluation process should have the opportunity to participate in all aspects of the evaluation, from supporting design to writing the final report.*** External evaluations may benefit from USAID participation in limited, targeted roles on an evaluation (for instance, to provide technical expertise during the design phase or to assist in the data collection phase of an evaluation). However, full integration is likely to be the most beneficial. This is particularly true for USAID team members who participate on

EVALUATION RESOURCE

evaluations in order to better understand the evaluation process. Moreover, USAID team members should not be excluded arbitrarily from participating in particular phases, such as writing the report, if they have committed the time to participate.

- ***USAID evaluation team members should not participate on an evaluation “part-time.”*** Regardless of how expansive or limited the role is of USAID team members, it is recommended that during the time period they have committed to be members of the evaluation team, the USAID team members serve as full-time members of the evaluation team, particularly if they are involved in data collection for the evaluation. It can be disruptive to the evaluation if team members are pulled away or unavailable at certain times. Supervisors should be made aware of and approve the expected level of effort required to fully commit to the evaluation role the USAID team member has agreed to.
- ***Team members should remain flexible and defer to evaluation team leaders on the conduct of the evaluation.*** Even the most thoroughly planned evaluations require adaptation in the field. Similarly, agreed-upon roles and responsibilities of USAID staff on the evaluation and the evaluation team leader may need to be adapted during implementation of the evaluation. In general, USAID team members should defer to the evaluation team leaders on matters related to the conduct of the evaluation and reporting of evaluation results (within ethical boundaries). If USAID evaluation team members disagree with evaluation findings, they have the opportunity, under USAID policy, to prepare a statement of differences.
- ***Communication between the evaluation team and the OU commissioning the evaluation should be conducted through the evaluation team leader and the USAID evaluation COR.*** USAID participation on an evaluation team should *not* be seen as an additional means for monitoring the external evaluation team. Back-channel communications between USAID evaluation team members and USAID staff in the OU commissioning the evaluation should be discouraged.

Disclosure of Real or Potential Conflict of Interest for USAID Evaluations

Instructions:

Evaluations of USAID projects will be undertaken so that they are not subject to the perception or reality of biased measurement or reporting due to conflict of interest.¹ For external evaluations, all evaluation team members will provide a signed statement attesting to a lack of conflict of interest or describing an existing conflict of interest relative to the project being evaluated.²

Evaluators of USAID projects have a responsibility to maintain independence so that opinions, conclusions, judgments, and recommendations will be impartial and will be viewed as impartial by third parties. Evaluators and evaluation team members are to disclose all relevant facts regarding real or potential conflicts of interest that could lead reasonable third parties with knowledge of the relevant facts and circumstances to conclude that the evaluator or evaluation team member is not able to maintain independence and, thus, is not capable of exercising objective and impartial judgment on all issues associated with conducting and reporting the work. Operating Unit leadership, in close consultation with the Contracting Officer, will determine whether the real or potential conflict of interest is one that should disqualify an individual from the evaluation team or require recusal by that individual from evaluating certain aspects of the project(s).

In addition, if evaluation team members gain access to proprietary information of other companies in the process of conducting the evaluation, then they must agree with the other companies to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.³

Real or potential conflicts of interest may include, but are not limited to:

1. Immediate family or close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated.
2. Financial interest that is direct, or is significant/material though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation.
3. Current or previous direct or significant/material though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project.
4. Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated.
5. Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated.
6. Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.

¹ USAID Evaluation Policy (p. 8); USAID Contract Information Bulletin 99-17; and Federal Acquisition Regulations (FAR) Part 9.5, Organizational Conflicts of Interest, and Subpart 3.10, Contractor Code of Business Ethics and Conduct.

² USAID Evaluation Policy (p. 11)

³ FAR 9.505-4(b)

Disclosure of Conflict of Interest for USAID Evaluation Team Members

Name	
Title	
Organization	
Evaluation Position?	<input type="checkbox"/> Team Leader <input type="checkbox"/> Team member
Evaluation Award Number <i>(contract or other instrument)</i>	
USAID Project(s) Evaluated <i>(Include project name(s), implementer name(s) and award number(s), if applicable)</i>	
I have real or potential conflicts of interest to disclose.	<input type="checkbox"/> Yes <input type="checkbox"/> No
If yes answered above, I disclose the following facts: <i>Real or potential conflicts of interest may include, but are not limited to:</i> <ol style="list-style-type: none"> 1. Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated. 2. Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation. 3. Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project. 4. Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated. 5. Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated. 6. Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation. 	

I certify (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	
Date	



USAID
FROM THE AMERICAN PEOPLE

VERSION I | JUNE 2013

TECHNICAL NOTE

Conducting Mixed-Method Evaluations

Monitoring and Evaluation Series

This document provides guidance on using mixed-methods for evaluations.

Technical Notes are published by the Bureau of Policy, Planning and Learning and provide key concepts and approaches to USAID staff and partners related to the Program Cycle.

INTRODUCTION

Over the past three decades, evaluators and others in the development field have increasingly recognized that incorporating multiple methods into a single evaluation often results in a stronger, more complete evaluation than conventional evaluation approaches relying on only one method. This trend has led to a rapidly growing interest in mixed-method evaluations among both practitioners and evaluators. At least two journals dedicated to mixed-method evaluations have been launched, and the number of books on the subject is growing steadily. Notably, USAID's *Evaluation Policy* strongly endorses mixed-method evaluation approaches: "Given the nature of development activities, both qualitative and quantitative methods yield valuable findings, and a combination of both often is optimal." (p. 4) This Technical Note provides guidance to USAID staff and partners on how mixed-method evaluations are conducted and important considerations when managing a mixed-method evaluation.

DEFINITION

A mixed-method evaluation systematically integrates two or more evaluation methods, potentially at every stage of the evaluation process, usually drawing on both quantitative and qualitative data. Mixed-method evaluations may use multiple designs, for example incorporating both randomized control trial experiments and case studies. They also may include different data collection techniques such as structured observations, key informant interviews, household surveys, and reviews of existing secondary data. In short, a mixed-method evaluation involves the systematic integration of different kinds of data, usually drawn from different designs. As a result, mixed-method evaluations require advanced planning and careful management at each stage of the evaluation process.

RATIONALE

The three main cases in which mixed-method designs help to strengthen an evaluation are:

- (1) When different evaluation questions require different methods, or when a single evaluation question requires more than one method to answer all components.

- (2) When different methods are used to answer the same elements of a single question, increasing confidence in the validity and reliability of the evaluation results.
- (3) When the results from one method are used to help design future phases of the evaluation using other methods.

In addition to these three main reasons, there are other benefits that can be realized by using mixed-method designs or data collection strategies. For example, mixed-methods approaches:

- Are more likely to reveal unanticipated results.
- Can provide a deeper understanding of why change is or is not occurring as planned.
- Often capture a wider range of perspectives than might be captured by a single method.

(1) USING DIFFERENT METHODS TO ANSWER DIFFERENT QUESTIONS OR TO ANSWER DIFFERENT PARTS OF THE SAME QUESTION

In many cases, one evaluation method will be insufficient to answer all of the questions included in an evaluation statement of work (SOW). For example, suppose an SOW involves an evaluation of a project that includes a new teaching technique, and includes two questions: #1. “Was there a statistically significant difference between female and male students’ academic achievement test scores?” and #2. “How did students’ parents perceive the effects of the project?” A single method will likely be insufficient to adequately answer both of these questions.

To answer the first question, the evaluator might choose a quasi-experimental design that uses existing test scores from before project implementation and new test scores from after completion to compare the performance of male and female students. This approach would address the question of whether the program resulted in differences in test scores between females and males. But these methods would not help to answer question #2. To understand parent perceptions, the evaluator likely would use individual or focus group interviews of a sample of parents, and perhaps conduct an evaluative case study in order to more deeply understand how parents view the program.

Sometimes, one evaluation question may contain multiple parts, and it may be necessary to use different methods to address each part. For example, question #1 could be changed slightly to ask, “Was there a statistically significant difference between female and male students’ scores? And, what explains possible gender differences in test scores?” In this case, simply relying on test scores in a quasi-experimental design would be insufficient. To understand the mechanism behind differences in female and male scores would require a different method, such as key informant interviews with teachers or focus group discussions with students.

(2) USING DIFFERENT METHODS TO ANSWER THE SAME QUESTION: TRIANGULATION

Even if an evaluation question can be answered using only one method, often it is preferable to combine multiple methods to answer the same question in order to gain a more complete understanding of the issue and more confidence in the findings. By approaching the same question from more than one perspective or by using more than one technique, evaluators can then compare and contrast the results from these different methods. This process is known as triangulation. If the findings from the different methods are similar, or reinforce one another, then users can have greater confidence in the findings than if they are based on only one method.

A NOTE ON TERMINOLOGY

In the literature on evaluation, “method” is sometimes used to refer to a data collection technique (interviews, surveys, observations), and other times to an evaluation design or approach (experimental, quasi-experimental, non-experimental). Though the definition is not completely settled in the literature, this Technical Note treats evaluations that combine methods in either sense as mixed-method evaluations.

If the findings from different methods vary significantly, the user and evaluator must carefully consider what might have happened to produce these divergent findings. One possible explanation could be bias in one set of data. Triangulation can help to minimize bias in cases like these, with data from one method acting as a check or balance against data from another method. For example, evaluators may use secondary data from the Ministry of Economy to measure changes in exports related to a trade facilitation project. But they may also suspect that firms are underreporting their exports to the government in order to pay less in taxes. To help

QUICK ADVICE FOR THE USAID EVALUATION MANAGER

Mixed method evaluation involves the systematic integration of different kinds of data, usually drawn from different evaluation designs. As a result, mixed-method evaluations require advanced planning and affect the evaluation budget.

Mixed method evaluations yield valuable findings, and a combination of both is often optimal given the nature of development activities. Mixed methods are more likely to reveal unanticipated results (a key advantage of evaluation over performance monitoring), which can provide a deeper understanding of why change is or isn't taking place as planned. It often captures a wider range of perspectives than might be captured by a single method.

At the planning stage, the evaluation manager must decide which methods to use and how to combine them. These decisions will be based primarily on the purpose of the evaluation and the key evaluation questions, but evaluation managers must also take into account factors such as time and cost. An evaluation manager should seek advice from a colleague with evaluation design expertise, as needed.

Lastly, using mixed methods requires mixed skills on the evaluation team. It is important to consider during the planning stage what kind of skills will be needed in order to conduct each aspect of the evaluation successfully, and then to select team members accordingly.

mitigate the risk of bias caused by this underreporting in the government data, the evaluation team may distribute a survey to supported firms and also conduct in-depth interviews with key informants from a sub-sample of firms in order to obtain a more accurate picture of how the project has influenced exports. Evaluators looking to answer one question with multiple methods often combine them using the parallel process described below. It is important for the evaluation manager to understand how this process of triangulation will work, because it has implications for the resources needed to carry out such an evaluation.

(3) USING ONE METHOD TO INFORM THE DESIGN OF ANOTHER METHOD

In some cases, one method can be used to help guide the use of another method, or to explain the findings from another method. In the first case, imagine an SOW for the evaluation of a youth vocational training project including the evaluation question: "Why do youth choose to participate in project activities?" The evaluator may wish to conduct a survey of participants, but be unsure how to word the questions, or what answer choices to include. By first conducting individual and focus group interviews with participants and non-participants, the evaluator may be able to identify some common reasons for participation among the target population, and then use these data to construct the survey. In this way, the qualitative methods (individual and focus group interviews), conducted first, can inform the quantitative method (survey), that comes afterward. Because this use of mixed-method evaluation requires each method to be sequenced, one after the other, these methods are often incorporated into mixed-method evaluations using sequential processes. Again, this design choice has time and resource implications, as discussed below.

HOW TO MIX METHODS

As mentioned above, evaluators must consider carefully how they will integrate the different methods used into a coherent, thoughtful evaluation design. This section outlines three of the most common ways in which

methods can be combined to accomplish the purposes described above: parallel combinations, sequential combinations, and multi-level combinations. (Table 2, at the end of this document, lists two additional techniques: conversion and data synthesis; for purposes of this note, however, the three techniques discussed in the text are sufficient to provide an overview of the issues.) A single evaluation might use more than one, or even all, of these combination patterns at different points during the evaluation process.

Each of these variations involves important management decisions on issues such as which data sources to rely on and how to gain access to them, what sites should be included in the evaluation (all project sites or a sample, and if just a sample, how the sample should be drawn), how much depth of explanation is needed, which stakeholders' views are most relevant, and so on. These kinds of questions normally need to be addressed in the SOW so that the evaluators can develop designs that answer the questions of interest. And while the evaluators may help to clarify questions and provide guidance on practical implications of design choices, it is the responsibility of the USAID managers to have made the key decisions, such as what questions need to be answered and what kinds of evidence are needed, when developing the SOW. These choices have important implications for time and resources. Mixed-method evaluations must be carefully planned with a thorough understanding of why and how each method is to be used to answer the questions. One must carefully consider the integrity of the design of the evaluation. Adding more methods to a design with the idea that "more is better" can lead to unnecessary complications and cost.

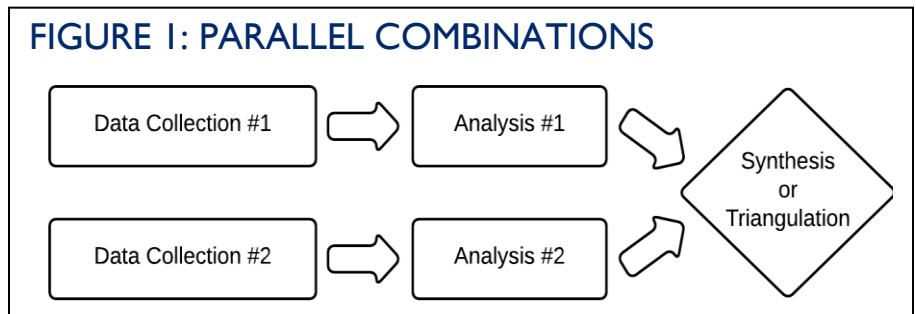
PARALLEL COMBINATIONS

In parallel combinations, methods are used separately and the findings are integrated after the data are analyzed. The same evaluation team might be involved in implementing multiple evaluation methods, and the actual data collection and analysis can happen over the same period of time, or at different times. **The key point is that in parallel combinations, each method is conducted in its entirety, separately from the other methods,** as shown in Figure 1.

For example, a team evaluating an alternative development project could collect data from government statistics on the number of acres of land converted from illegal to legal crops, and then analyze this quantitative data to estimate the impact of the project. At the same time, the same team also might conduct individual and focus

group interviews with farmers to better understand their choice of what to grow. These two methods could take place simultaneously, and the data analyzed separately. Then, if the data from the two methods were intended to answer the same question, the findings could be triangulated. If they were intended to answer different questions, then the results would be combined, or synthesized, in the evaluation report.

FIGURE 1: PARALLEL COMBINATIONS



SEQUENTIAL COMBINATION

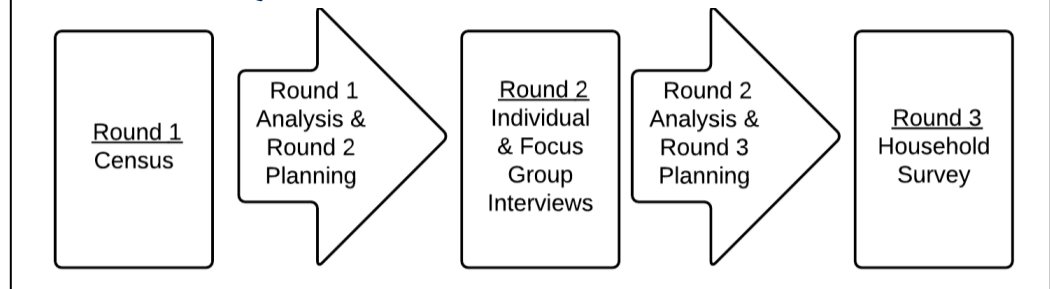
Multiple methods also can be used at different times and in a specific order. **With sequential combinations, methods are employed one after the other, with the findings from methods used earlier in the evaluation informing the design and implementation of methods used later in the evaluation.**

Round 1 data collection:

Based on the evaluation design, the evaluator constructs data collection instruments, conducts a first round of data collection, and analyzes that data. For example, as illustrated in Figure 2, if the evaluation is intended to determine

whether a water, sanitation and hygiene (WASH) project is leading to higher rates of hand-washing in a particular community, the evaluator may first collect and analyze quantitative data from the most recent census to gather information about the population of the community and any relevant demographic characteristics, such as age, socio-economic status, and ethnicity.

FIGURE 2: SEQUENTIAL COMBINATIONS



Use of Round 1 findings to inform Round 2 data collection: The evaluator then can use the findings from the first round of data collection to inform the second round. In this example, the census data would help the evaluators identify the demographic characteristics that should be represented among the second-round interviewees. These interviews would help bring to light the common reasons why the affected population chooses to wash their hands, or not, and what obstacles may exist that prevent them from doing so.

Use of Round 1 and 2 findings to inform Round 3 data collection: Based on qualitative information collected during the interviews, the evaluator would design the content of a household survey. This survey would help to answer the central question of what proportion of villagers has changed their hand-washing behavior, as well as why their behavior changed. The household survey would be conducted in the community, possibly following a sampling strategy based on Round 1 quantitative data. The data collected during the Round 3 household survey would directly address the original evaluation questions. While it might be possible to do only a survey, the mixed methods approach would have a number of advantages: thanks to the Round 1 analysis the sample would be more representative of the total population of the village, and the survey questions would be more appropriately tailored to the local context and to the diversity of the community thanks to the Round 2 analysis.

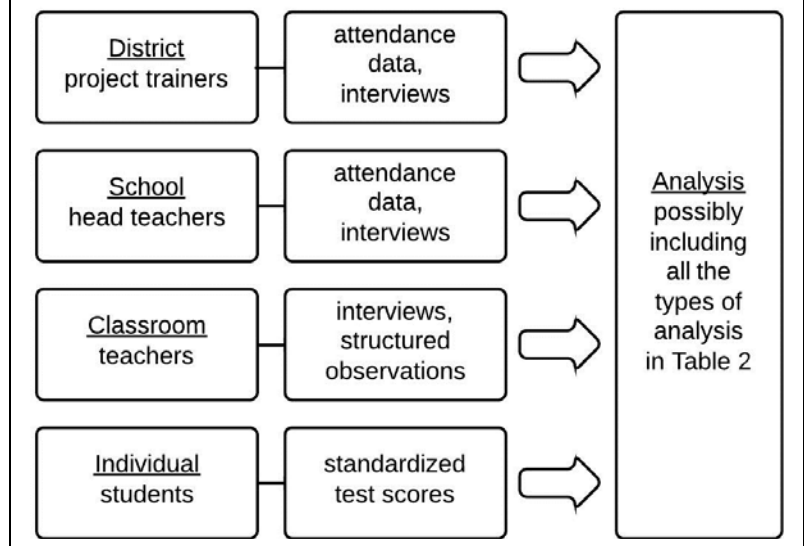
MULTILEVEL COMBINATIONS

Many projects involve systems with multiple levels, and the evaluators of these projects often must collect data and draw conclusions about each of these levels in order to have a clear understanding of the overall performance of the project. Not surprisingly, the evaluator may conclude that different kinds of methods are best suited to collecting and analyzing information from different levels.

For an example of this type of evaluation, consider an education project which is intended to raise student literacy by introducing more effective teaching strategies in a set of project schools. The project is designed so that the head teacher from each project school is trained by project trainers at the district level, and these teachers then return to their own schools to train their fellow teachers.

