THREE KEYS TO GETTING EVALUATION QUESTIONS RIGHT: FEASIBILITY, SCOPE, AND CLARITY

Evaluation questions form the heart of every USAID performance evaluation. A set of clear, realistic questions clearly linked to the evaluation purpose can result in detailed findings and actionable recommendations that improve program outcomes. Inversely, questions that are unwieldy, unrealistic, or difficult to understand can make a mess of even the best methodology. Research conducted by The Cloudburst Group at the request of USAID's Bureau for Democracy, Human Rights, and Governance (DRG), identified three key elements—**feasibility, scope, and clarity**—where performance evaluation questions commonly fall short.

The following guidance provides tips and examples for how to improve performance evaluation questions across all three elements. For more information about developing successful performance evaluation questions, check out the Addressing Learning and Evaluation Challenges - Performance Evaluation Questions findings report and accompanying glossary of common evaluation question terms.

FEASIBILITY | CAN THE QUESTION BE ANSWERED WITH THE METHODS AVAILABLE?

There is a common tension in developing evaluation questions between asking questions evaluators can feasibly answer with a PE methodology and asking questions that USAID needs answers to for upcoming programmatic decision-making. From one perspective, USAID should only ask performance evaluation questions that are answerable with the available methodologies, but from the other perspective, USAID staff should ask the questions for which they need answers. While both perspectives are valid, the more difficult the question is to answer with the available methods, the more likely that evaluation findings will have low levels of confidence and might lead to incorrect conclusions.

Traditional PEs are typically a combination of formative and summative evaluations. These evaluations can answer questions about processes and outcomes, comparative questions, and questions about program assumptions and theories of change. They cannot answer causal questions about program effects, which require an impact evaluation. If evaluation commissioners ask questions that cannot be feasibly addressed through a PE, the evaluation team's answers will be based on inadequate evidence and will likely be vague, inconclusive, or incorrect.

Question types that are feasible under most performance evaluation methodologies include:

- Questions about program **processes.**
- Questions about program **outcomes for program participants** (provided they can be measured with available methods or pre-existing data—ask an evaluation specialist to confirm).

- Questions about program **assumptions and theories of change.**
- Questions that **compare program aspects.**
- Questions about **bottlenecks, challenges, and opportunities**.
- Questions about **lessons learned.**

The questions below are methodologically feasible but require data sources that go beyond the scope of most performance evaluations, which is addressed in the next section. Ask these questions with caution and ensure there are adequate resources to address them!

- Questions about **outcomes for indirect participants or non-program participants (again provided outcomes can be measured)**
- Questions **comparing the evaluated program to other similar programs.**

SCOPE | ARE USAID PROJECT ACTIVITIES SPECIFIED IN REASONABLE QUANTITIES/TIMEFRAMES?

USAID staff often want to understand the program as a whole or want the evaluation to speak to the many aspects of a program. This can lead to evaluation findings that are shallow and have limited confidence and limited usefulness. By focusing questions on specific aspects of a given program, theory of change, or implementation that are most important for informing decision-making, the evaluation team can look deeply into the most important issues and report their findings with confidence. To do this, commissioners should:

1. **Identify the decisions the evaluation will inform and write questions aimed at providing this information.** For example, if a program team will use an evaluation to design the follow-on program, these are some of the decisions they may want data to inform

   - Which program activities should be continued?

   - Are there any gaps in the current programming?

   - Is the program working with the correct stakeholders? What other partnerships should be explored?

   - Is the program reaching its target population? If not, how could targeting improve?

   - Could the theory of change be strengthened? Are there any gaps in knowledge about the theory of change?

2. **Prioritize collecting information about the program components where there is most uncertainty.** For example, if monitoring, evaluation, and learning (MEL) data provides clear information about the effectiveness of some activities, focus questions on other activities. If there is uncertainty about the theory of change, ask a question to validate the assumptions of a part of the theory of change.

3. **Questions that require in-depth comparisons to other donors, organizations, or programs can be valuable —but require more resources to answer.** If this

information is critical for decision-making, ask the question, but ensure there are adequate resources to answer it by either reducing the number of other questions asked or adding resources to the evaluation.

4. **If asking a question about lessons learned, make sure it is tailored to specific activities or program components most relevant to future decision-making.** Avoid general questions about lessons learned, which frequently duplicate the findings and recommendations from other questions, and provide clarity on the types of lessons of interest. For example, ask about lessons learned in addressing a known and difficult programmatic challenge.

5. **Limiting to 3–5 questions per evaluation is a good rule of thumb, but the data sources and methods for answering each question matter more.** If multiple questions can be answered from the same data source (KIIs with the same stakeholders, a survey, MEL data, etc.), the evaluation team can answer more questions. If a key question has a larger scope, such as comparisons to other programs or a question about the whole of the activity, even 1–2 questions may be all the evaluation team can answer satisfactorily.