To understand whether the project has been effective, the evaluator will need to collect data on student literacy, likely based on the scores on a standardized test. But to understand *why* the project has been effective or not will also require data from additional levels. As shown in Figure 3, the evaluator may also want to collect attendance data from the trainings that occurred at the district level and conduct in-depth and focus group interviews with the project trainers responsible for that training. At the school level, the same type of data could be collected from both the teachers who received the district training, and from those other teachers who received the trainings in their own schools. The evaluator may also conduct structured observations of classrooms in project schools in order to see firsthand whether teachers are using the literacy techniques promoted by the trainings. Incorporating different types of data at these different levels provides the evaluator with a more complete, holistic understanding of how the project operates and how it achieved, or did not achieve, its goals.

FIGURE 3: MULTILEVEL COMBINATIONS



Multi-level mixed-method evaluations can be combined with either parallel or sequential processes, or a combination of the two, depending on the levels of the project and the purpose of the evaluation. In the example above, there is a parallel process embedded in the multi-level design. However, integrating different types of data at different levels like this does make the overall evaluation design more complex, and requires additional planning, coordination, and management of the evaluation to make sure that all the data collected are analyzed and incorporated into the final report. While much of this responsibility necessarily falls to the evaluators, the implication for managers is that they must provide careful oversight of this complex evaluation process to ensure that the efforts are timely and well-coordinated, and carried out within time and budget constraints. Thus, from the manager's point of view there is a trade-off between the quality of these mixed-method evaluations in terms of accuracy and completeness, and the time and resource costs necessary to carry them out.

INTEGRATING MULTIPLE METHODS INTO THE EVALUATION

In order to get the most out of a mixed-method evaluation, the evaluation manager and the evaluator must consider carefully what purpose each method is intended to fulfill, and how they will be combined, at each stage of the evaluation, to most efficiently and effectively accomplish these purposes.

PLANNING THE EVALUATION

In order to conduct a successful mixed-method evaluation, the evaluator must start at the planning stage. At this point, the evaluator or evaluation manager must decide which methods to use and how to combine them. These decisions will be based primarily on the purpose of the evaluation and the key evaluation questions, but evaluation managers must also take into account factors such as time and cost.

When drafting the evaluation SOW (Scope of Work), the evaluation manager must first decide which questions need to be answered and how rigorous the evaluation needs to be. If the evaluation is intended to estimate impact that can be attributed to a specific intervention, then the evaluation will include some kind of experimental or quasi-experimental design which typically includes a control or comparison group. This

decision must be made at the planning stage of the project so the project design and management will allow for this kind of evaluation design. If the evaluation is intended to answer questions more oriented toward project management, then non-experimental designs likely will be sufficient. In cases in which an evaluation needs to answer both attribution-related and other types of questions, the evaluation manager may choose an evaluation that incorporates aspects of both experimental (or quasi-experimental) and non-experimental designs. If at the planning stage the managers decide that the evaluation requires mixed-methods, they need to be aware that it will require careful management throughout the evaluation to ensure they accomplish their dual purposes.

To see how combining these kinds of methods, or evaluation designs, can strengthen an evaluation, imagine a project that aims to increase farmers' income by training them in improved farming techniques. The project design stipulates treatment and control groups in order to allow an evaluation to measure the contribution of the project to increased farmer income. In this case, the evaluation manager might decide on a quasi-experimental design with a treatment group, who receive the training, and a control group, who do not. At the same time, the evaluation also may aim to determine whether the project was implemented according to its original design, and this part of the evaluation may rely on a non-experimental design, focusing on a description of project operations, to fulfill this aim. The quasi-experimental component of the evaluation may reveal no significant difference in income between farmers in the treatment group and those in the control group, which would indicate that the project had no impact. Some stakeholders may also conclude from this finding that training projects of this kind do not increase farmer's incomes. At the same time, the results of the non-experimental component of the evaluation may reveal that only a small number of farmers in the treatment group actually received the intended training, or that they all did receive the training and then shared the improved farming techniques with the farmers in the control group. In either case, this additional information would help to explain the findings of the quasi-experimental part of the evaluation: it is possible that training of this type can lead to increases in farmers' income, but because of faulty implementation, it is impossible to know for sure from this evaluation.

CHOOSING THE RIGHT EVALUATION TEAM

Mixed-method evaluations require specialists with different kinds of skills. Even if an evaluator can construct a statistical regression model, this does not mean that the same person will be able to conduct a structured observation or effectively facilitate a focus group interview. It is important to consider during the planning stage what kind of skills will be needed in order to conduct each aspect of the evaluation successfully, and then to select team members accordingly.

Once the evaluation team has finalized the overall evaluation design, the next step is to plan for data collection. In some cases, the evaluator may choose to begin using multiple data collection methods even during the planning stage, to help design tools or methodologies that then will be used during the data collection or data analysis stages of the overall evaluation. This use of mixed methods follows the sequential pattern of combining methods described above. A review of relevant literature and secondary data is one method commonly employed at the planning stage. Stakeholder consultations, also a common part of the planning stage, can take the form of individual and focus group interviews. The evaluator also could decide to conduct a mini-survey of some of the stakeholders to quickly get data from a larger group of people than is possible with in-depth interviews. The range of methods selected during the planning stage will depend in part on what data is already available.

For example, if the data the evaluator needs are available for the whole population in secondary data sources, such as government statistics, then there may be no reason to do a mini-survey. Alternatively, it may be beneficial to conduct focus group interviews with certain groups of stakeholders, but there may not be enough time before the start of the evaluation to do so. The evaluation manager must think carefully about the information needed to plan the rest of the evaluation, how these different methods would contribute to this

planning, and how best to spend the resources available for the evaluation. Using a mix of methods at the planning stage can greatly strengthen evaluations, and if appropriate, the evaluation manager should include guidance for their use at this stage in the evaluation SOW.

Whether a mix of methods is applied during the planning stage or not, planning for mixed-method evaluation still requires the evaluation manager to approve the methods to be used, including data collection tools, data sources, sampling strategies, data management, analysis techniques, reporting techniques, and how these different pieces of the different methods will be woven together into a coherent mixed-method design that meets all the requirements of the evaluation. One tool that can be helpful during the planning stage is an evaluation design matrix like the one in Table I. Such a matrix, completed by the evaluation team as part of the planning process, can assure the evaluation manager that the team has considered what it needs to do to answer the evaluation questions, and has a plan for carrying out those tasks.

SELECTING THE APPROPRIATE MIX OF METHODS: A HYPOTHETICAL EXAMPLE

This section presents an example of how multiple methods can be applied in the context of a mid-term evaluation of a multi-year gender equity project. The project goals are to increase the number of women who graduate from secondary school. The project attempts to achieve these goals by combining scholarships and mentoring provided by the host government's Ministry of Education (MoE) with support and training from the project staff.

The evaluation questions ask:

- (1) Has the project been successful in meeting its targets for number of scholarships distributed and number of mentors connected with scholarship recipients?
- (2) In the targeted secondary schools, did the project increase the number of female graduates?
- (3) What changes could be made to the project to increase its effectiveness?

TABLE I: SAMPLE MIXED-METHOD EVALUATION DESIGN MATRIX

Q#	Data collection method	Data collection instrument	Sample questions on the instrument	Data source
1	Desk review	Annotated bibliography	NA	Project managers
1	Data quality review	Checklists, project reports	NA	Project staff
2	Review of School statistics	School data form	Enrollment records by year, grade, and gender	Project staff, government officials, school administrators
2	Semi-structured interview	Interview protocol	What are the reasons you stayed in or left school?	Students, teachers, mentors
2	Focus group interview	Focus group protocol	How do you decide whom to send to school?	Parents
3	Semi-structured interview	Interview protocol	What worked? What did not work?	Project staff, government staff, school administrators, students, teachers, mentors
3	Focus group interview	Focus group protocol	What worked? What did not work?	Students, teachers, parents, mentors
3	Survey	Survey instrument	Did the project receive enough money?	Project staff, government staff, school administrators, Students, teachers, mentors

The first question deals with contractual targets, while the second aims to test the primary theory of the project—namely that scholarships and mentoring for girls leads to increases in female secondary student graduation rates. The third question asks evaluators to determine where there are weaknesses in the project and suggest ways to strengthen them. Answering these three questions will require multiple research methods. By first examining each question individually, the evaluator will then be able to choose methods to collect all the data necessary to answer all of the questions.

To answer the first question, a document review of quarterly and annual reports could establish the up-to-date numbers of scholarships granted and of mentors trained. These could be cross-checked through examinations of project records at multiple sites, including at government offices, at the project head office and field offices, and at target schools. The quality of the management information system and monitoring processes is a central concern when answering this type of question.

To answer the second question, the evaluation manager must decide how rigorous the research should be, given time, resources, and conditions in the field. If available, multi-year enrollment records for a sample of the targeted schools from before and during the project could be collected using student records. These data could then be compared to relevant regional and national trends, or to a more systematically identified comparison group in an impact evaluation design to more confidently address attribution of outcome changes to the project. Evaluators also could conduct interviews with students, teachers, and parents to triangulate the findings from the project reports, and also to see if any other factors influenced decisions about schooling for girls. A deeper level of analysis would examine the retention and graduation data for the entire school population—scholarship and non-scholarship students—and compare them. Non-scholarship recipients and their families could also be interviewed.

To answer the third question, evaluators must determine among all the stakeholders involved with the project who should be consulted. Evaluators likely would conduct semi-structured interviews with key project staff, teachers, mentors, students, and their parents. A survey tool could be developed as well and administered to select groups of beneficiaries to rate multiple aspects of the project.

TAKING INTO ACCOUNT PRACTICAL CONSTRAINTS

In many cases, the ideal mix of methods will not be possible, either due to constraints on the evaluation itself, such as limited time or funding, or due to contextual factors, such as a challenging geographic or political environment. All evaluations must deal with these types of constraints, but they are especially relevant in deciding among different methods and ways to integrate them into a mixed-method design. In some cases a mixed-method approach can help overcome some of these obstacles. For example, if the evaluation ideally would include a large-scale household survey, but this method is too costly or time-consuming, an analysis of existing census data could be used to select a small purposive sample of informants to interview that included members of all the relevant social groups. Or if a certain portion of the population of interest is inaccessible due to security concerns, statistical matching techniques based on secondary data can be used to identify a similar, alternative group to include in the evaluation sample.

After examining the questions, it is apparent that either a parallel or multi-level approach can work. The questions could be investigated separately but simultaneously, and then the findings compared in the analysis stage. All three questions, however, share some common populations and sites, and so it is likely more efficient in this case to use a multi-level design. The mix of different tools that would be used in this evaluation model—project statistics, school data forms, semi-structured interviews for four different populations, focus group protocols, and a survey—would be influenced by the location and population of the different research sites, in this case sites at the federal and state government, schools, and households.

DATA COLLECTION

Having designed the evaluation and identified data needs and sources, the next step is to carry out the data collection strategy. Mixed-method designs usually require multiple data collection methods to accommodate the different kinds of data needed to carry them out. This can be complex, time-consuming, and costly. However, with careful planning, it may be possible to design data collection instruments and sampling methodologies so that different kinds of data can be collected with a relatively small additional investment of time and other resources.

For example, if an evaluation of an education project requires the evaluator to administer a short, standardized test to a sample of students in project schools, it might be possible to add a small number of closed-ended or open-ended survey questions to the end of the test, asking, for example, what the student had for breakfast that morning, so the evaluation also could examine how student nutrition affected educational outcomes. The evaluator could also coordinate with school officials to schedule a number of classroom observations, a focus group interview with teachers from the school, and an in-depth interview with the principal, all on the same day. In this way, several different types of data can be collected in only one visit, minimizing the disruption to the students and teachers, as well as travel time and transportation costs. Another example is a household survey which, in addition to asking questions of family members, also instructs enumerators to observe and note characteristics such as the building materials the house is made of,

which often correlates with family wealth. With some forethought and creativity, data collection instruments can be systematically integrated to help capture a broad range of data effectively and efficiently.

DATA ANALYSIS

It is important to design the data analysis strategy before the actual data collection begins because the way that the data will be analyzed and used can affect both the choice and the order of data collection methods. Analyzing data collected from a mixture of methods is often more complicated than analyzing the data derived from one method, as the evaluator must integrate



© Photo by Amer Sawalha

multiple data analysis methods in order to determine and understand key findings. There are several different general techniques to analyze data from mixed-method approaches, including parallel analysis, sequential analysis, and multilevel analysis, corresponding to the patterns of combining methods described above, as well as conversion analysis and data synthesis. Table 2 briefly describes these different analysis techniques and the situations in which each method is best applied. More specific data analysis techniques, such as key word coding or theme analysis for qualitative data, and cross-tabulations or regression modeling for quantitative data, can also be used within the framework of any of these more general data analysis techniques. As with all evaluations, the choice of analytical techniques depends on the purpose of the evaluation and the type of data involved, as well as time and resources available.

When Findings Do Not Converge

In cases where mixed-method evaluations employ triangulation, it is not unusual that findings from the separate analysis of each data set do not converge to support the same conclusions. If this occurs, the evaluator must try to resolve the conflict among divergent findings. This is not a disaster. Often this kind of situation can present an opportunity to generate more nuanced explanations and important additional findings that are of great value.

One method to use when findings from different methods diverge is to carefully re-examine the raw qualitative data through a second and more in-depth content analysis. This step is taken to determine if there were any factors or issues that were missed when these data were first being organized for analysis. The results of this third layer of analysis can produce a deeper understanding of the data, and can then be used to generate new interpretations. In some cases, other factors external to the project might be discovered through contextual analysis of economic, social, or political conditions or an analysis of operations and interventions across project sites.

Another approach is to reanalyze all of the disaggregated data in each data set separately, by characteristics of the respondents as appropriate to the study, such as age, gender, educational background, socio-economic status, or locale of respondents. The results of this analysis may yield other information that can help to resolve the divergence of findings. This further analysis will provide additional explanations for the variances in findings. While most evaluators build this type of disaggregation into the analysis of the data during the design phase of the evaluation, it is worth reexamining patterns from disaggregated data.

Sometimes data quality issues, such as the validity of secondary data sources or possible errors in survey data from incomplete recording or incorrect coding of responses can cause dissonance in results. If the evaluators are still at the project site, it is possible to resolve data quality issues with limited follow-up data collection by, for example, conducting in-depth interviews with key informants.

The data analysis approaches noted above underscore the need for USAID evaluation managers to plan adequate time and resources for data analysis in evaluation SOWs, as it is a critical component of any evaluation, and perhaps even more so in one using a mixed methods approach.

Dealing with Divergent Findings: An Example of Triangulation

Imagine an evaluation to assess the impact of a school feeding project on school attendance rates. School records showed that daily enrolment had increased by ten to fifteen percent after the school feeding project began. A review of records from other local schools without the feeding project did not find any similar increase. A household survey before and after the project detected only a much smaller increase. Interviews with key informants provided different opinions. NGOs involved with the project, or who ran other school feeding projects, reported the project had been successful, whereas several other informants said that the school might deliberately over-report the increase to convince sponsors to continue the feeding project. How would triangulation be used to obtain the best estimate on the basis of this conflicting information?

First, the evaluators responsible for managing each kind of data collection would meet to explore possible explanations for the differences and to understand any possible sources of bias in the data. How reliable are attendance records? Does someone check names each day while the students are there, or does the teacher try to recall at some later point in the day after the students have gone home? Is there any reason why schools would intentionally inflate the number of children attending? When is attendance recorded? If it is at the start of the day, might some children just come for the breakfast and then leave – but be reported as attending? Are there any reasons that parents might misreport the number of their children attending school? Might key informants have any reason to over- or under-estimate the effects of the school feeding project?

Second, if this discussion does not fully explain the differences, as it probably would not in this case, the evaluators might agree on the additional kinds of data that would help clarify the situation and which would be feasible to collect. The most obvious approach would be to examine in more detail the school attendance records to address the questions mentioned above.

Third, a return to the field would be organized (where feasible). When revisiting the schools, the evaluators also would compare the attendance records with the number of children in the classroom after the breakfast had ended. They also might identify some of the children who were not in school but whose parents had reported in the survey that they attended regularly. Some of the families might then be revisited to check on the consistency between parents' reports of attendance and actual school attendance. They would then conduct informal interviews to try to understand reasons for the discrepancies.

These follow-up returns to the field are extremely valuable, but they can only take place if time and resources have been budgeted in advance. Consequently, it is recommended that, the evaluation budget should include time and money for this purpose as these kinds of inconsistencies are very commonly detected during data analysis.

REPORTING

Through the use of mixed-method evaluations, findings and conclusions can be enriched and strengthened. Yet there is a tendency to underuse, or even not to use, all the data collected for the evaluation. Evaluators can rely too heavily on one particular data source if it generates easily digestible and understandable information for a project manager. For example, in too many cases quantitative data that can be presented easily in graphs or tables are emphasized, whereas possibly more important but harder to present data generated from qualitative methods are insufficiently analyzed and reported. Evaluation reports should strive for interesting graphical presentations of findings from qualitative data as well.

One way to prevent underutilization of findings is to write a statement of work that provides the evaluator sufficient time to analyze the data sets from each method employed, and hence to develop valid findings, explanations, and strong conclusions that a project manager can use with confidence. Additionally, statements of work for evaluation should require evidence of, and reporting on, the analysis of data sets from **each** method that was used to collect data, or methodological justification for not having included analyses from any data sources used.

TABLE 2: TECHNIQUES FOR ANALYZING MIXED-METHOD DATA¹

Type	Analysis Technique	Analytical Output
Parallel	Two or more data sets collected using a mix of methods (quantitative and qualitative) are analyzed separately. The findings are then combined or integrated.	Triangulating designs to look for convergence of findings when the strength of the findings and conclusions is critical, or to use analysis of qualitative data to yield deeper explanations of findings from quantitative data analysis.
Conversion	Two types of data are generated from one data source beginning with the form (quantitative or qualitative) of the original data source that was collected. Then the data are converted into either numerical or narrative data. A common example is the transformation of qualitative narrative data into numerical data for statistical analysis (e.g., on the simplest level, frequency counts of certain responses).	Extending the findings of one data set, say, quantitative, to generate additional findings and/or to compare and potentially strengthen the findings generated from a complimentary set of, say, qualitative data.
Sequential	A chronological analysis of two or more data sets (quantitative and qualitative) where the results of the analysis from the first data set are used to inform the analysis of the second data set. The type of analysis conducted on the second data set is dependent on the outcome of the first data set.	Testing hypotheses generated from the analysis of the first data set.
Multilevel	Qualitative and quantitative techniques are used at different levels of aggregation within a study from at least two data sources to answer interrelated evaluation questions. One type of analysis (qualitative) is used at one level (e.g., patient) and another type of analysis (quantitative) is used in at least one other level (e.g., nurse).	Evaluating cases where organizational units for study are nested (e.g., patient, nurse, doctor, hospital, hospital administrator in an evaluation to understand the quality of patient treatment).
Data Synthesis	A multi-step analytical process in which: 1) a rating of project effectiveness using the analysis of each data set is conducted (e.g., large positive effect, small positive effect, no discernible effect, small negative effect, large negative effect; 2) quality of evidence assessments are conducted for each data set using “criteria of worth” to rate the quality and validity of each data set gathered; 3) using the ratings collected under the first two steps, develop an aggregated equation for each outcome under consideration to assess the overall strength and validity of each finding; and 4) average outcome-wise effectiveness estimates to produce one overall project-wise effectiveness index.	<p>Providing a bottom-line measure in cases where the evaluation purpose is to provide a summative project-wise conclusion when findings from mixed-method evaluations using a triangulation strategy do not converge and appear to be irresolvable, yet a defensible conclusion is needed to make a firm project decision.</p> <p>Note: there may still be some divergence in the evaluation findings from mixed data sets that the evaluator can still attempt to resolve and/or explore to further enrich the analysis and findings.</p>

¹ See Teddlie and Tashakkori (2009) and Mark, Feller and Button (1997) for examples and further explanations of parallel data analysis. See Teddlie and Tashakkori (2009) on conversion, sequential, multilevel, and fully integrated mixed methods data analysis; and McConney, Rudd and Ayers (2002) for a further discussion of data synthesis analysis.

ADDITIONAL RESOURCES

The following resources can be used as samples or templates, or to provide more information on the topics reports and on evaluation in general. Some other resources exist but are out-of-date with current USAID guidance. Where information differs, the USAID Evaluation Policy and the USAID ADS (Automated Directives System) 200 series take precedence over other resources.

Bamberger, Michael. *Integrating Quantitative and Qualitative Methods in Development Projects*, World Bank Publications, 2000.

Greene, Jennifer C. 2007. *Mixed Methods in Social Inquiry*. Sage Publications.

Journal of Mixed Methods Research. Sage Publications.
<http://mmr.sagepub.com>.

Mark, Melvin M, Irwin Feller and Scott C. Button. "Integrating Qualitative Methods in a Predominantly Quantitative Evaluation: A Case Study and Some Reflections." *Advances in Mixed-Method Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms*. Greene and Caracelli eds. *New Directions for Evaluation*. Jossey-Boss Publishers, No. 74 Summer 1997. pp. 47-59.

McConney, Andrew, Andy Rudd and Ayres Robert,. "Getting to the Bottom Line: a Method for Synthesizing Findings Within Mixed-Method Program Evaluations." *American Journal of Evaluation*. Vol. 3. No. 2. 2002. pp. 121-140.

USAID. Automated Directives System (ADS), Chapter 203, Assessing and Learning.



USAID
FROM THE AMERICAN PEOPLE

VERSION 1.0 | NOVEMBER 2013

TECHNICAL NOTE

Evaluative Case Studies

Monitoring and Evaluation Series

This Note provides guidance on using case studies in USAID evaluations.

Technical Notes are published by the Bureau for Policy, Planning and Learning and provide key concepts and approaches to USAID staff and partners related to the Program Cycle.

INTRODUCTION

Evaluative case studies can be used as a valuable method or complement to other methods for evaluating USAID development projects and activities. This Technical Note provides practical information to assist evaluation managers and development practitioners to manage evaluations employing case studies, including guidelines for identifying when a case study has been conducted well and used effectively.

DEFINITION

According to the widely-used U.S. Government Accountability Office (GAO) definition:

“Case study as an evaluation method is a means of learning about a complex instance, based on a comprehensive understanding of that instance obtained through extensive description and analysis of that instance taken as a whole and in its context”.

This definition involves three key elements of interest for USAID evaluations. First, evaluative case studies involve in-depth description and analysis of a **particular instance or aspect** of a program, project, or other development activity. It is this focus on the particular that separates case studies from other methods, which usually try to understand what happened with many, if not all, of the people or sites involved. Second, the aim of case studies is an **in-depth** understanding of the particular case or cases being studied. In other words, a case study is not a quick site visit and brief report of impressions, but requires enough time on the ground, observing, talking to people, and collecting other data to gain a detailed picture of the project being evaluated. Third, this understanding must take account of the case **as a whole** and its surrounding context, not just specific, selected pieces of the case. The objective is to understand at a deep level what is happening in a particular place and why.

RATIONALE

Case study is a frequently used evaluation method in international development. Its appeal for both managers and evaluators includes:

1. *Flexibility of use:* Case studies can be used in a variety of forms, and in a variety of settings. For example, a case study might be undertaken to illustrate the benefits of a national environmental protection project or to investigate how local cultural beliefs affected the impact of a community immunization program.
2. *Efficiency:* Compared with many other evaluation methods, case studies can be relatively low-cost. Travel to project sites is necessary, but a case study does not require the same amount of preparation time as a large-scale survey, for example and can be completed relatively quickly.
3. *Dealing with multiple interventions:* Many development projects involve combinations of several component interventions. Case studies can be used to investigate and describe how a set of interventions works together (or not) to effect changes.
4. *Addressing context:* Where a project or activity was implemented can matter just as much as *how* it was implemented. Case studies examine a project or activity in relation to its surrounding context.

DECIDING TO USE AN EVALUATIVE CASE STUDY

Whether the case study is an appropriate method for a given evaluation depends on the main evaluation questions of interest. The evaluative case study is best used when the major questions are “how” or “why” questions. For example, the questions might ask how a complex intervention has been implemented; why the intervention has achieved or not achieved its intended (or unintended) effects; or why the effects are not larger or smaller. On the other hand, if the evaluation questions aim to answer “how many” or “to what extent” questions, and need to answer these questions about the whole population involved in the program, then the evaluative case study will likely be unable to answer these questions by itself. A case study would not be the best evaluation method to answer questions such as “How many farmers were trained?” or “To what extent are farmers implementing the techniques on which they were trained?” If the question is “Why do farmers choose to use the techniques promoted by the training?”, then it may be useful to include an evaluative case study as part of the evaluation design.

Time and cost considerations also may affect the decision to adopt the case study method. Case studies often can be designed faster, and at lower cost, than other methods, such as surveys or other forms of data collection. This decision will depend on the specific circumstances of the evaluation, of course, but if the questions are amenable to case study, it frequently is the lowest-cost, most timely option.

NOT AN EVALUATIVE CASE STUDY

Teaching case study: Used widely today in law, medicine, and education, as well as other fields, the intent of the teaching case study is to establish a framework for debate and discussion among students. A case is presented that illustrates a particular problem and need for a decision on a course of action. Students discuss possible courses of action and make a recommendation that may be compared to the actual decision. These types of case studies help to teach a concept, not evaluate a project or program.

Site visit: Site visits are associated with the evaluative case study method, but a “fly-through” site visit is not an evaluative case study. While there is no set rule about how long one has to be at a site for it to be a case study, it must be long enough to achieve a comprehensive and in-depth understanding.

USE THE CASE STUDY METHOD WHEN:

- The evaluation question is a “how” or “why” question.
- The evaluation focuses on an ongoing intervention or one recently completed.
- There is sufficient time on-site for in-depth data collection.
- There is no need for a statistically representative sample of a larger population.
- The intervention is complex and cannot be clearly separated from the context in which it is embedded

CONSIDER OTHER EVALUATION METHODS WHEN YOU:

- Are focused on “how many,” or “to what extent” questions
- Are doing a retrospective evaluation of an intervention that ended a considerable time ago
- Do not have much time for on-site data collection
- Need to generalize findings statistically from a sample to a population
- Can isolate the variables of interest

DETERMINING THE TYPE OF CASE STUDY

The following paragraphs describe the four main types of case studies. The type of case study used for a given evaluation will depend on the key questions to be addressed.

- **The illustrative case study:** This type of case study is descriptive in character and usually is intended to add realism and provide in-depth examples to supplement information gathered through other evaluation methods such as the survey method. The illustrative case study may describe a typical case, best cases, largest cases or the like. It might ask a question such as how a large urban secondary school is implementing the national school reform program.
- **The exploratory case study:** Also descriptive in nature, this type of case study is used to identify issues for later broader investigation. They are often done as part of mixed-methods evaluations particularly to inform the content of a large-scale survey when little is known about a project’s implementation. Exploratory case studies might pose questions such as: “What issues or challenges have surfaced in the program’s implementation in highly rural areas?”
- **The critical instance case study:** Again, descriptive in nature, this type of case study examines in detail an instance that is unusual or unique. It might ask a question such as, “Why are HIV/AIDS rates falling in a particular country when rates in neighboring countries are rising?”
- **The explanatory case study:** There are two types of explanatory case studies. The first focuses on program implementation. This type of case study investigates how a program operates, either at one or multiple sites. A program implementation case study might ask questions such as, “In early implementation of the voucher program, what program aspects are working well or less well and why?” The second type of explanatory case study focuses on program effects. This type of case study examines the outcomes and impacts—both positive and negative and planned and unplanned—of an intervention and seeks to provide explanations by making causal inferences about the reasons for success or failure. A program effects case study might ask questions such as: “How has a women’s microenterprise project increased the economic wellbeing of participants in the three selected villages?” or “Why does an inner-city school, serving the children of low-income families, have the highest science achievement scores of all schools in the province?” Explanatory case studies may include multiple sites, but their findings are not generalizable. (See below, Generalizing from Evaluative Case Studies.)

DETERMINING THE UNIT OF ANALYSIS

The unit of analysis refers to the entity at the center of each case study, and choosing the unit of analysis is an important step in the evaluative case study process and requires careful consideration. At its most fundamental, choosing the unit of analysis means deciding whether each case study will focus on an individual (a student, an entrepreneur), a small group (a family, a team), a community (a village, a nation), or some other category. Case studies may also be done on events or entities other than individuals, such as decisions, organizations, or critical incidents. A clearly defined unit of analysis will also include geographic and time boundaries. For example, an evaluation manager may decide to include only cases of people living with HIV/AIDS currently participating in a health program in a particular province, or may choose the unit of analysis to be refugee families displaced by a drought that has gone on for the past two years, though they may be currently living in a number of different places.

In any case, the evaluation questions will always guide the choice of the unit of analysis; the evaluation manager should make sure that the information collected based on the unit of analysis will enable the evaluator to answer the evaluation questions posed. It is also possible that the evaluator and evaluation manager may decide to modify the unit of analysis during the process of the evaluation, based on their evolving understanding of the situation or changing realities on the ground. When making these decisions, they should always keep in mind how the data will be analyzed. For example, cross-case analysis requires that the units of analysis are consistent across cases. If part of the purpose of the evaluation is to respond to prior research, the evaluation manager may choose a unit of analysis corresponding to the unit of analysis used in this existing literature.

Clearly outlining the unit of analysis and the reasons for choosing it strengthens the systematic and purposeful nature of the evaluative case study, as well as its overall transparency. At the same time, it also contributes to a key strength of the evaluative case study, its ability to capture both characteristics of the unit itself and also the context in which it exists. In other words, though it is crucial to have a clearly defined unit of analysis, this clear definition does not invite or compel the evaluator to take the unit of analysis out of its context and ignore important factors that emerge through the course of the data collection process (but that may not have been explicitly included in its initial definition). In fact, it is only by clearly delineating the boundaries between the unit of analysis and its context that the evaluator can begin to identify and place the various factors acting within and on the unit of analysis.

CASE STUDY SAMPLING TECHNIQUES

There are many different ways to select cases for an evaluative case study. Some of the most common are outlined below, but there are others, and the final decision will always depend on the purpose of the evaluation, as well as the budget and the time available.

SINGLE ILLUSTRATIVE CASE

If the purpose of the evaluative case study is illustrative, even one case study might be sufficient. With survey data on the physical condition of a nation's rural health clinics, for example, an evaluation team could identify the characteristics of the typical clinic and then conduct a single illustrative case study of a typical clinic, the one which best meets those characteristics.

THREE MYTHS ABOUT CASE STUDIES

Case studies take a long time.

While case studies as an evaluation method are not rapid appraisals, they may take only weeks rather than months or years. A case study of the nature and extent of implementation of state-wide school reform in one school, for example, may be accomplished in a few weeks. The key is whether the requirements of the case study definition—a comprehensive understanding of the implementation of the intervention in the school, taken as a whole and in its context—have been met. A rapid appraisal methodology would not likely be compatible with case study methodology.

Case studies only use qualitative methods.

It is true that qualitative methods are predominant in case study evaluations, as evaluative case studies seek to obtain understanding through “extensive description,” but quantitative methods also can be used. Many case study evaluations use a combination of qualitative and quantitative methods.

Case studies cannot be used to generalize or make causal inferences.

Evaluative case studies should not be used to make statistical generalizations from a case to a population of interest, but they can modify a generalization, including those related to causal inferences, by refuting it (it only takes one case) or by providing more detail about it. They also can add positively to a body of evidence supporting a theory behind an intervention. (See below GENERALIZING FROM EVALUATIVE CASE STUDIES.)

BRACKETING

In contrast, if the purpose of the evaluation is to better understand the conditions under which a project works or fails, the evaluator might select extremes cases, a strategy sometimes called bracketing. For example, if the initial results from survey data or background research showed that some entrepreneurs benefited greatly from an economic growth project, while some did not seem to benefit at all or were actually harmed by the project activities, the evaluator might choose to conduct case studies of one or two of the most successful participants, as well as one or two of the least successful participants. This sampling strategy likely would provide both managers and evaluator with a better understanding of the success factors and why it was sometimes effective and sometimes ineffective, than a single illustrative case.

EXISTING GROUPS

Often, groups will already exist within the context of the evaluation, and the evaluator may have good reason to believe that these groups have experienced the project differently. In these cases, it often is important to capture the different experiences of the various groups, and the evaluator may choose one or more cases from within each group. At the level of the individual, these groups could be based on sex, age, or ethnicity, to name a few of the most common groupings. The evaluator, therefore, might choose to do one illustrative case study of an adult male beneficiary and contrast it to one of an adult female beneficiary; one of a child, one of a middle-aged adult, and one of an older adult; or one of a person from each ethnic group. Alternatively, if a project was conducted in five different provinces, and the evaluator suspects that the project was implemented differently in different provinces (due to different levels of support from the provincial government, for instance), then she might design the evaluation to include a case study in each province, in order to try to capture and understand those differences.

TYOLOGY

Frequently for multi-site evaluative case studies, sites are selected based on typologies. To explain how this might work, imagine a project that promotes new legislation on providing in-school support to children with disabilities. The evaluator may determine that the wealth and the degree of urbanization of the area where a school is located have important influences on the success of programs in a given school. By analyzing all of the schools in the country based on these two characteristics, it becomes clear that there are three main “types” of schools affected by the project: wealthy urban schools, poor urban schools, and poor rural schools. The evaluator then may choose to do one case study of each of these three different school types. In reality, many more factors are likely to be considered, including size, ethnic makeup, and past

student performance, resulting in more, and more specific, school types.

PURPOSIVE RANDOM SAMPLING

Random sampling can be used to select cases for an evaluative case study, but it is important to recognize that using random sampling does not mean that the cases selected are statistically representative, or generalizable to the total population. Random sampling can, however, help reduce the possibility of bias in the selection once other purposive selection techniques no longer offer guidance. For example, if the evaluator has used existing groups or typologies to narrow down the groups from which the case study will be chosen, random selection could then be used to make the final choice. This final step helps prevent the evaluator, or another stakeholder, from inappropriately choosing the case study for unstated reasons, introducing bias.

CONDUCTING THE EVALUATIVE CASE STUDY: USING KEY CASE STUDY PRINCIPLES

There are several key practices that characterize and guide an evaluative case study that make it valuable as an evaluation approach. Some of most important are briefly explained below.

THICK DESCRIPTIONS

The goal of an evaluative case study is to obtain as full an understanding as possible, and this understanding is gained by extensive description and analysis. Through rich, extensive, sometimes called “thick” description, the case study evaluator tries to give the reader a sense of what the experience of being at the site would convey.

ASSERTIONS

At the beginning of the case study, the evaluator will develop assertions related to the evaluation questions which then help guide the direction of the evaluation moving forward. These assertions will be based at first on any background data collected, or on the first parts of the case study. The evaluator will then seek more information to refute or support these assertions, and as new information is obtained, support, modify, discard, or replace the original assertions, depending on the new information.

PROGRESSIVE FOCUSING

Throughout this process of refining key assertions, the evaluator must continually adjust the focus of the evaluation. Progressive focusing does not mean the key evaluation questions are changed completely, but the evaluator may choose to modify some of the data collection instruments, or add additional data sources, to obtain all of the information needed for a complete, in-depth understanding of the issues involved. In many cases these adjustments will mean delving more deeply into topics that the evaluator already knows are important to answer the evaluation questions. In other cases, the results from one interview or observation, for instance, may reveal new, unanticipated information. This new information may raise additional questions that the evaluator must answer in order to completely understand the project and comprehensively answer the evaluation questions. Both the evaluator and the evaluation manager recognize that a certain degree of flexibility is necessary as the case study progresses, to allow for these kinds of modifications.

USING APPROPRIATE DATA COLLECTION METHODS AND PROCEDURES FOR CASE STUDIES

Evaluators can use the full range of data collection methods to construct a case study, including desk reviews of secondary data and exiting literature, surveys, focus group interviews (see Technical Note on Focus Group Interviews), semi-structured interviews, direct and participant observation, photos and drawings and case histories. Using a combination of these methods helps to capture a rich and comprehensive picture of the project being evaluated. At the same time, the evaluator must carefully consider what type of data collection methods are best suited to capture the data necessary, and will often have to balance these needs against constraints of time, resources and access to data.

RECORDING AND ORGANIZING DATA

During the data collection stage, case studies produce a large amount of different kinds of data. In order to store and organize these data, the evaluation team must create a database. The exact format will depend on the different kinds of data contained in it, but it should be able to store and categorize all of the data collected in an easily searchable way. If possible, the evaluator should enter the data into the database as soon as possible after they are collected and make backup copies of the database whenever new data are entered. Adhering to this protocol will help reduce the risk of data loss, and also will make it easier for the data collectors to remember all of the important details that they otherwise might forget. The evaluation team also should conduct regular quality assurance checks to make sure that the data are being entered correctly and consistently. The Evaluation SOW should include a requirement that data be organized into a database that can be shared with USAID upon completion of the evaluation, along with the protocols needed to extract information from that database.

PROTECTION OF HUMAN SUBJECTS

All those responsible for planning and carrying out evaluations of development projects constantly must be aware of the potential risks to the people involved. Because case studies often are intensely personal and hands-on exercises, protecting the privacy and the rights of those involved is particularly relevant and important. Evaluators have the obligation to think through the ethics of the situation and make sure that the necessary steps are taken to respect privacy and do no harm. Evaluation managers should include an explanation of the procedures an evaluator will use to protect human subjects as part of the evaluation proposal.

ANALYZING CASE STUDY DATA

Data analysis actually begins during the data collection stage of an evaluative case study, as the evaluator goes through the continuous process of developing assertions and progressively focusing on relevant issues. Once data collection ends and the more formal data analysis begins, several key processes are likely to be involved.

CODING

Coding involves labeling or categorizing passages or parts of text (or pictures or videos, etc.) so that the data can be readily retrieved, searched, compared and contrasted. For example, during an evaluation of an agricultural project that promotes the adoption of improved farming techniques, evaluators conducting case studies may identify the availability of extension workers as a major issue in the context of the evaluation. While coding, they would label and track every instance when an informant or other data source raises the issue of the availability of extension workers. Using this database of coded data, they could then more easily identify what proportion of stakeholders raised this issue, which kinds of informants or sources raised it, and how many times, on average, these groups raised it, for example. If the database is relatively small, some evaluators may choose to code their qualitative data without the assistance of a software program. However, often the sheer volume of data involved becomes too large to rely on an evaluator, or even a team of evaluators, reading separate documents, highlighting passages, and tracking and tallying all of the references by hand. In these cases, there are a number of computer software packages (Atlas ti, Ethnograph, NVivo, to name only a few) that can help code this type of data automatically. Even with the assistance of a software program, coding and analyzing the type and amount of data that case studies often produce can take a significant amount of time, and the evaluation manager should allow for this time when constructing the SOW. Coding the data also requires an intimate knowledge of the project, the population and the context, and so the evaluators usually code the data themselves. The evaluation manager should know in advance if the evaluation team will use software to analyze the data and how data records will be transferred to USAID.

TRIANGULATION

Because case studies often incorporate data from multiple sources and multiple methods of data collection, evaluators are able to compare and contrast the findings based on these different sources and methods. This process is called triangulation, and it can help to reduce potential bias and increase confidence in the evaluation's overall findings and conclusions. If, for example, data are collected from semi-structured interviews, structured observations, and analysis of existing records and all three sources provide similar characterizations of the project being evaluated, there is less likelihood that the findings are due to chance, evaluator bias, or other unexplained factors. On the other hand, if different sources and methods lead to contradictory findings, then the evaluator must carefully consider what may have led to these findings, and may wish to reexamine the data collected, or even return to the field to collect additional data. In either case, triangulation helps to construct a more complete, more reliable picture of the project. Triangulation can happen not just among different methods and among different sources; evaluators also can triangulate results among different data collectors. If a team of enumerators helps to conduct a mini-survey, or multiple moderators are used to facilitate a series of focus group interviews, then the evaluator should also compare the responses that one data collector received with the responses that the other data collectors received, to check for any systematic differences, which could indicate some form of bias.

CROSS-CASE ANALYSIS

When an evaluation includes multiple studies of similar cases (individuals, groups, locations, etc.), evaluators will conduct two different stages of analysis. First, they will analyze each case individually to understand it as a separate entity, as described above. After that, they will conduct a cross-case analysis, comparing and contrasting the results from the different cases. This process is similar to triangulation and also yields similar benefits: greater confidence in findings and the minimization of potential bias. The difference is that while triangulation can happen within a single case study, cross-case analysis requires more than one comparable case study within a single evaluation.

GENERALIZING FROM EVALUATIVE CASE STUDIES

As described above, evaluators cannot use case studies to make generalizations about an entire population based on a statistically representative sample, as they can with surveys. However, case studies still can help to advance or refute general ideas about how a project is working or not working. For example, imagine that the evaluators of the agricultural project mentioned above wish to answer the question of whether farmers will be more likely to employ improved farming practices if access to an extension worker is increased. By designing a study including four case studies focusing on communities with no access to an extension worker, four case studies of communities where there was limited access to an extension worker, and four case studies of communities where there was complete access to an extension worker, they could then conduct a cross-case analysis of the twelve cases. If the farmers in the communities with complete access to an extension worker overwhelmingly adopted the improved farming practices, but the farmers in the communities with limited or no access to extension workers generally did not adopt these same practices, then the evaluator could present these results as support for the assertion that greater access to extension workers led to greater adoption of improved farming techniques. On the other hand, if the farmers with less access to an extension worker are just as likely, or more likely, to adopt the farming techniques as those with complete access, these findings would refute the assertion that greater access is necessary for greater adoption.

This type of evidence is fundamentally different from evidence obtained through a large-scale, statistically representative survey, and it may not, by itself, fulfill the needs of many evaluations. At the same time, a high-quality case study produces systematically collected, thoughtfully targeted data, and therefore the results of an evaluative case study represent more than just a haphazard collection of uninformed opinions. The evaluators in this case also must be conscious of factors other than access to an extension worker that may influence the

farmers' choice to adopt the farming practices, such as geography and type of crops grown, gender, class, or ethnicity. As this example illustrates, however, explanatory case studies can contribute to the process of supporting or refuting assertions.

POTENTIAL FOR BIAS

As with all kinds of evaluation methods, evaluative case studies are vulnerable to a number of types of bias.

- **Evaluator bias:** Given the evaluator's central role throughout the process of designing and conducting a case study, any bias this person may possess has the potential to influence the results, whether it is unconscious or intended. Bias may be introduced, for example, when initial assertions are formulated, when collecting the data, or when interpreting the results. An evaluator using a case study only to substantiate a preconceived position is an example of evaluator bias.
- **Selection bias:** The results of an evaluative case study are dependent upon the set of cases chosen for inclusion in the study. In most cases, evaluators choose cases based on which they believe will provide the most valuable, relevant data for the evaluation. If either the evaluator or another stakeholder chooses the cases based on any other criteria, this action may lead to biased results. To guard against this bias, evaluators should provide a clear justification for case selection, and may also employ techniques such as purposive random selection, described above, to further limit the influence or appearance of selection bias.
- **Alternative explanations:** Evaluators must also be careful to account for alternative explanations of observed phenomena. This is addressed in evaluative case studies through the process of assessing and revising assertions, as described above.

It is important to remember that these and other types of bias are a potential concern for all evaluation methods, and it is the responsibility of all professional evaluators and evaluation managers to recognize and limit their influence. By systematically collecting data and carefully documenting the analysis and rationales for all decisions made, the evaluator can help make the entire evaluative case study process more transparent and its conclusions more reliable.

REPORTING FINDINGS

The evaluation manager and evaluators should have a plan for how the results, including raw data, summaries, findings, conclusions and recommendations, will be reported and shared at the conclusion of the case study. Given the wide range and depth of case study data, clearly and succinctly presenting all of the data can be challenging. The box below provides four general strategies for organizing and presenting case study findings.

WAYS TO PRESENT CASE STUDY FINDINGS

Natural history: This type of report presents a chronological presentation of the case, similar to a narrative story.

Critical incident: The report is built around a single incident that captures the evaluative message of the report.

Thematic: These reports are organized around the key themes that have emerged from the case study, sometimes based on the key evaluation questions.

Within and between: Reports on multiple case studies should contain both the individual case study summaries ("within") and the results of the cross-case analysis ("between"). Sometimes the cross-case analysis will be included in the main body of the report, while the individual case study summaries are included as annexes.

TABLE 1: CASE STUDY QUALITY CHECKLIST

*For evaluation managers conducting a review of an evaluation draft report employing case study methods:
These case study checklist criteria should be checked in addition to broader evaluation quality criteria.*

Questions	Yes or No?
Is the rationale clearly presented for use of the evaluation case study method?	
Is it clear which type of case study has been conducted?	
Is the unit of analysis clear?	
Is the justification clear for the number of case studies conducted?	
Is the rationale provided for the selection of the specific case(s) or site(s)?	
Is the context in which the intervention is embedded described?	
Are data collection methods within the case study clearly described?	
If a multisite case study is undertaken, is there clear explanation of the within site analyses and cross-site analysis?	
If an explanatory case study is undertaken, are alternative rival explanations for findings identified and examined?	

ADDITIONAL RESOURCES

The following resources can be used as samples or templates, or provide more information on the topics reports and on evaluation in general. Some other resources exist but are out-of-date with current USAID guidance. Where information differs, the USAID Evaluation Policy and the USAID ADS (Automated Directives System) 200 series take precedence over that in other resources.

Campbell, D. (1975). "Degrees of freedom and the case study". *Comparative Political Studies* 8: 178-18.

Datta, Lois-ellin (1990). *Case Study Evaluations*. Washington, DC: U.S. General Accounting Office, Transfer paper 10.1.9.

Davey, Lynn (1991). "The application of case study evaluations." *Practical Assessment, Research & Evaluation* 2(9).

Gomm, Roger, Martyn Hammersley, and Peter Foster, Eds. (2000). *Case Study Method: Key Issues, Key Texts*. Thousand Oaks, CA: Sage.

Merriam, Sharan B. (2009). *Qualitative Research: A Guide to Design and Implementation*. San Francisco, CA: Jossey-Bass.

Miles, Matthew B., and Huberman, A.M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks, CA: Sage.

Morra, Linda G., and Friedlander, Amy C. (Undated). *Case Study Evaluations*. Washington, DC: The World Bank Operations Evaluation Department.

Patton, Michael Quinn (2002). *Qualitative Research & Evaluation Methods*. Third Edition. Thousand Oaks, CA: Sage.

Platt, J. (1992). "Case Study" in American methodological thought". *Current Sociology* 40: 17-48.

Richards, Lyn (2009). *Handling Qualitative Data: A Practical Guide*. Second Edition. Thousand Oaks, CA: Sage.

Scholz, Ronald W. and Olaf Tietje. (2002). *Embedded Case Study Methods: Integrating Quantitative and Qualitative Knowledge*. Thousand Oaks, CA: Sage.

Social Impact (2006). *Monitoring, Evaluation and Learning for Fragile State and Peacebuilding Programs: Practical Tools for Improving Program Performance and Results*.

Stake, Robert E. (1995). *The Art of Case Study Research: Perspectives on Practice*. Thousand Oaks, CA: Sage.

Yin, Robert K. (2009). *Case Study Research: Design and Methods*. Fourth Edition. Thousand Oaks, CA: Sage.

Yin, Robert K. (1994). "Discovering the Future of the Case Study Method in Evaluation Research". *Evaluation Practice*. 15(3): 283-290.



TECHNICAL NOTE

Focus Group Interviews

Monitoring and Evaluation Series

This Note provides guidance on the use of focus group interviews within evaluations.

Technical Notes are published by the Bureau for Policy, Planning and Learning and provide key concepts and approaches to USAID staff and partners related to the Program Cycle.

INTRODUCTION

Focus group interviews (FGIs, also often called focus group discussions) are a key tool for collecting data to support many USAID activities. This data collection technique is widely used in international development planning, monitoring and evaluation. They are quick, versatile and typically inexpensive, so they can be used with a broad range of stakeholder groups in a wide variety of settings. Because they allow evaluators to obtain a wealth of in-depth qualitative information, they often are combined with quantitative methods, such as surveys, to help explain the “why” or the “how” of observed statistical trends. FGIs typically offer greater descriptive depth than what is provided through surveys and a broader base of perspectives than often is obtained through individual interviews. For these reasons, FGIs are commonly included as part of both impact and performance evaluations.

DEFINITION

A focus group interview is a data collection technique in which a small group of people, usually between six and twelve individuals, is guided by a moderator to discuss specific topics in a structured way. The moderator raises issues identified in a discussion guide and uses probing questions to solicit opinions, ideas, and other information, often as it relates to a project or other activity. (For brevity, in the rest of this note “project” will be used to include projects and other activities.)

BACKGROUND

FGIs are commonly used in the fields of marketing and politics to quickly and inexpensively assess public opinion. What makes them useful in these settings also makes them valuable in the international development field. USAID’s staff and partners frequently collaborate with local organizations and communities, and FGIs are one way to quickly gauge the concerns or opinions of groups like these. Having this improved understanding makes it possible to better tailor USAID’s activities to the needs and goals of these communities, ultimately making these projects more successful and sustainable.

RATIONALE AND LIMITATIONS

This section describes the key reasons why FGIs are used in development evaluation, as well as their key limitations.

ADVANTAGES

When considering including FGIs in an evaluation design, they are often compared to large surveys and individual interviews. With these two alternatives in mind, the key advantages of FGIs are that they are:

1. **Interactive and Informative.** FGIs enable participants to interact with each other, sharing and reacting to each other's ideas. This process allows for new ideas to emerge that may not have been expressed by any one person. In other words, an FGI is more than just interviewing several people at the same time; it is more than the sum of its parts. Neither individual interviews nor surveys allow for this kind of interactivity.
2. **Time and Resource Efficient.** One FGI can collect rich, substantial data from multiple stakeholders in one or two hours. Though an individual interview allows the evaluation team to delve deeper into the situation of one individual, and surveys can capture data from statistically representative samples, FGIs offer an efficient option for obtaining detailed qualitative data from up to a dozen people in a short period of time, without the need for any special equipment.

ACTIVITIES WHICH ARE SIMILAR TO, BUT DIFFERENT FROM FOCUS GROUP INTERVIEWS:

Community interviews are intended for gathering data from a relatively large group of people (25-50 individuals). Typically they feature quite basic questions about community conditions or delivery of program services.

Group informant interviews often take place at the spur of the moment in the field. They involve administering an individual interview protocol with more than one person at a time. In other words, the interviewer may ask a question, and then each person will answer it in turn, but it lacks the structured, purposeful, guided interaction among participants that makes FGIs unique and valuable.

LIMITATIONS

There are also several limitations to FGIs that are important to consider.

1. **Require advanced planning.** FGIs call for getting together a group of people at the same place and time, which usually requires invitations to be distributed and confirmed in advance. Surveys and individual interviews also require advance planning to design the instruments and schedule the data collection sessions. However, the fact that FGIs require a group often implies greater coordination.
2. **Limited generalizability.** Usually, the prospective number of interviewees involved with a project is relatively large, and collecting the viewpoints of the limited number of people included in FGIs does not give evaluators a sample of sufficient size to be able to draw conclusions about the whole population. Therefore, FGIs are not intended to answer questions of "how much" or "to what degree" a service, result or opinion exists in the whole population. Surveys or other quantitative data usually fill this role.
3. **Vulnerable to moderator bias.** The moderator plays a central role throughout the FGI process, especially when guiding the discussion and interpreting the results. Thoughtful design of the moderator's guide, accurate recording, systematic coding and involvement of more than one individual in the data analysis can help minimize, but not eliminate, the risk of bias. On the other hand, in-person and telephone surveys, as well as individual interviews, also are susceptible to this kind of bias.
4. **Limited confidentiality.** While there are rare instances in which people may be more forthcoming in a group discussion than they would be on an individual basis, in most cases participants will be less likely to share very personal or sensitive information when they are speaking with a group. In order to collect such data, evaluators can use surveys or individual interviews, where privacy of the data collection can help to protect confidentiality.

DECIDING TO USE FOCUS GROUP INTERVIEWS

FGIs can be a useful data collection tool in many circumstances, but evaluators and evaluation managers still must think carefully about what the evaluation questions are, what the different ways for collecting the data necessary to answer those questions are, and how the data from FGIs can contribute to those answers. In situations where time or other resources are limited, or when group perspectives, not just individual opinions, are particularly important, FGIs may be appropriate and valuable. In many cases, FGIs are combined with other data collection techniques as part of a mixed-method design. (See the Technical Note on Conducting Mixed-Method Evaluations.) Doing so allows the evaluation to take advantage of the strengths of FGIs to accomplish specific purposes within the context of the evaluation, while at the same time supplementing them with other tools like surveys and individual interviews, in order to capture the most comprehensive picture possible of a project's performance or impact. Collecting information about the same issue but from different sources or different methods also allows for triangulation, the process whereby different sets of data are compared and contrasted to check for consistency, helping to increase confidence in the evaluation results.

FGIs often serve as a key source of data in performance evaluations, and often do so in combination with other data collection techniques such as reviews of secondary data and individual interviews. For example, when evaluating an education project based in schools, evaluators may conduct an individual interview with each school principal, because there is only one in each school and her or his perspective is likely to be especially important, and potentially different, from those of the teachers. Evaluators could then conduct an FGI with a group of teachers from each school, capturing a broader range of viewpoints in a shorter time than would be possible through individual interviews. In the same evaluation of this education project, evaluators may have identified certain trends in student test scores based on a review of existing secondary data. They could then use FGIs with teachers or with students to help explain these trends in the quantitative data. FGIs can also be particularly valuable when a performance evaluation aims to capture information about levels of stakeholder satisfaction, project strengths and weaknesses, and recommendations for improvement.

FGIs often play an important role in impact evaluations, even though, as noted above, data from FGIs are rarely representative enough to make generalizations about the overall impact of a project. Early on in an impact evaluation, FGIs can be used to collect qualitative information about the local environment which can

WRITING A STATEMENT OF WORK INCLUDING FOCUS GROUP INTERVIEWS

Once the evaluation manager has decided to include FGIs as part of an evaluation, there are several important considerations that she or he must keep in mind when writing the evaluation Statement of Work (SOW). The evaluation SOW should clearly indicate what deliverables will be expected, including how the raw FGI data will be submitted as part of the final report. The SOW should require the evaluator to provide an explanation of the following:

1. How the FGIs will help to answer the evaluation questions
2. Who will conduct the FGIs and why they are qualified to do so
3. What the rationale and method are for deciding the number, timing, and location of the FGIs
4. How the participants will be selected and recruited
5. How the FGIs will be recorded
6. How the FGI data will be analyzed and presented
7. What the expected deliverables are and which will need approval

An explanation of all of these items should be included in the evaluation design section submitted with the work plan, and also in the final report. The evaluation SOW should also specify that the evaluator must provide the data collection protocols, in this case, the FGI moderator guide(s), to the evaluation manager for approval prior to the start of data collection.

FGIs often play an important role in impact evaluations, even though, as noted above, data from FGIs are rarely representative enough to make generalizations about the overall impact of a project. Early on in an impact evaluation, FGIs can be used to collect qualitative information about the local environment which can

then be used to inform the design of the large-scale quantitative survey instruments. After such a survey is completed, FGIs can be used to explain patterns in the survey data and capture specific details and anecdotes which, when combined with the survey data, present a richer, more complete picture of the impact of the project.

PLANNING FOR FOCUS GROUP INTERVIEWS

SELECTING THE MODERATOR AND ASSISTANT MODERATOR

It is extremely difficult for one person to effectively facilitate and thoroughly document an FGI simultaneously. On the other hand, having too many non-participants in the room during an FGI may make participants feel uncomfortable or self-conscious. Therefore, FGIs are often conducted by a team of two or three people: a moderator and one or two assistant moderators. The **moderator** facilitates the FGI by firmly but unobtrusively guiding the group through a discussion that stays on topic, is rich in information, and maintains a comfortable environment where participants feel free to express their perspectives. The responsibility of the **assistant moderator** is to document the FGI, including taking thorough notes and making sure that all technology, such as audio or video equipment, is working properly. Both the moderator and the assistant moderator must be knowledgeable about the topic being discussed and very familiar with the language and culture of the participants. Given the importance of language in FGIs, moderators should be fluent in the language of the FGI participants. If that is not possible, a skilled interpreter should be present and the consequences of conducting the FGI through an interpreter should be described in the methodological limitations section of the evaluation report. The moderator should also know about the social, political and economic realities of the communities in which the FGI is being held, and about the design and implementation of the project to date. This knowledge allows the moderator to guide the discussion so that it addresses and explores the issues that are most relevant for the evaluation.

SELECTING THE PARTICIPANTS

After identifying the FGI team, the next step for the evaluator is to decide how many FGIs to conduct, and whom to include as participants. Whom the evaluation team selects to include will depend on the questions to be answered, but **in forming an FGI group it is important for all of the participants to be as similar as possible with regard to the characteristics most relevant to the questions asked.** For instance, if it is likely that males and females experience a project differently, perhaps because of traditional gender roles, then it is valuable to have separate FGIs—one with men and one with women. Similarly, the evaluator should take care to avoid situations where some participants in the group will be reluctant to share honestly because of the presence of other people in the group. Participants might be intimidated by having a community leader or a professional supervisor in the group, for example. In addition to sex, FGIs could also be separated (or disaggregated) by age group, ethnic group or role in relation to the project.

The ideal number of FGI participants is between six and twelve. Because the purpose of an FGI is to generate a focused discussion on a specific topic, it is important not only to have enough people to hold an engaging discussion, but also to limit the size of the group so that all participants can express their views. In some cases, it is easy to identify the ideal group of participants for an FGI because the total number of participants available is between six and twelve. In many other instances, however, there will be a large number of potential participants to choose from, even after the evaluation team has divided up the stakeholders based on relevant characteristics. In these cases, the team will have to use an alternative method to select the FGI participants. There are a few appropriate FGI sampling methods to choose from, each with its own strengths and weaknesses.

NOMINATION

Using this method, key stakeholders nominate people they think would make good participants. Nominees are likely to be familiar with and able to speak about the topic, and will also likely be willing to participate in the FGI. However, if the stakeholder who is nominating participants wants the results of the FGI to reflect positively upon the project, she or he may deliberately avoid nominating individuals who have had negative experiences or are dissatisfied with the project. On the other hand, an opponent of the project might only nominate participants who have negative things to say about it. In this way, relying on key stakeholders to nominate FGI participants could introduce bias into the data and prevent the moderators from capturing a comprehensive and accurate picture of the project.

VOLUNTEERING

If there are a large number of potential participants, and the selection criteria are broad, volunteers can be recruited with flyers, newspaper ads, or announcements at religious or civic meetings. This sampling method is simple, but can lead to biased findings as the people who respond to these notices may be systematically different from the general target population. For instance, they may be beneficiaries who had particularly good or bad experiences with the project, and therefore are motivated to take the time to participate. If the sampling methodology is based only on volunteering, participants may also be disproportionately comprised of individuals who have more leisure time, while other potential participants who have more commitments, such as work or family responsibilities, may be less likely to volunteer.

RANDOM SAMPLING

If there is a large but defined group of potential participants, such as beneficiaries of a project, it may be possible to randomly select FGI participants. Randomly selecting participants can make FGIs more representative of the total population of potential participants by minimizing the sources of bias that can result from other sampling methods. In order to use random sampling, the evaluation team must have a list of all of the potential participants. The team can then assign a number to each individual and randomly choose as many numbers as the evaluation design says are needed. Randomly selecting FGI participants does introduce the risk that some participants selected may not be willing or able to participate. Pressuring unwilling individuals to participate in an FGI will not only be difficult for the moderator, but it is also likely to decrease the quality of the data collected. An unwilling participant may not fully engage in the discussion, or may respond to questions more negatively than they otherwise would, which also could negatively affect the tone of the overall discussion.

REPRESENTATIVENESS: GENERALIZING FROM FOCUS GROUP INTERVIEWS

It can be easy to misunderstand the role focus groups can play in the analysis of opinions of a large group of people. Focus group participants ideally should be representative of the larger relevant population—for instance, with individuals coming from various geographical districts, from various age groups, representing both genders, and so on. But it is important to keep in mind that this “representativeness” is only for strengthening the diversity of perspectives to be gathered from focus group participants; this method of selection does not allow the evaluator to make statistically valid generalizations about the opinions of the whole population.

If the evaluation team gathers data from several focus groups and systematically asks similar questions within these sessions, the evaluation report may appropriately refer to patterns *across the focus groups*. (“Seven of the nine focus groups included positive comments about the quality of training.”) But evaluators should avoid summarizing these findings *across participants* using percentages (“Fifty percent of participants supported the training”) or fractions (“Two thirds of participants were satisfied with the program”), as these kinds of statements are likely to encourage readers to incorrectly interpret such data as applicable to the larger population.

DETERMINING THE NUMBER, TIMING AND LOCATION OF FOCUS GROUP INTERVIEWS

When there is a small number of potential FGI participants, the evaluator can divide all of the potential participants into groups of between six and twelve, and the number of FGIs to conduct will be clear. However, when the number of potential FGI participants is large, and the evaluator could conduct multiple FGIs within the same target group, it is less clear as to how many FGIs should be conducted. **In the absence of other constraints, it is best to continue FGIs with different participants until no new perspectives or information are being discovered, or in other words, until the moderator starts to hear the same things repeatedly. In reality, however, it may take many FGIs to reach this point, and constraints such as time, cost, and access to participants may limit the number of FGIs that can reasonably be completed.** In situations such as these, the evaluation team will have to determine the best solution, based on the evaluation objectives, the characteristics of the population, and the constraints of the project and the environment. For example, while a team may wish to conduct FGIs with six different groups (males and females of three different age groups) the team may have available resources for only five FGIs. In this instance, the evaluation team may choose to combine the two youngest age groups into one mixed-sex group, because up until adolescence, boys' and girls' experiences with the project are relatively similar, but older women and men are affected quite differently by the project. While this compromise is not ideal, and important information may be lost, it is typical of the kind of decision that often must be made in real-world contexts.

FGIs should be scheduled at times and in places that are comfortable and convenient for the participants. If participants feel comfortable and at ease in the environment, they will be more likely to respond openly and honestly to the questions asked. FGIs with project and ministry staff should be held in locations where they feel their comments will not be overheard by supervisors or other non-participant colleagues. The location should be private and, with the exception of the moderator and assistant moderators, there should be no non-participant observers (such as USAID evaluation managers) present. In the case where participants may be accompanied by their children, childcare should be provided, if feasible, so that participants will be able to devote their undivided attention to the discussion. Seating should be arranged in a circle or oval to emphasize that the opinions of all participants are equally valuable. Finally, especially if the participants have had to travel to the FGI, the moderator should consider providing refreshments and compensation for travel expenses.

PREPARING THE MODERATOR'S GUIDE

The moderator's guide helps the moderator introduce the purpose and process of the FGI to the participants, and then outlines the key questions or topics to be discussed during the FGI.

INTRODUCTION, GROUND RULES, CONFIDENTIALITY AND INFORMED CONSENT

The moderator should first **introduce herself or himself**, then welcome and thank all of the FGI participants. After that, the moderator should **explain the purpose of the FGI**, how the participants were selected, any potential benefits or risks to participating in the FGI, how long it will take, and whether there will be any compensation for participating. After discussing these logistical issues, the moderator also should **address expectations**, or ground rules, for the FGI. The ground rules will vary depending on the FGI, but in general they will include:

- Everyone is encouraged to share their ideas, and the FGI is strengthened if everyone participates.
- There are no wrong answers, and everyone's perspective is equally valued.
- The ideas shared during the FGI should not be shared outside the FGI with non-participants in order to respect participants' privacy.
- Disagreements about ideas can be valuable and productive, but personal attacks will not be tolerated.

After establishing these ground rules, the moderator should ask if there are any questions or concerns participants have, and these issues should be addressed and consensus reached as a group before moving on. The question of **confidentiality** is also important to address, and the approach to protecting confidentiality as data are gathered, stored and reported should be discussed and agreed upon between the evaluation manager and the evaluation team during the design phase. The moderator should clearly describe how the data collected will be used, including with whom it will be shared, and crucially, whether names or other personal or identifying information will be included with the data. Many times, the experiences and opinions shared during a focus group will include sensitive information, and participants may not feel comfortable sharing openly if they feel it could have negative consequences for them in the future. The moderator must be honest about how the data will be used, but should also reassure the participants that the data will be treated sensitively and that their privacy will be respected to the greatest degree possible given the needs and purposes of the evaluation. After providing this information, it is important to describe what will or will not happen if they choose not to participate. To ensure the data collected are reliable, participation in an FGI should be entirely voluntary and there should be no consequences for declining to participate. After informing participants of all of this information, **the moderator must ask each member to confirm that they consent to participate in the FGI**. Often, to be consistent across FGIs, the language communicating these points is written in the moderator's guide.

TYPES OF FGI QUESTIONS

Good discussion questions initiate exchanges between group members and elicit multiple points of view on a topic. The best questions are simple, single-topic, and use language familiar to, if not commonly used by, the participants themselves. There are many different kinds of questions that can be useful at different points in a discussion as well as for eliciting different kinds of information. Thought and care should be put into how questions are asked, and in what order. Follow-up questioning and other moderation strategies are described in the "Conducting and Recording the Focus Group" section below.

- **Closed-ended questions** are those that can be answered "yes" or "no" (e.g., "Do you collect water from a well?"), or questions that have a limited range of answers (e.g., "At what time of day do you collect water?"). Questions that ask participants to respond by raising their hands also fit into this category. These questions generally do not elicit discussions and should be used sparingly, but they can be useful at the beginning of a discussion to "break the ice," or "warm up" the participants before posing more complex questions.
- **Open-ended questions** have a wide range of possible responses, and are therefore more likely to generate discussion among the participants. Questions such as "What do you think of the new program?" "How has the program benefitted or harmed the community?" or "How did you feel when the program was ending?" are all examples of open-ended questions. These types of questions are valuable because they spur a variety of responses, but the moderator must also be careful to guide the responses so that they do not stray too far from the main topics of the FGI.
- **Recall and hypothetical questions** are types of open-ended questions that can be very productive for discussions. Recall questions, such as "Tell me about the first time the program community advocate came to visit you," elicit stories from group members and can yield rich and detailed data. Hypothetical questions, such as "If you were to improve the program, what would you change and what would you keep the same?" also are productive ways to initiate discussion and are good for exposing the values behind participants' practices and opinions.
- **Activities** also can be valuable tools for eliciting discussion and energizing participants. Activities can include role plays, making lists, or drawing maps or illustrations, all of which can be done by individuals or small groups and then shared with the whole group for comment. For example, the moderator could ask a group of community advocates to draw a map of the community and mark the areas where they met the

most resistance to program implementation. During an FGI with twelve community members, participants could break into four groups of three and create lists of the new skills or information they learned from the community advocates.

Because FGIs involve discussions of sometimes complex concepts by multiple individuals, it is important to **clearly define the words and concepts being discussed**. For example, terms like “poverty,” “rights,” and “development” likely have different meanings for different people, even within the same community. Even seemingly simple words like “children,” can be interpreted differently. Thus, while everyone is likely to agree that a five-year-old girl or boy is a child, what about a fifteen-year-old, or an eighteen-year-old? So, if the moderator wants to know about a program’s impact on children in the community, she or he may first discuss and agree on a working definition of “children” to be “boys and girls up to fifteen years old,” and then ask questions like, “Has the program benefitted children in your community?” If the moderator understands the local context and language, it will obviously be much easier for him or her to choose appropriate words to use in the FGI.

CONDUCTING AND RECORDING THE FOCUS GROUP INTERVIEWS

After welcoming the participants, explaining the purpose and ground rules of the FGI, and attaining participant consent, the moderator will begin the discussion. The moderator’s ability to effectively and efficiently guide the discussion is key to a successful FGI. The following are some strategies and techniques moderators often use.

INCLUDING ALL PARTICIPANTS

The moderator should take note of who has not yet spoken, and who has spoken a lot, and respectfully steer the discussion in a way that allows for all to be heard. The moderator may need to respectfully interrupt some talkative participants or gently coax others who are reluctant to share. Cultural and gender awareness and sensitivity of the moderator are critical factors here, since participant expectations and responsiveness to moderator cues will vary considerably according to cultural context. The moderator also should be aware of



the way power relations may contribute to this dynamic, as shyness may be a result of fear to speak in the presence of certain individuals or groups. In this case, it may be necessary to reevaluate the makeup of the groups, or to offer to speak with certain individuals in private after the FGI has concluded.

EMPLOYING FOLLOW-UP QUESTIONS

Skilled use of follow-up questions to comments made by participants is one of the most important qualities of a successful FGI moderator. Participants may

make statements that are confusing, contradictory, incomplete, or unclear. Discussions may drift away from the focus of the FGI, or may stall if participants are reluctant to speak. It is at these points that moderators can employ follow-up questions to help revive and guide the conversation.

- **Clarification questions**, such as, “What did you mean when you said X?” ask a respondent to define terms or phrases whose meaning may not be clear, may be contested, or is central to the discussion topic.
- **Probing questions** ask for more detail in an answer. For instance, the moderator may say, “You mentioned that the community advocate didn’t always visit your family. Tell me more about that.”
- **Origin questions** ask about where an opinion or idea came from, such as “What led you to this idea?”
- **Conflicts and consequences questions** ask about the implications of an idea, such as “Is there anyone who disagrees with what was just said?” or “What might be the positive and negative impacts of doing things the way you describe?” Disagreements within the group are not necessarily bad, but allowing two participants to argue back and forth without reaching a consensus, for example, can be distracting and cause other participants to become less engaged.

RECORDING THE DISCUSSION

A wealth of valuable information can be raised during an FGI, but in order for it to be analyzed and used, the assistant moderator must accurately and comprehensively record it. Different data from the FGI can be recorded in different ways, but generally, it will involve some kind of note-taking and also possibly audio or video recording. How the discussion will be recorded should be agreed upon between the evaluation team and the evaluation manager at the design phase. Regardless of the final determination, the following points are important to note:

- **Consistency and clarity:** Having a standardized recording form tailored for each set of FGIs can help the assistant moderator capture and organize this data.
- **Group member characteristics:** Characteristics of participants which are most relevant to the evaluation objectives, should be noted, as this information may be crucial to understanding and analyzing the FGI.
- **Key points and themes:** Assistant moderators should record the key points and themes that arise for each discussion topic as thoroughly as possible, whether they are expressed by only one or multiple participants.
- **Word-for-word quotations:** As often as possible, the assistant moderator should record participants’ exact quotations, as these are often the most useful and powerful data derived from FGIs.
- **Non-verbal observations:** In addition, the assistant moderator also should take note of participants’ body language, such as head nodding, eye contact, voice volume, or emotions that would indicate intensity of agreement or disagreement, the importance of particular topics, or the power dynamics among participants.

Recording the sound or images from an FGI can be a valuable supplement to written notes. Recording discussions also makes it possible to transcribe them and to code the content, which allows evaluators to conduct more detailed analyses of the data.

ANALYZING THE DATA

The analysis of FGI data actually begins during the FGI and continues in a systematic way until the results are reported.

STAGES OF ANALYSIS

- **During each FGI**, the moderator is constantly listening to, processing, and responding to the participants’ statements in order to guide the conversation and explore emerging ideas.

- **Immediately after each FGI**, the moderator and assistant moderator should discuss their initial impressions and interpretations of the ideas raised during the discussion, and should note any key themes they have already identified.
- **As soon as possible after each FGI**, the moderator and assistant moderator should assist with the transcription of any recordings made and should prepare a field report for each FGI. Field reports should summarize the responses to each key question and highlight quotations that illustrate particularly important ideas. These reports can be coded for analysis using software designed specifically to analyze such data.
- **After the last FGI**, the evaluators must then begin comparing and contrasting the results of different FGIs, if more than one was conducted, looking for patterns and trends. (If only one FGI was conducted, then the evaluator can look for patterns or trends among participants' expressed ideas.) These findings may be identified and organized based on key research questions, but the evaluator should also be conscious of the possibility of discovering unanticipated findings that do not fit neatly within any of the original research questions.

TECHNIQUES FOR THE ANALYSIS OF QUALITATIVE FOCUS GROUP INTERVIEW DATA

The way that the data from the FGI is analyzed will depend on the purpose of the evaluation and the role of the FGI within the evaluation design.

TRANSCRIBING

If the assistant moderator created an audio or video recording of the FGI, the evaluator may use the recording to make a transcript. A transcript is a word-for-word record of everything that was said during the FGI. Producing a transcript can be very time-consuming, but once it is done, the evaluator has many more options for analysis, including coding.

CODING

Often, and especially in evaluations including more than one FGI or other sources of qualitative data in addition to an FGI, the evaluator will decide to code the FGI data so that it is easier to compare different data sets and to identify any themes or patterns present in the data. Coding involves labeling or categorizing passages or parts of transcripts or other data sources so that the data can be readily retrieved, searched, compared and contrasted. If there are a large number of FGIs or a large amount of data to code, there are a number of computer software packages (for example, Atlas ti, Ethnograph, NVivo) that can help code this type of data automatically. Even with the use of a software package, coding qualitative FGI data requires a substantial amount of time, and the evaluation manager should allow for adequate time in the SOW. The evaluator should also make clear in the methodology section of the work plan whether and how the data will be coded as part of the analysis process, and whether and how this coded data will be included in the final report. (See "Report the Findings" below.)

TRIANGULATING

Triangulation is the process through which evaluators compare and contrast findings related to the same question but drawn from different sources and methods. Evaluators often use data from FGIs to triangulate findings from other methods to help to reduce potential bias and increase confidence in the evaluation's overall findings and conclusions. For a more detailed discussion of triangulation, please see the Technical Note on Conducting Mixed-Method Evaluations.

REPORTING THE FINDINGS

FGIs often produce a large amount of different kinds of data, including quotations, observations, and general impressions of the moderators. It can therefore be difficult to organize and present this data to other interested stakeholders. Presenting FGI results can be further complicated by the fact that FGIs are often one of several data collection techniques employed as part of an evaluation, so the evaluation report must not only clearly present the results from the FGI, but also combine the FGI results with those from the other techniques in a way that is coherent and that supports the overall findings, conclusions and recommendations of the report. Generally, there are three key factors to gauge the presentation of FGI findings in an evaluation report:

1. **Relevance to the evaluation questions:** Do the results presented help to answer the evaluation questions?
2. **Sufficient methodological information:** Does the report provide enough information about how the FGIs were conducted so that the reader can have confidence in the results?
3. **Succinct and comprehensive analysis:** Are the results presented clearly so that the reader can quickly grasp the most important themes and findings?

It is best practice to provide a separate summary of FGI results, including separate summaries for the different relevant groups of participants, in the evaluation report. For example, if an evaluation included three sets of FGIs, one set with students as participants, one with their parents as participants, and a third with their teachers as participants, then the evaluation report should address what the main findings were for each group. **The raw data obtained from the FGIs also must be submitted to USAID as part of the evaluation files** (along with the FGI moderator's guides described above). The section below outlines several different ways the evaluator can include and incorporate these different components into the final report.

WAYS TO PRESENT FOCUS GROUP INTERVIEW FINDINGS

There are several ways to present FGI findings as part of an evaluation report.

1. **In a separate section in the main body of the report:** The authors may choose to dedicate a chapter or section in the main body of the report to the FGI results. In this case, it should include the main findings that emerged from the FGI data, disaggregated by relevant FGI groups, as well as actual quotations from FGI participants which help to highlight and support these findings. When the evaluation design uses FGIs, and only FGIs, to address a single evaluation question, this strategy of presenting FGI results may be most appropriate.
2. **Summarized in the main body, details in a separate annex:** The authors may choose to include some discussion of the FGI results in a summary in the main body of the evaluation report, but still use an annex to go into greater detail with a complete summary of this data.
3. **Interspersed throughout the main body, details in a separate annex:** As FGI data are often used to complement other data to answer the same questions, it is often useful to present all of the data relevant to a single evaluation question, including the FGI data, together, integrated into the same section of the main report. For example, the findings section of an evaluation report may contain a table with collected survey data, followed by a paragraph which interprets the quantitative data in the table. This

REPORTING FOCUS GROUP INTERVIEW METHODS AND EVIDENCE

When evaluations use FGIs, the following key components of an evaluation report should provide transparent information on FGI methods and evidence:

1. **Methodology:** Include rationale for using FGI, selection strategy, limitations, analysis plan, etc.
2. **Data Collection Instruments:** Include all FGI moderators' guides in the annex.
3. **Findings:** Disaggregate findings by all relevant FGI participant groups.
4. **Data Sources:** Include raw FGI data in an annex in the form of transcripts, detailed summaries, or audio recordings as determined by the SOW.

paragraph could then be immediately followed by a paragraph discussing how the data collected from the FGIs support, refute or provide context to the survey data. The authors should also include relevant quotations from FGI transcripts in this section in order to enrich this discussion of results. Presenting FGI results in this way adds richness to the discussion of the evaluation results and clearly demonstrates how the different data sets help to complement each other. However, a separate section communicating the results of the FGIs, as an annex, is still recommended.

Regardless of the broader placement of FGI findings in the evaluation report, the authors should make optimal use of the actual words participants use during the FGIs. Direct quotations from FGI transcripts can make the evaluation report more interesting and persuasive, and can be inserted directly into the relevant paragraphs or included in text boxes to set them apart and highlight them as primary data. A quotation should not disclose the identity of the individual speaking, but it often is useful to make clear the individual's role within the project, for example, “-A small business owner in West Java”.

WAYS TO PRESENT FOCUS GROUP INTERVIEW RAW DATA

As noted above, USAID Evaluation Policy does require that the raw data obtained from FGIs be submitted to USAID along with the final evaluation report. Sometimes, this data will come in the form of full transcripts, plus moderator notes, from every FGI conducted as part of the evaluation. Though full transcripts provide the best written record of an FGI, producing them requires a considerable amount of time on the part of the evaluation team, and the evaluation manager should make sure, if they are required to be submitted as a part of a deliverable, that the SOW includes sufficient time and resources to allow for this process. In many cases, when the evaluation includes multiple FGIs, the transcripts from all of the FGIs may total many pages, so that including all of them as part of the evaluation report may be overly costly and impractical. In cases like these, the evaluation manager may decide to ask for a written summary of each FGI to be submitted instead of complete transcripts. Including the full set of raw data with the final evaluation report increases the transparency of the overall evaluation and allows the reader to explore FGI results in greater depth.

ASSESSING FGI DATA QUALITY

If the evaluation manager, or any interested stakeholder, is provided with all of the required components discussed above (detailed methodology including limitations, moderator's guides, summary of results disaggregated by relevant groups, and raw data), he or she can then essentially retrace the steps of each FGI and in doing so, assess the quality of the FGI data and the resulting findings. The assessor could first review the moderator's guides, checking to see whether the best practices described above are incorporated into the guides and whether the questions included in the moderator's guide are appropriately crafted to help answer the evaluation questions. By reviewing the raw data (either the FGI summaries or transcripts), the assessor could also make note of what proportion of participants actively spoke and responded to questions during the FGI; if only a small fraction of the respondents actively share their ideas during the FGI, then the resulting data is likely to be less informative and less reliable than data from an FGI where all group members actively participate. If a disproportionate number of participants in an FGI belong to one group (many more men than women; many more elderly adults than younger adults), and this was not the intention of the evaluator, then this skewed makeup is also likely to introduce bias into the results. **The moderators and the evaluators should clearly explain any limitations to the methodology and sources of bias that arose during implementation in the final evaluation report.**

ADDITIONAL RESOURCES

The following resources can be used as samples or templates, or provide more information on the topics reports and on evaluation in general. Where information differs, USAID's ADS (Automated Directives System) 200 series take precedence over that in other resources.

"Can You Call It A Focus Group?", Iowa State University- University Extension, accessed February 22, 2013. <http://www.extension.iastate.edu/publications/pm1969a.pdf>.

"Focus Groups", Better Evaluation, accessed February 22, 2013. <http://betterevaluation.org/evaluation-options/FocusGroups>.

R. Krueger and Mary Anne Casey, *Focus Groups: A Practical Guide for Applied Research* (3rd ed.). Thousand Oaks, CA: Sage, 2000.

International Program for Development Training, "Toolkit 5: Focus Groups," IPDET Handbook, Module 8 (Data Collection Methods), pp. 395-413. http://www.worldbank.org/oed/ipdet/modules/M_08-na.pdf.



HOW-TO NOTE

Preparing Evaluation Reports

Monitoring and Evaluation Series

**Create
evaluation
reports that
are clear,
credible, and
useful.**

How-To Notes

are published by the Bureau of Policy, Planning and Learning and provide guidelines and practical advice to USAID staff and partners related to the Program Cycle. This How-To Note supplements USAID ADS Chapter 203



INTRODUCTION

This Note supplements USAID Automated Directives System (ADS) Chapter 203 and provides current good practice in preparing evaluation reports, the main deliverable for most evaluations. Following these practices will help to establish clear expectations for evaluation reports during the preparation of evaluation statements of work and the in-briefing of the evaluation team. These practices also serve as a guide for reviewing the quality of draft evaluation reports submitted by the evaluation team. This Note is also a resource for USAID partners and independent evaluators of USAID programs and projects. An [evaluation report template](#) and [sample evaluation report covers](#) are available as additional resources.

BACKGROUND

The most important outcome of an evaluation is that it is used to inform decisions and improve USAID projects and programs. A key factor in using evaluation findings is having a well-written, succinct report that clearly and quickly communicates credible findings and conclusions, including easy-to-understand graphics and consistent formatting.

REQUIREMENTS

USAID's Evaluation Policy and ADS 203 provide guidance on evaluation report structure and content, and steps in the process of creating a report. These are listed in Table 1. The report must present a well-researched, thoughtful and organized effort to objectively evaluate a USAID activity, project or program. Findings, conclusions and recommendations must be based in evidence derived from the best methods available given the evaluation questions and resources available. The evaluation methods, limitations, and information sources must be documented, including by providing data collection tools and the original evaluation statement of work as annexes to the main report. Finally, the findings should be shared transparently and widely, to ensure accountability and to promote learning from USAID's experience.

TABLE 1: EVALUATION REPORT REQUIREMENT

(from the USAID Evaluation Policy and ADS 203)

Report should be...	A thoughtful, well-researched, well-organized, and objectively evaluate what worked, what did not, and why.
Executive Summary	Include a 3 to 5 page Executive Summary that provides a brief overview of the evaluation purpose, project background, evaluation questions, methods, findings, and conclusions.
Evaluation Questions	Address all evaluation questions in the statement of work.
Methods	<ul style="list-style-type: none"> ▪ Explain evaluation methodology in detail. ▪ Disclose evaluation limitations, especially those associated with the evaluation methodology (e.g. selection bias, recall bias, unobservable differences between comparator groups, etc.). <p>NOTE: A summary of methodology can be included in the body of the report, with the full description provided as an annex.</p>
Findings	<ul style="list-style-type: none"> ▪ Present findings as analyzed facts, evidence and data supported by strong quantitative or qualitative evidence and not anecdotes, hearsay or people's opinions. ▪ Include findings that assess outcomes and impacts on males and females.
Recommendations	<ul style="list-style-type: none"> ▪ Support recommendations with specific findings. ▪ Provide recommendations that are action-oriented, practical, specific, and define who is responsible for the action.
Annexes	<p>Include the following as annexes, at minimum:</p> <ul style="list-style-type: none"> ▪ Statement of Work. ▪ Full description of evaluation methods. ▪ All evaluation tools (questionnaires, checklists, discussion guides, surveys, etc.). ▪ A list of sources of information (key informants, documents reviewed, other data sources). <p>Only if applicable, include as an annex <i>Statement(s) of Differences</i> regarding any significant unresolved differences of opinion on the part of funders, implementers, and/or members of the evaluation team.</p>
Quality Control	Assess reports for quality by including an in-house peer technical review with comments provided to evaluation teams.
Transparency	<ul style="list-style-type: none"> ▪ Submit the report to the Development Experience Clearinghouse (DEC) within three months of completion. ▪ Share the findings from evaluation reports as widely as possible with a commitment to full and active disclosure.
Use	Integrate findings from evaluation reports into decision-making about strategies, program priorities, and project design.

STEPS IN THE PROCESS

1. Define Report Requirements in the Evaluation Statement of Work and Final Work Plan

All evaluation statements of work (SOW) should clearly define requirements and expectations for the final evaluation report. All of the items in Table 1 must be included as requirements for the final report. Ensure that all requirements in the SOW are also included in the final evaluation work plan that is put in place once the evaluation team is on board. Adjustments can be made at this time, as long as the minimum requirements are met, and additions can be included such as defining when the first draft will be due, how many days USAID will have to review and provide comments, and when the final report will be submitted.

2. Review First Draft

Program Offices must ensure that evaluation draft reports are assessed for quality by management and through an in-house peer technical review and comments provided to the evaluation teams. USAID staff may consider including implementing partners and other direct stakeholders in the review process. Tools such as the USAID Evaluation Report Checklist can be used.

3. Final Draft and Statement of Differences

Evaluation reports are independent products and therefore the evaluation team leader reviews the comments and determines which to incorporate into the final draft. Once the final draft is submitted to the USAID mission or office, the content should not be changed without the permission of the evaluation team leader. USAID, other funders, implementing partners, and other members of the evaluation team can decide to include a *Statement of Differences* as an annex to the report, if there are differences related to the evaluation findings or recommendations.

4. Submit to DEC and Share Findings Widely

USAID Program Offices must ensure that evaluation final reports (or reports submitted by evaluators to USAID as their final drafts) are submitted within three months of completion to the Development Experience Clearinghouse at <http://dec.usaid.gov>. The actual submission can be done by USAID staff or by the evaluation team with USAID concurrence (once an opportunity has been provided for USAID or others to include a *Statement of Differences*, if appropriate). In addition to submission to the DEC, USAID should also consider how to share the evaluation report widely to facilitate broader learning. This could include posting the report on the USAID mission website, translating a summary into local language, and hosting presentations of the evaluation findings.

5. Use Evaluation Findings to Inform Decisions

USAID must integrate evaluation findings into decision making about strategies, program priorities, and project design. While the Program Office in a mission should ensure this happens, it is the responsibility of all USAID staff.

CONTENT AND STRUCTURE

General Style

When writing a report, the evaluation team must always remember the primary audience: project and program managers, policymakers, and direct stakeholders. The style of writing should be easy to understand and concise while making sure to address the evaluation questions and issues with accurate and data-driven findings, justifiable conclusions, and practical recommendations.

Report Sections and Content

At a minimum, all reports should include the following sections: Executive Summary (3 to 4 pages); Evaluation Purpose and Questions (1 to 2 pages); Project Background (1 to 3 pages); Evaluation Methods and Limitations (1 to 3 pages, with full version provided in an annex); Findings, Conclusions and Recommendations (15 to 25 pages); and, Annexes. Reports may include additional content, split the sections up differently, or present the sections in a different order.

- **Executive Summary**

The Executive Summary, between three to four pages in length, should stand alone as an abbreviated version of the report. All content of the full report should be summarized, and the Executive Summary should contain no new information.

- **Evaluation Purpose and Questions**

The evaluation purpose should be clearly defined at the beginning of the report. It should describe in about one page or less why the evaluation is being conducted now, how the findings are expected to be used, what specific decisions will be informed by the evaluation, and who the main audiences are for the evaluation report. The evaluation questions are linked to the purpose, and should be listed here. Good practice is to limit the questions to three to five that are clear, focused, and that will directly inform specific decisions.

- **Project Background**

This section should summarize the project being evaluated in one to three pages, including the original problem the project is designed to address, any changes that have occurred since

the project was started, a description of the beneficiary population, geographic area of the project, and the underlying development hypothesis, or causal logic, of the project or the broader program of which the project is a part. If a results framework (for strategies, objectives or programs) or logical framework (for projects) is available, this should be included here. For projects designed under the [project design guidance](#) released in 2011, the evaluation team should have access to the final Project Appraisal Document and related annexes (which includes a logical framework and original monitoring and evaluation plans, among other things). This information provides important context for understanding the evaluation purpose, questions, methods, findings and conclusions.

- **Methods and Limitations**

This section should provide a detailed description within one to three pages of the evaluation methods and why they were chosen. If more space is needed, additional detailed information on the methods should be provided in an annex. The reader needs to understand what the evaluation team did and why to make an informed judgment about the credibility of the findings and conclusions and the underlying evaluation design including the data collection and analysis methods.

Evaluation methods should correspond directly to the questions being asked and should generate the highest quality and most credible evidence possible, taking into consideration time, budget and other practical considerations.

This section should provide information on all aspects of the evaluation design and methods, including tradeoffs that led to selection of specific data collection and analysis methods, a description of data availability and quality, and sampling strategies (purposive, random, etc.), including how interview subjects or site visits were selected. Just as important as describing the evaluation methods is describing any limitations in data collection and analysis, data quality, access to data sources, or any other factors that may result in bias. To show the

relationship between the evaluation questions and methods, it is useful to include a chart that lists each evaluation question, the corresponding evaluation method to be used for data collection and analysis, data sources, sample sizes, and limitations.

- **Findings, Conclusions and Recommendations**

Findings, conclusions, and (if requested in the evaluation statement of work) recommendations, make up the main body of the report, synthesizing what was learned during the evaluation and presenting it in an easy to understand and logical fashion. Findings are empirical facts based on data collected during the evaluation and should not rely only on opinion, even of experts. Conclusions synthesize and interpret findings and make judgments supported by one or more specific findings. Recommendations, if applicable, are specific actions the evaluation team proposes be taken by program management that are based on findings and conclusions. The reader should be able to discern what evidence supports the conclusions and recommendations. Whenever possible, data should be presented visually in easy to read charts, tables, graphs, and maps to demonstrate the evidence that supports conclusions and recommendations. All graphics must have a title, be clearly labeled, and include a caption.

- **Annexes**

All evaluation reports must include the following as annexes: 1) the Evaluation Statement of Work, 2) Detailed description of the evaluation design and methods, 3) copies of the actual data collection tools such as survey or interview questions, 4) a list of information sources (including documents reviewed, sites visited, and key informants, assuming they gave permission to be identified), and 5) disclosure of any conflict of interest by including a signed statement by evaluation team members that attests to a lack of conflict of interest or describes an existing conflict of interest relative to the project being evaluated. Additional annexes can be included at the discretion of the evaluation team and USAID, and in some cases implementing partners, including, if applicable, any *Statements of Differences* with the evaluation conclusions.

FORMAT AND GRAPHIC STANDARDS

Reminder on USAID Graphic Standards

The USAID Graphic Standards Manual is available at http://pdf.usaid.gov/pdf_docs/PNADB334.pdf.

Evaluation report authors and reviewers should be familiar with the USAID Graphic Standards and apply them consistently. These include requirements and guidance related to USAID branding, choice of typography, and color palette.

Cover

The cover of an evaluation report should be attractive and provide enough information that a reader can immediately understand what was evaluated. To make evaluation reports distinct from other types of USAID publications all evaluation report covers should:

- Follow USAID Branding and Graphics Standards.
- Include a title block in USAID light blue background color with the word “Evaluation” at the top and the report title underneath. The title should also include the word “evaluation.”
- Include the following statement across the bottom: “This publication was produced at the request of the United States Agency for International Development. It was prepared independently by [list authors and/or organizations involved in the preparation of the report].”
- Feature one high-quality photograph representative of the project being evaluated. The photo should be high resolution, visually simple, colorful, and in focus. Include a brief caption on the inside front cover explaining the “who, what, when, where, and why” of the photo and with photographer credit.

Title and Title Page

While titles are determined by the evaluation team and the USAID Mission or operating unit commissioning the evaluation, all evaluations will be submitted to the DEC and therefore titles should be clear to the general reader. A review of recent evaluation titles leads to the following suggestions for good practice:

- Compose a title that is informative, clear and compelling (e.g., “Improving Community Health in Fredonia: Evaluation of the USAID/Fredonia Community Health Project”).
- Avoid acronyms and do not use implementing partner names (e.g., “XYZ LTD Evaluation”).
- Include the word “evaluation” in the title. This will help the DEC correctly archive the document.

The report title should be repeated on the title page, the first right-hand text page of the report. The title page also includes the subtitle, if any, and the standard disclaimer for publications by external authors: “The author’s views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.” It may also include the date of the report, a short abstract summarizing the report, or other information.

Acronyms

Keep the use of acronyms to a minimum, and define all acronyms used in the report by including an acronym list in the beginning of the report

Table of Contents

This comes before any content referenced in the table. Sufficient detail should be provided to guide the reader through the report, including page numbers.

Length

Evaluation reports should be no more than 30 pages in length, not including any annexes and three to four pages for an Executive Summary. This should be sufficient to provide a summary of the evaluation purpose and approach, key findings, conclusions and recommendations. Additional detail can be provided as annexes.

COMPANION PRODUCTS

USAID staff should consider other products to include in the evaluation statement of work that can complement the report and aid in disseminating evaluation findings to a broader audience. These could include photos documenting the evaluation, a short video that combines footage from the evaluation with a summary of the findings, a short fact sheet, a local language translation of the executive summary of the evaluation report, or a presentation via webinar of the evaluation report. Some products are not appropriate to ask as a deliverable from an evaluation team, such as “Success Stories” as this would put the evaluation team’s objectivity and independence into question.

ADDITIONAL RESOURCES

The following resources can be used as samples or templates, or provide more information on evaluation reports and on evaluation in general. Some other resources exist but are out-of-date with current USAID guidance. Where information differs, the USAID Evaluation Policy and the USAID ADS (Automated Directives System) 200 series take precedence over that in other resources.

- Evaluation Report Template: <http://kdid.org/kdid-lab/library/sample-eval-report-template>
- Evaluation Cover Samples: <http://kdid.org/kdid-lab/library/sample-eval-report-covers>
- Sample Disclosure of Conflict of Interests Form: <http://kdid.org/kdid-lab/library/DisclosConflictInterest>
- USAID Graphic Standards Manual: http://pdf.usaid.gov/pdf_docs/PNADB334.pdf
- USAID’s Center for Development Information and Evaluation Publications: Style Guide: Guidelines for Project Managers, Authors, and Editors, December 2001
http://pdf.usaid.gov/pdf_docs/PNACN266.pdf

EVALUATION REPORT AND REVIEW TEMPLATE

The Evaluation Report Checklist and Review Template are tools to assist in developing and reviewing USAID Evaluation reports. The checklist provides a quick guide to understanding the minimal standards for an evaluation report, while the Review Template provides additional criteria for assessing the quality of the draft report during a peer review. For further guidance on developing an evaluation report, see the Evaluation Report How-to Note and Template.

	Evaluation Report Compliance Checklist	Evaluation Report Review Template
Correct Usage	Determine if required, essential, or highly recommended elements are present in an evaluation report and compliant with the USAID Evaluation Policy	Assess the quality of a draft evaluation report against evaluation standards
User	Mission or Operating Unit's Evaluation point of contact (or designee) in the Program Office	Peer reviewer (individual who does not have a conflict of interest or who did not participate in the evaluation)

EVALUATION REPORT COMPLIANCE CHECKLIST

Evaluation Report Compliance Checklist

This Checklist is for determining if required, essential, or highly recommended elements are present in an evaluation report. It is not a means for assessing the quality of these elements. For assessing quality of a draft evaluation report as part of a peer review process, please see the Evaluation Report Review Template. For guidance on developing an evaluation report, see the Evaluation Report How-to Note and Template.

Evaluation Title:	
Evaluation Report Checked By:	Date:

I. Structure and Content	✓	COMMENTS
1. Does the opening section of the report include:		
1.1. A title that identifies the evaluation as either an impact or performance evaluation, per the definitions in Automated Directives System (ADS) 203?		
1.2. An executive summary 3–5 pages in length that summarizes key points (purpose and background, evaluation questions, methods, findings, and conclusions)?		
1.3. Table of contents?		
1.4. List of acronyms?		
2. Does the main body of the report include:		
2.1. Description of evaluation purpose, including information on:		
2.2. Why the evaluation is being conducted (purpose)?		
2.3. Who will use the results of the evaluation (audience)?		
2.4. Description of activity/project/program and background, including information on:		
2.4.1. Award number(s)?		
2.4.2. Award dates (start and end dates)?		
2.4.3. Funding level?		
2.4.4. Implementing partner(s)?		
2.4.5. Country and/or sector context?		
2.4.6. The specific problem or opportunity the evaluation was designed to address?		
2.4.7. How the activity/project/program addresses the problem?		
2.5. List of the evaluation questions?		
2.6. If an impact evaluation, are the evaluation questions about measuring the change in specific outcome(s) attributable to a specific USAID intervention?		
2.7. Description of the evaluation methods?		
2.8. Description of the limitations to the evaluation methodology?		
2.9. If an impact evaluation, does the evaluation use experimental or quasi-experimental methods to answer the impact evaluation questions?		
2.10. Findings and conclusions?		

EVALUATION REPORT COMPLIANCE CHECKLIST

I. Structure and Content	✓	COMMENTS
2.11. Does the report address all evaluation questions in the statement of work (SOW), or if not, is the exclusion of a question documented as having been approved?		
2.12. If recommendations are included, are they separated from findings and conclusions?		
3. Do the annexes include:		
3.1. The Evaluation SOW?		
3.2. A description of evaluation methods (recommended to be included in an annex when methods are not described in full in the main body of the report)?		
3.3. All tools, such as questionnaires, checklists, survey instruments, and discussion guides?		
3.4. All sources of information properly identified and listed?		
3.5. Statements of differences (if any were prepared)?		
3.6. Signed disclosure of conflict of interest forms of evaluation team members? (These are required to be signed, but only recommended to be included in the report annex.)		
3.7. Abridged bios of the evaluation team members, including qualifications, experience, role on the team, and ethical certifications (optional, required for PEPFAR evaluations)?		
II. Format and Graphing Standards		
4. Does the cover include:		
4.1. USAID standard graphic identity/brand in left area in a white field?		
4.2. A high-quality photograph representing the activity/project/program being evaluated?		
4.3. A title block in USAID light blue background color with the word "Evaluation" at the top and the report title (also including the word "evaluation") underneath?		
4.4. Month and year of the report?		
4.5. The statement "This publication was produced at the request of the United States Agency for International Development. It was prepared independently by [list authors and organizations involved in the preparation of the report]."?		
4.6. Does the title avoid acronyms that are not spelled out?		
5. Does the inside front cover page include:		
5.1. A caption explaining the "who, what, when, where, and why" of the cover photo, with photographer credit?		
6. Does the title page include:		
6.1. The report title repeated from the cover?		
6.2. The month and year of the report repeated from the cover?		
6.3. The standard disclaimer for publications by external authors: "The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development."?		

EVALUATION REPORT REVIEW TEMPLATE

Evaluation Report Review Template

This Review Template is for use during a peer review of a draft evaluation report for assessing the quality of the report. For each section of the evaluation report, the Template provides a series of questions to prompt considerations of quality during the review. A box is provided to check if the section under review should be revised, and a space is provided for comments. In providing comments during a peer review, reviewers should be familiar with what was asked of the evaluation team in the Evaluation SOW and provide actionable comments appropriate to the drafting stage of the evaluation report.

For checking if required elements of an evaluation report are simply present, please see the Evaluation Report Checklist.

Evaluation Title:		
Evaluation Report Review By:		Date:

Executive Summary	Check if revisions needed	
<i>Does the executive summary provide an accurate reflection of the most critical elements of the report, including the evaluation purpose, questions, background information, methods, limitations, findings, and recommendations? The executive summary should not add new information or contradict the evaluation report.</i>		
Comments:		

Introduction and Purpose	Check if revisions needed	
<i>Does the evaluation purpose represent the management intent (as described in the SOW)? Is it clear why the evaluation was conducted and who the primary and secondary audiences are?</i>		
Comments:		

Information and Background	Check if revisions needed	
<i>Is the information provided about the country and/or sector context for the activity/project/program sufficient to provide a reader without prior knowledge a clear understanding of the subject of the evaluation? Are the basic characteristics of the activity/project/program being evaluated adequately described? Is the geographic scope clear (preferably with a map)? Are the interventions clearly described, and is the activity/project/program's theory of change sufficiently described (preferably with a graphic and narrative description)?</i>		
Comments:		

Evaluation Questions	Check if revisions needed	
<i>Do the evaluation questions reflect the evaluation questions from the SOW? If they have been modified, does the report state that there was written approval for changes in the evaluation questions? If changed, are the new questions limited, clear, and researchable?</i>		
Comments:		

EVALUATION REPORT REVIEW TEMPLATE

Methodology	Check if revisions needed	
<p>Methodology: Does the methodology section (in report or annex) describe specific data collection and analysis methods in detail? Is it clear which methods are used to address each evaluation question (preferably through a design matrix)? Are the methods sound and appropriate for each of the evaluation questions (i.e., are the methods up to the task set forth by the evaluation questions)? Are the methods those that would generate the highest-quality and most credible evidence that corresponds to the questions being asked, taking into consideration time, budget, and other practical considerations? Are the methods based on social science methods and tools that reduce the need for evaluator-specific judgments? Does the documentation of the methods offer sufficient expectation that if another team applied the same methods, they would generate the same findings?</p> <p>Limitations: Are limitations to the methods used presented clearly and fully? Is it clear what has been done to mitigate limitations or to restrict findings to what is permissible given the limitations?</p>		
Comments:		
Findings and Conclusions	Check if revisions needed	
<p>Are all evaluation questions addressed in the main body of the report? Are findings credible—presented as analyzed facts logically linked to evidence, rather than anecdotes, hearsay, and unverified opinions? Are findings specific, concise, and supported by quantitative and qualitative information that is reliable and valid? Is it clear which quantitative and qualitative information supports which findings? Are the findings objective, such that if a different, well-qualified evaluator were to undertake the same evaluation, he or she would arrive at the same or similar findings and conclusions? If normative judgments are presented, is it clear what criteria were used to make those judgments? Are the findings clearly distinguished from conclusions and recommendations? Are the conclusions directly based on findings and evidence already presented in the report?</p>		
Comments:		
Recommendations (if included)	Check if revisions needed	
<p>Are recommendations supported by findings and conclusions? Are they clearly separated from findings and conclusions? Are recommendations action-oriented, practical, and specific? Do the recommendations assign or designate the executor of each recommendation?</p>		
Comments:		
Annexes	Check if revisions needed	
<p>Sources of information: Is the listing of sources of information in the annex clear and complete, including documents reviewed and individuals interviewed?</p> <p>Data collection tools: Are data collection tools included in the annex complete? Do they match what is described in the methods section?</p> <p>Statements of Differences: If any statements of differences are included, do the statements have merit? Did the evaluation team respond appropriately?</p> <p>Evaluation team: Is sufficient information provided about the evaluation team, including disclosure of conflict of interest statements? Are any potential conflicts of interest described, along with how they were mitigated?</p>		
Comments:		
Gender	Check if revisions needed	
<p>Do evaluation methods incorporate attention to gender relations in all relevant areas? Do findings and conclusions address gender where relevant and appropriate? If person-level outcome data are assessed, are they sex-disaggregated?</p>		
Comments:		
Overall	Check if revisions needed	
<p>Is the report structured effectively and formatted appropriately? Is it well-written and clear? Overall, is the report a thoughtful, well-researched, and well-organized effort to objectively evaluate the activity project, or program?</p>		
Comments:		

EVALUATION RESOURCE

MANAGING THE REVIEW OF A DRAFT EVALUATION REPORT

All draft evaluation reports are required to go through a review process. This document shares good practices for conducting and managing this review. An effective review process is an important step to successfully sharing, applying, and learning from evaluations.

Policy

Automated Directives System (ADS) 203.3.1.8 establishes the minimum criteria for all evaluation reports. ADS 203.3.1.2 requires that Program Office staff organize, and Technical Office staff participate in, peer technical reviews to assess quality of evaluation draft reports. The Standardized Mission Order on Evaluation outlines standards for the peer review process, but each mission may have customized aspects of the peer review. Final content of the evaluation is determined by the evaluation team.

Guidance

The Evaluation Toolkit provides several tools to assist in the process to monitor compliance of evaluation reports with the USAID Evaluation Policy (see How-To Note: Preparing Evaluation Reports; the Evaluation Report Template; and the Evaluation Report Checklist and Review Template). In addition, the Toolkit contains guidance on documenting Statements of Difference (which is part of the review process).

Types of reviews

Similar to the process for reviewing an Evaluation statement of work (SOW), there are different kinds of reviews for draft evaluation reports. The processes and purpose of these reviews often vary. Good practice is to recognize different processes that exist.

- **Compliance review.** This review is usually best conducted by the Contracting Officer's Representative (COR) of the Evaluation in the Program Office or the Evaluation Point of Contact (EPOC) to ensure the evaluation meets the standards established in the Evaluation Policy (Tool: Evaluation Report Checklist).
- **Peer review.** The COR of the Evaluation should coordinate this review. The review will include two kinds of peer reviewers:
 - *Individuals who have not directly participated in the evaluation, project, or activity but bring subject matter and/or technical expertise to an evaluation.* Working with the Technical Office, the COR of the Evaluation may choose individuals from regional and/or Washington bureaus to participate in the peer review. Staff from the Program Office may also contribute to this review, identifying issues, questioning analyses and providing suggestions to the evaluation team on the merits of methods (Tool: Evaluation Report Review Template).
 - *USAID staff managing and implementing the project or activity being evaluated:* As outlined in the Standardized Mission Order on Evaluation, no more than half of the peer reviewers should be from the Development Objective (DO) team. These individuals should review the document for factual clarifications of findings and conclusions (Tools: Evaluation Report Review Template and Statements of Difference) and ask questions about recommendations that will assist implementers to take actions based on the recommendations.

EVALUATION RESOURCE

- **Stakeholder review** (including implementing partners, alliance partners, host-country government partners, and others). Similar to the review by USAID staff involved directly in the project or activity, this group of reviewers should identify factual clarifications or address any limitations that the evaluators may have noted in the report (e.g., missing documentation, data, etc.). Likewise, they should be able to ask questions about recommendations relevant to their work. These reviewers may prepare a Statements of Difference following the review of the draft report.

What is the value of the review of the draft evaluation report?

There are numerous reasons to engage stakeholders, USAID staff, and peers in a review beyond its requirement by USAID. These include:

- Ensuring that the required elements are included.
- Improving the overall quality of the evaluation. The peer review process can involve experts in evaluation and the technical subject area of the evaluation.
- Increasing the independence and objectivity of the evaluation. By bringing in additional staff members from other parts of the mission or Agency, the response to an evaluation can benefit from perspectives that are not too close to the activity, project, or program being evaluated, thereby promoting a more neutral and unbiased perspective.
- Facilitating buy-in from internal stakeholders regarding the content of the evaluation. Evaluations are only as worthwhile as their utility to the users. By bringing in the primary audiences who may use the evaluation results, the review can help ensure that, without creating bias, the evaluation report meets the needs of these audiences.

Before the draft evaluation report is received

There are several steps that the individual managing the evaluation in the Program Office can take to streamline the review process:

1. Document any minor adjustments made in carrying out the evaluation from the original SOW or through a letter modification by the Contracting Officer. This is critical, given that a peer reviewer assessing the quality and compliance of the evaluation report will not necessarily know all of the decisions that have been made amid changing circumstances, problems with reaching sites, etc.
2. Engage with the Program Office, Technical Offices, and Contract Office to update them on the evaluation process.
3. Share the Statements of Difference guidance with the stakeholders who are reviewing the draft evaluation report so they understand how to incorporate their comments into a draft evaluation report.
4. Set aside time to synthesize and integrate the comments that will be sent to the evaluation team leader.

Managing the review process

Once the evaluation report is received and is ready to be shared beyond those most directly involved in the drafting, the COR of the Evaluation should organize the review process. While there is no standard way of conducting a review, some standard practices are outlined in the Standardized Mission Order on Evaluation. Mission staff should consult their own Mission Order on Evaluation for peer review practices specific to their mission. Below are some key questions to answer when conducting a peer review:

When to conduct the reviews?

When planning the reviews, missions should aim for conducting it after a **full draft of the evaluation** has been completed and incorporates the evaluation team's full analysis. Furthermore, the COR for the Evaluation should avoid circulating an evaluation draft report that he or she finds incomplete.

EVALUATION RESOURCE

Instead, he or she should consider returning the evaluation to the evaluation team with instructions for completing the report. A complete draft will streamline the review and ensure that the reviewers focus on substantive issues within the report rather than on formatting or missing sections.

While divergent comments may be useful to an evaluation team in formulating a final evaluation report, the evaluation manager should avoid sending comments that are contradictory and/or fit the criteria of statements of difference. When scheduling the peer review, the COR for the Evaluation should ensure that sufficient time is allowed after the peer review to compile, consolidate, and prioritize comments prior to returning the draft to the evaluation team.

How many and which individuals will be involved in the peer review?

As noted in the Standardized Mission Order on Evaluation, the mission's Evaluation POC in the Program Office (or their designee) will lead a peer review. This is normally the COR for the Evaluation. He or she should take lead responsibility for ensuring that the evaluation report meets the procedural standards and requirements of ADS 203 and the Evaluation Policy so that other peer reviewers can focus on content.

The peer review should include no fewer than two individuals in addition to the COR for the Evaluation or EPOC (or designee). Emphasis should be placed on finding at least one peer reviewer with evaluation methods expertise. Peer reviewers may include individuals from the DO team and Program Office as well as USAID/Washington regional and technical bureaus, the Bureau for Policy, Planning, and Learning Office of Learning, Evaluation, and Research, and local partners. It is best practice, however, for no more than half of the peer reviewers to be from the Technical Office that oversees the project or activity being evaluated. USAID/Washington regional bureaus have a particular responsibility to participate in peer reviews when requested by missions. Mission staff should consult the [M&E POC's List](#) for Washington bureau contact information.

How much time will be provided for the peer review?

The Mission Order on Evaluation in each mission should specify the length of time peer reviewers will have to review the draft evaluation report. This timeline is often established in the contract with the evaluation team. Best practice is to allow 10 business days for comments. Some USAID/Washington offices may have their own standards for how long they typically take to review a draft evaluation report, so check with them if considering including USAID/Washington staff members in your peer review.

How will comments be received for the peer review?

There are a variety of ways of structuring the peer review process. Missions may choose to have a peer review meeting where individuals can discuss their comments on the draft evaluation report, request written comments on the draft, or both. However, note that circulating the document in Microsoft Word for reviewers to insert their comments often leads to excess comments and/or contradictory comments.

Many missions choose to ask the peer reviewers to fill out standard review sheets or checklists (see Evaluation Report Checklist and Review Template) while others prefer reviewers to send comments utilizing a standardized format.

Regardless of the method chosen, the COR for the Evaluation should provide clear instructions to the peer reviewers regarding the means for providing comments on the draft evaluation report.

EVALUATION RESOURCE

After the review

The COR for the Evaluation should consolidate and share the comments with the Evaluation Team. The use of the Statement of Difference documentation should typically be used for comments from USAID staff and implementing partners involved directly in the project or activity.

As noted in the Standardized Mission Order on Evaluation, if Statements of Difference are received, the COR for the Evaluation should share them with the evaluation team. The evaluation team should be given the opportunity to revise and/or respond to the statements in the evaluation report.

For more information

For more information on the peer review process, check out the following:

- Standardized Mission Order on Evaluation
- ADS 203.3.1.8: Documenting Evaluations

From the Toolkit:

- Guidance: [How-to Note: Preparing Evaluation Reports](#). This Note provides current good practice in preparing evaluation reports, the main deliverable for most evaluations. These practices also serve as a guide for reviewing the quality of draft evaluation reports submitted by the evaluation team.
- Tool: [Evaluation Report Template](#). This Template is an optional tool to help improve consistency of the evaluation report with USAID formatting standards.
- Tool: Evaluation Report Checklist and Review Template. This Template includes two tools: a checklist for compliance and a peer review template. The tool includes guidance for users on each tool and the correct usage.
- Guidance: [Statements of Difference](#). Each USAID evaluation report should include any Statements of Difference as an annex. These statements describe any significant unresolved difference of opinion on the part of funders, implementers, and/or members of the evaluation team.

EVALUATION RESOURCE

USAID EVALUATIONS: STATEMENT OF DIFFERENCES QUESTIONS AND ANSWERS

Each USAID evaluation report should include any Statement of Differences as an annex regarding any significant unresolved differences of opinion on the part of funders, implementers and/or member of the evaluation team.

Who may write a statement of differences?

At minimum, the following may write a statement of differences and have it included as an annex to the evaluation report.

1. Members of the Evaluation Team
2. Funders of the Evaluation (including the USAID Operating Unit funding the evaluation and/or other organizations funding the evaluation).
3. Funders of the project or activity being evaluated (including the USAID Operating Unit funding the project or activity and/or other organizations funding the project or activity).
4. The organization implementing the project or activity that is being evaluated.

The evaluation COR/Manager or the operating unit leadership may determine if other stakeholders are relevant and should have the opportunity to write a statement of differences

When should the opportunity to prepare a statement of differences be provided?

In order to ensure that these organizations and individuals have an opportunity to review the report and prepare a statement of differences, it is appropriate to provide them with a final draft of the evaluation report. There is no set amount of time, during which a statement of differences must be received, but providing one or two weeks following the receipt of the report is a reasonable time frame.¹ Evaluation CORs/Managers should ensure that adequate time has been budgeted to allow for appropriate stakeholders to prepare a statement of differences and for the evaluation team to prepare a response, if needed.

What should USAID missions tell the relevant stakeholders about writing a statement of differences?

There are a few points evaluation managers may note to those potentially writing a statement of differences to help keep the process manageable:

- The draft report that is the basis for a statement of differences is subject to change at the discretion of the authors. Those organizations who have received a draft report should not share or distribute the draft outside their organization.
- Writing a statement of differences is optional and no individual or organization is required to provide one.
- The statement of differences should be made in writing, signed, and submitted electronically.
- The statement of differences is a one-time only opportunity. Statements from a single funder or implementer should be a single, corporate response. Letters from individuals, other than evaluation team members, or additional statements at a later stage in the evaluation process will not be accepted.

¹ For comparison, GAO generally gives an agency from 7 to 30 calendar days to comment on a draft report.

EVALUATION RESOURCE

- The statement of differences should focus on errors of fact and differences regarding the findings, conclusions, and recommendations stated in the evaluation report. The basis for any differences with the evaluation report should be described.
- The statement of differences will be included as an annex to the report.
- The statement of differences should be limited in length. A 10 page limit should be sufficient in most cases.

Should the evaluation team respond to the statement(s) of differences?

Evaluation teams are not *required* to respond to any statement of difference that have been received. However, the evaluation team should be given the opportunity to revise the evaluation report and/or respond to the statement(s) of differences in the evaluation report, if they so choose. Ideally, the evaluation workplan will include time to review and respond to any statement of differences submitted to the evaluation team. After the evaluator responds to the statement(s) of differences, the statement of differences process stops. Only the original Statement of Differences and the response from the evaluator should be included in the evaluation report annexes.

If factual errors have been noted in a statement of differences and the facts are acknowledged as true, it is reasonable to ask the evaluator team to correct them in the main body of the report before submitting the final version. In these cases, the evaluator's response to the statement(s) of differences should then note that steps were taken to correct the factual errors. However, if the factual errors are disputed by the evaluator or there is a difference of interpretation about facts, findings, conclusions, or recommendations, the evaluator should be given the liberty to decide if they consider it appropriate to change the report or not.

EVALUATION RESOURCE

Sample Letter

USAID Mission

Date

Dear [.....],

We are pleased to provide you with a final draft of the Evaluation report titled [.....]. As a [funder of the project being evaluated/funder of the evaluation/implementer of the project being evaluated], we are providing this review copy of the evaluation prior to its completion and dissemination. Please do not share or distribute this draft outside your organization.

If, upon reviewing this evaluation report, you find errors of fact or you have differences of opinion regarding findings, conclusions, and recommendations, you may write a *Statement of Differences* that will be appended to the final report.

In completing a Statement of Differences, please be aware of the following.

- This draft report is not completely final and is therefore subject to change at the discretion of the authors.
- Writing a statement of differences is optional and no individual or organization is required to provide one.
- The statement of differences should be made in writing, signed, and submitted to USAID electronically.
- The statement of differences is a one-time only opportunity. Statements provided after the deadline listed below or additional statements will not be accepted.
- Statements of differences should be a single, corporate response. Letters from individuals will not be accepted.
- The statement of differences should focus on errors of fact and differences regarding the findings, conclusions, and recommendations stated in the evaluation report. The basis for any differences with the evaluation report should be described.
- The statement of differences should be no more than 10 pages in length.

Statements of differences will be accepted until [Date]. Any statements provided after that date will not be accepted.

EVALUATION RESOURCE

SUBMITTING A USAID EVALUATION TO THE DEVELOPMENT EXPERIENCE CLEARINGHOUSE

Following the completion of a USAID evaluation report, the Program Office of the Operating Unit responsible for the evaluation must ensure that the final evaluation report is posted on the [Development Experience Clearinghouse](#) (DEC) no later than three months after completion. Exceptions to this requirement are granted in only very rare circumstances (see guidance on Exemptions to Public Disclosure of USAID-Funded Evaluations).

It is highly recommended that USAID staff post the evaluation to the DEC rather than relying on a contracted evaluation firm to post the evaluation report to the DEC.

To submit an evaluation to the DEC:

1. Go to <https://dec.usaid.gov>
2. Click the “Submit” tab in the center of the page.
3. Select “Submit Documents” on the left of the page.
4. Read and follow the guidelines provided. For evaluations with Annexes in multiple files, please submit according to the special guidelines provided.
5. Fill out the form with the information from the document, including Title (as it appears on the title page), Publication Date, and Language. If you have other information about the document, such as a contract number, please fill in the appropriate fields on the form. This helps DEC staff with review and processing.
6. Select the correct “Document Type” from the drop-down menu:
 - “Final Evaluation Report” for end-of-activity/project evaluation reports;
 - “Special evaluation” for mid-term evaluations or any other evaluation reports.
7. Upload the document! (People frequently forget this step.)
8. Don't forget to click “Save” when you are finished. (People also frequently forget this step.)

After the document has been uploaded, the DEC will create a document page with the information you entered about the evaluation report. Copy the URL from the “View Document” field of this page. Share the URL with mission staff, the evaluation team, implementing partners, and other stakeholders. Also, keep the URL in the Program Office records so that it is readily available for input into the Evaluation Registry during the annual PPR preparation.

Policy Reference

Automated Directives System (ADS) 203.3.1.10 states:

“Evaluation reports must be provided to the [DEC]: dec.usaid.gov within three months of the evaluation’s conclusion. The evaluation reports will be accessible for use in planning and assessing other programs.”

Exception: In cases where national security considerations and/or proprietary information may be involved, USAID Missions/Offices may request an exception from this requirement. Exception requests should be submitted to the Bureau for Policy, Planning, and Learning, Office of Learning, Evaluation, and Research.”

EVALUATION RESOURCE

EXCEPTIONS TO PUBLIC DISCLOSURE OF USAID-FUNDED EVALUATIONS

USAID is committed to full and active disclosure of evaluation reports, methods, findings, and data produced by the Agency or partners receiving USAID funding. This is guided by Agency policies and directives, including the [USAID Evaluation Policy](#) and [ADS 540 – Development Experience Information](#). These direct that evaluation final reports (or reports submitted by evaluators to USAID as their final drafts) must be submitted within three months of completion to the Development Experience Clearinghouse (DEC). PPL/LER will work with the Office of the General Counsel to review and clear any exceptions to this requirement.

Principled exceptions to this requirement will be rare, must be requested and justified in writing, and should be based on one or more of the following criteria:

In the judgment of the USAID Mission or Office that commissioned the evaluation,

1. Disclosure would impair foreign assistance objectives.
2. Disclosure would involve releasing otherwise proprietary information owned by third parties.
3. Information disclosed could put individual safety at risk or release private personal information.

Exceptions to public disclosure will **not** be made for the following reasons:

- Findings are unexpected, negative and/or embarrassing to USAID.
- USAID or partners disagree with the findings. In this case, a Statement of Differences may be attached as an annex explaining the disagreement.

Requests for an exception from making an evaluation report publically available on the DEC should be sent as a decision memo to the director of PPL/LER. Please see [ProgramNet for email contact information](#).

Memo requests should:

- Be from the relevant USAID Mission Director or Washington-based Office Director.
- Provide a clear and compelling written justification based on one or more of the three criteria.
- Be accompanied by a copy of the Evaluation report (electronic attachment is fine) with the information of particular concern identified or highlighted.
- Specify the requesting office's preferred outcome (redact report, submit to SBU side of DEC, other)

PPL/LER will provide one of the following decisions after reviewing the request:

1. The request is denied.
2. The request is approved by PPL/LER and GC. The requesting office will be asked to do one or a combination of the following: submit a redacted version of the report to the public DEC; submit the report in its entirety or in redacted form to the SBU side of the DEC; the report will not be submitted to either the public or SBU side of the DEC.

EVALUATION RESOURCE

WAREHOUSING EVALUATION DATA

USAID's Evaluation Policy and ADS 203.3.1.10 state that "All quantitative data collected by USAID or one of the Agency's contractors or grantees for the purposes of an evaluation must be uploaded and stored in a central database... The data should be organized and fully documented for use by those not fully familiar with the project or the evaluation."

USAID uses the Development Data Library (DDL) as the central database for storing quantitative evaluation data. Per ADS 579.3.2, the DDL is the Agency's repository of USAID-funded, machine-readable data created or collected by the Agency and its implementing partners. While evaluation reports are to be submitted to the Development Experience Clearinghouse (DEC), quantitative datasets should *not* be submitted to the DEC, but instead submitted to the DDL.

1. **How do I submit data?** Begin the process by completing the form here:

<http://www.usaid.gov/data/DDLsubmissions>

2. **What evaluation data should I submit?**

Data Type	Examples
Performance monitoring and USAID-funded evaluation data	Lot quality assurance, cluster samples, facility surveys, school attendance, crop and weather monitoring
Quantitative data underpinning evaluation findings	Baseline, midterm, final surveys
Qualitative unstructured data (Optional)	Interview responses or focus group notes
Supporting Documentation	Codebooks; Data dictionaries; Forms, templates, and data gathering tools; Explanations of redactions; Notes on data quality, data limitations, or data context; Data gathering methodologies

3. **What format should data be in?** Datasets submitted to the DDL must be in non-proprietary, machine-readable format.

Do submit: Spreadsheets in a non-proprietary format such as comma-separated values (CSV), extensible markup language (XML), JSON

EVALUATION RESOURCE

Don't submit: Documents using proprietary software owned by companies such as Adobe, Intuit, Lotus, or Microsoft, unless you have **also** included the file in a non-proprietary format. Datasets should not be submitted using word processing software, hypertext markup language (HTML), or portable document format (PDF). Supporting documentation (e.g. codebooks, interview notes, etc.) may be submitted using word processing software.

4. **Who can access the data?** Please see our Frequently Asked Question #34 for more information on this topic: <http://www.usaid.gov/data/frequently-asked-questions#Q34>
5. **What is the clearance and publication process?** After completing the submission form, you will receive instructions on uploading the data and initiating the clearance process. The USAID operating unit is responsible for obtaining clearance from:
 1. Office of Security
 2. Bureau for Management, Office of the Chief Information Officer, Information Assurance Division / Compliance and Risk Management Branch
 3. Bureau for Management, Office of Management Services, Information and Records Division
 4. Operating Unit of Origin and any other relevant operating units
 5. Office of General Counsel or Resident Legal Officer

Prior to public release of information, the COR/AOR or designee will redact any sensitive data specified by the clearing officials above, particularly if it violates any law or policy, or jeopardizes privacy, confidentiality, or national security.

6. **Who should submit the data?** USAID staff, contractors, and recipients of USAID assistance awards must submit any dataset created or collected with USAID funding to the DDL, in accordance with the terms and conditions of their awards. Operating Units are responsible for including datasets as deliverables in USAID awards, when writing statements of work or program descriptions. They must appoint and maintain a [Data Steward](#) plus one alternate. Operating units or implementing partners can submit data to the DDL.
7. **Where do I go if I have questions?** Please first consult our [Frequently Asked Questions about USAID's Open Data Policy](#). We also encourage anyone to post questions that may benefit from a public response to our site on [StackExchange](#). USAID staff will be notified when a question is posted and will promptly reply with a public response. All other questions can be addressed to opendata@usaid.gov.

Is there any additional guidance? Further Guidance is available in [ADS 579 – USAID Development Data](#). Guidance on specific data-related topics (e.g. data protection, data dictionaries, etc.) will be released directly on www.usaid.gov/data as it becomes available.

EVALUATION RESOURCE

THE USAID EVALUATION REGISTRY: *QUESTIONS AND ANSWERS*

What is an evaluation registry?

The Evaluation Registry is an annex to the annual Performance Plan and Report (PPR). It includes information on completed evaluations during the previous fiscal year and ongoing and planned evaluations for the current fiscal year, plus two out years. The annual PPR guidance from the Office of the Director of U.S. Foreign Assistance (F) indicates the specific information that is to be supplied about each evaluation.

How do I populate/update the PPR Evaluation Registry? Where do I start?

The PPR Evaluation Registry should draw from a Mission's Evaluation Plan, which is a required part of the Mission-wide Performance Management Plan.

What types of evaluations should be included in the Mission's Evaluation Plan and in the PPR Evaluation Registry?

Missions should include **all types** of evaluations in the Mission Evaluation Plan and, correspondingly, in the PPR Evaluation Registry: required and non-required; external and internal; performance and impact. Do not include assessments, reviews, studies, etc. that are not evaluations as defined by the Automated Directives System (ADS) 203 and the Evaluation Policy.

Should I really include internal evaluations in the PPR Evaluation Registry?

Yes. There are two types of internal evaluations at USAID. The first type of internal evaluation is one that is led by a USAID staff member. The second type of internal evaluation is one that is led by someone with a fiduciary responsibility to the implementing partner conducting the project or activity being evaluated, for instance when an implementing partner subcontracts a firm to conduct an evaluation. As long as USAID funds are being expended to conduct these evaluations, they are still USAID evaluation and both types of these internal evaluations should be included in the Evaluation Registry along with external evaluations.

How do I enter data on an impact evaluation that spans many years?

For an impact evaluation (or any evaluation) that spans multiple years, the status of the evaluation should be listed as ongoing for the fiscal year during which the evaluation began and for subsequent years until the year in which the evaluation was completed, at which point the status changes to "completed." Baseline data collection and other deliverables that are conducted as part of an impact evaluation should not be reported separately. The entire costs of the evaluation, including data collection, should be included in the budget field under one entry in the Evaluation Registry.

Is there a way to distinguish between various types of evaluations in the PPR?

Missions must distinguish required vs. non-required evaluations in the PPR Evaluation Registry. In addition, the FY2015 PPR Evaluation Registry will require missions to distinguish between external and internal evaluations.

Why does the PPR Evaluation Registry include a field for an estimated "evaluation budget"? Is it really important to fill that out?

Estimating the resources required to undertake an evaluation is essential to good planning and should be a key component of a Mission's Evaluation Plan. But, since evaluation budget estimates likely will change due to any number of factors, some Missions have hesitated to include this information in the PPR

EVALUATION RESOURCE

Evaluation Registry. Nevertheless, filling out this field is critical for several reasons: When aggregated across the Agency, the budget estimates help USAID get a sense of the totality of resources being devoted to evaluation, track trends over the fiscal years, and help us demonstrate to external stakeholders, such as the White House's Office of Management and Budget, the priority that USAID places on evaluation (albeit using a monetary proxy). Once an evaluation is complete, the budget number should be updated with the actual costs.

Is the PPR Evaluation Registry being used to set USAID Forward targets and determine actuals?

Yes. Since FY2013, the PPR Evaluation Registry is used to track USAID Forward targets and actuals. For evaluations, the data will be pre-populated. For example, the FY2015 USAID Forward evaluation target will be calculated using data entered into the previous year's PPR Evaluation Registry for planned and ongoing evaluations, and the FY2015 actuals will use this year's Evaluation Registry to count the number of evaluations completed. More information will be included in the annual PPR Guidance. Please also see Cindy Clapp-Wincek's 12/2013 blog post on this topic, which is still relevant:

<https://programnet.usaid.gov/blog-entry/some-words-advice-preparing-your-pprs-and-evaluation-targets> (*accessible for USAID staff only*).

Policy references regarding the Evaluation Registry:

The Evaluation Policy states, "Operating units will provide information online in a fully searchable form about the initiation of evaluations and expected timing of release of findings. This information will be included in the annual PPR and communicated to the public on the USAID website."

ADS 203.3.1.10 states, "USAID Missions/Offices will provide information through FACTS [Foreign Assistance Coordination and Tracking System] Info about completed evaluations and the initiation of evaluations and expected timing of release of findings. This information will be included in the annual [PPR] Evaluation Registry and communicated to the public on the USAID website."

ADS 203.3.14.3 states, "All USAID Missions/Offices and Washington Operating Units are required to submit an inventory of evaluations conducted during the previous year in their annual [PPR] in the Evaluation Registry (an annex to the PPR in FACTS Info). The Registry also requires planned evaluations and estimated budgets for the coming fiscal year, plus two out years. This is in addition to the requirement to submit all evaluation reports to the Development Experience Clearinghouse."

EVALUATION RESOURCE

UTILIZING AND LEARNING FROM EVALUATIONS

The value of an evaluation is in its use. Evaluations should inform decision-making, contribute to learning, and help improve the quality of development programs. This guidance is intended to assist USAID program and evaluation managers in ensuring that evaluations serve these purposes and support improvement and learning.

Different types of use

The use of evaluations can take different forms. The literature identifies several types of use:

- **Instrumental use** involves the direct use of an evaluation to make decisions or to adapt or change some aspect of a program. For example, USAID may use the findings from a mid-term evaluation of an activity to update some aspect or aspects of the program for the remainder of its implementation, or the findings from the final evaluation of an activity may be used in the design for a follow-on project or activity.
- **Conceptual use** refers to the development of enhanced knowledge or improved understanding of how programs work. For example, an evaluation may contribute to the learning agenda for a Mission's Development Objective (DO) team or Technical Office.
- **Process use** involves gains for those who participate in the evaluation, regardless of the findings, and could include a better understanding of the program being evaluated or the process of evaluation. USAID may use an evaluation to help build the capacity of a local government counterpart or its own staff—for example, by including select individuals on the evaluation team.
- **Symbolic use** refers to token use, intended only to fulfill a requirement or provide justification for actions already taken. Symbolic use implies a “box-checking” exercise, and is not a type of use generally encouraged by USAID or other donors.

These types of uses are not mutually exclusive: A single evaluation may be used in several ways.

Planning for use

Use can happen at various stages in the evaluation process. Evaluation use may occur before or during the evaluation, shortly after it is completed, or long after the findings have been presented. At whatever stage it occurs, if evaluations are to be used as intended, utilization should be planned for.

Identify users and uses. The first step in planning for use is to identify the intended users and the potential uses for the evaluation. Often, USAID commissions evaluations to inform the design of a project or activity (instrumental use), to influence the future direction of an ongoing activity (instrumental use), or to determine the effectiveness of a particular approach

EVALUATION RESOURCE

(conceptual use). Sometimes, the evaluation results may be used to influence policy (instrumental use). In these cases, the intended users are USAID, implementing partners, and policymakers. In other cases, intended users may include civil society organizations, local communities, or other donors.

The planning team should consider conducting a stakeholder analysis to identify the range of potential users and uses. They may also find it helpful to distinguish between primary, secondary, and tertiary users and uses. For example, if USAID/Zambia were planning to conduct a final evaluation of one of its health activities, the Mission may identify their health office, the associated DO team, and the implementing partners as the primary users of the evaluation. The Global Health Bureau in Washington may be a secondary user, and the American and Zambian public may be tertiary users.

Consider the timing of use. Consideration of use should affect the timing of the evaluation. For example, if an intended use is to inform the design of an upcoming project, the evaluation should be complete before USAID begins the project design process. Or if USAID hopes the evaluation results will influence policymaking in the host country, the evaluation should be timed so that results are available at the appropriate stage of the legislative calendar.

Engage users in evaluation planning. To the extent possible, the intended users should be involved in the planning process by developing or reviewing evaluation questions, raising timing considerations, and contributing to the development of the scope of work for the evaluators. In addition to bringing USAID staff, local government counterparts, partners, or others into the process as intended users, the planning team may consider, for example, co-funding an evaluation with another donor if that donor is an intended user or has relationships with other intended users.

Conduct simulations of use. The team may consider using simulations during the evaluation planning stage to further engage intended users and to enhance the likelihood of future use. This activity involves considering several potential outcomes of the evaluation and developing scenarios based on those outcomes (see box). The planning team and intended users would think through the decisions and action points triggered by the potential evaluation findings.

Simulation of Use

An evaluation planning team would follow these steps to conduct a simulation of use exercise:

1. Devise several potential findings based on the proposed evaluation design and knowledge of the program.
2. With the primary intended users, hold a guided discussion to interpret the potential findings and discuss the decisions or actions that might result.
3. Interpret the simulation to determine if changes to the evaluation design, the evaluation team, data collection, or other aspects of the process would increase utility.
4. With the primary intended users, make an explicit decision to proceed with the evaluation give the likely costs and uses—before data collection begins.

Source: Utilization-Focused Evaluation Checklist. Evaluation Checklists Project. Michael Quinn Patton, January 2002.

EVALUATION RESOURCE

Simulations can help users and stakeholders determine if they should adjust the evaluation questions or scope of work. They can also foster a greater sense of ownership over the evaluation process and prompt users to seriously consider whether an evaluation is worth doing given the costs and the realistic range of uses.

Facilitating use

USAID and other stakeholders are more likely to use the evaluation findings if utilization is facilitated. The following strategies can assist the team in facilitating utilization;

Engage potential users in conducting the evaluation. As the previous section notes, engaging intended users in the evaluation process helps increase the likelihood of future use. Evaluation specialist Michael Quinn Patton notes, “Intended users are more likely to use evaluations if they understand and feel ownership of the evaluation process and findings; they are more likely to understand and feel ownership if they’ve been actively involved; by actively involving primary intended users, the evaluator is... preparing the groundwork for use and reinforcing the intended utility of the evaluation.” (Utilization-Focused Evaluation Checklist, January 2002.) To enhance the feeling of ownership, the USAID team may want to include an intended user or other stakeholder on the evaluation team. Such participation may also facilitate process use, such as building an evaluation culture or strengthening the capacity of stakeholders to conduct evaluations.

Engage potential users in developing the evaluation recommendations. Most evaluations include recommendations, but developing recommendations need not be the sole responsibility of the evaluator. Recommendations developed by an external evaluation team often lack an insider perspective of what is politically or operationally feasible, thus hindering their chances of being used.

Evidence-based evaluations add considerable value to USAID’s learning efforts, but the recommendations for moving forward in response to the evidence can and often should be an inclusive effort. Including the evaluation users in developing recommendations will help “ground-truth” their feasibility and improve the chances that they will be acted upon. The USAID team commissioning the evaluation may consider a facilitated discussion on recommendations in which the evaluation team presents the evaluation findings and potential recommendation options, intended users review and reflect on the findings, and all parties consider and agree upon the final evaluation recommendations to be included in the evaluation report or a separate annex.

Track recommendations and follow up. Following up on recommendations is another way to facilitate and track utilization. USAID staff may use the Post-Evaluation Action Review Template to describe the actions associated with each accepted recommendation and assign responsibility for follow up. During subsequent Portfolio Reviews, the status of the action plans and their use in respective decisions is discussed and documented. An action plan tracker also helps to document how an evaluation was used, should that information be needed for reporting, audits, or other accountability purposes.

EVALUATION RESOURCE

Ensure that resulting products are appropriate for intended users. When considering the intended users during the planning process, the team should also think through the products that would be most appropriate for each audience. Are the intended users interested in reading the entire report? Or would a one-page summary be more helpful? Would the use of infographics or other data visualization techniques help convey the findings?

In addition to a report, the planning team may consider other types of products that might be helpful for intended users, including videos, one-or multi-page briefs, presentations, or posters. USAID may consider incorporating some of these activities—such as developing a multi-pager or creating key graphics—into the scope of the work for the evaluators, if appropriate. Consult the section of the Toolkit on Dissemination Plan Guidance for more information.

Reflect on the evaluation process. As with any other activity or effort that involves significant effort, USAID should consider holding an [After Action Review](#) (AAR) soon after the evaluation is complete to reflect on the experience, discuss what happened, and gather lessons to apply in future evaluations. AARs are especially valuable for evaluations that included process use as a goal so that participants in the evaluation have the opportunity to discuss and reflect on what they learned and how they plan to use that knowledge.

Consider use throughout the Program Cycle. The planning team should consider the Program Cycle and where and how an evaluation might contribute to the various phases. For example, the evidence from an evaluation could be used in the design of a project, be cited in a Country Development Cooperation Strategy, contribute to the evidence base for the development of an Agency policy or implementation guidance, or inform portfolio reviews.