6. **If you must ask a question—or set of questions—with a large scope, recognize the trade offs.** The evaluation team will have less confidence in the findings and the report will only go into limited depth. As such, it might not tell commissioners anything new. Increasing timelines and budgets can also make it possible to answer questions with a wider scope.
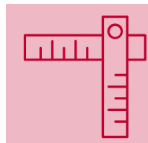
CLARITY | IS THE QUESTION PRECISE?

The third element of a successful question is **clarity**. If evaluation questions and terms used are not clear, evaluation teams might not provide the answers commissioners expect. To ensure questions are clear and precise, follow these guidelines.

1. **Express the question itself as a single clear, precise sentence free of jargon. Provide additional context alongside the question rather than as part of the question.** Include additional context to guide the evaluation team, including identifying lines of inquiry and potential hypotheses, providing examples, or giving additional information. This can be done in a paragraph placed under each evaluation question or included elsewhere in the evaluation scope of work.

2. **Define all terms in the question.** What do you mean by resilient? Sustainable? Effective? Include clear definitions of any key terms used in the question. For more definitions of common evaluation terms commissioners can insert into questions, refer to the illustrative glossary of common evaluation terms.

3. **Use compound questions and subquestions sparingly and thoughtfully.** Evaluation commissioners will sometimes join two separate questions as one. This can make sense if the two components are clearly asking about the same component (such as "why or why not" or "To what extent"), but in most cases, the two questions are better addressed separately, either as a subquestion, if the follow-on question can be answered with the same data sources and methods, or as an entirely new question.

TOPIC	ILLUSTRATIVE QUESTION
Outcomes, effectiveness, and program objectives	Were the specific approaches developed under Objective 1 very effective, effective, less effective, or not effective at producing engaging media content? *Focusing scope through "specific approaches developed under Objective 1," while clarifying measurement terms with four rating categories and defining "effective" as "producing engaging media content." The term "engaging" should be defined elsewhere.* What have been the bottlenecks and opportunities in the activity approach to building legitimate, trustworthy, and responsive relationships between police and communities? *Focusing scope through "approach to building […] relationships between police and communities," and providing clarity elsewhere with definitions of "legitimacy, trustworthy, and responsive."* Which collaborative activities were valued most by journalists participating in the activity? Which of the three program approaches appear to be the most promising at enabling participating children to live in family care? To what extent has the activity contributed to the institutional capacities of key justice sector institutions to address judicial corruption?
Sustainability	How likely is it that the activity's local government training and support program is able to be sustainably implemented by the national government one to three years after the end of the program? *Focusing scope through "local government training and support program," defining sustainability in terms of the national government implementing the program "one to three years after the end of the program." Note that for reasons of feasibility, this question is not asking for comparison to another donor's program.* What steps is the activity currently taking to build sustainability into its intervention and to what extent do these steps address existing constraints to sustainability?

TOPIC	ILLUSTRATIVE QUESTION
Theory of change and assumptions	To what extent did the mismatch between elements of the theory of change and the distribution of project human/financial resources affect implementation? *Focusing scope of factors affecting implementation to "the distribution of project human/financial resources," and enhancing feasibility by pointing to existing project data instead of e.g. a population survey.* Did the program have any unintended consequences on women's political participation? In what ways did changes in the context of political party competition and independent media capacity assumed by the theory of change generate obstacles or opportunities for the main activities under evaluation?
Inclusivity/ gender and targeting	To what extent were women and youth included in community outreach about participatory budgeting? *Focusing scope on "women and youth" and "community outreach about participatory budgeting," and enhancing feasibility by pointing to information available from project MEL data and interviews of project staff.* To what extent and how did USAID activities foster the participation of target groups in civic education opportunities? To what extent did the project's prioritization of marginalized groups align with the expectations of key stakeholders?
Implementation and adaptation	To what extent and how did the Mission use research and analysis, including previous evaluation findings, to make timely and effective programmatic changes to achieve its goals? To what extent is MEL data providing actionable information that is used to inform program adaptations? *Focusing scope on "program adaptations" and "MEL data," clarifying "actionable" elsewhere, and enhancing feasibility by pointing to information available from project MEL data and interviews of project staff.* What lessons should USAID and its partners draw from efforts to increase political will for anti-corruption reform? To what extent have scale-up efforts incorporated lessons learned from the initial pilot? To what extent have lessons learned been identified, shared and incorporated across sites and sub-awardees?

TOPIC	ILLUSTRATIVE QUESTION
Partnerships and stakeholder engagement	To what extent, if any, has the activity's existing partnership structure with the Ministry of Community Development and the Ministry of Finance contributed to the success or failure of the program to enable the sub-national governance system to expand revenue?
	To what extent are Objective 2 activities engaging with the most relevant institutional stakeholders, including teachers, school boards, and government officials, at the provincial and national levels?
	Focusing scope on "Objective 2 activities" and "most relevant institutional stakeholders [...] at the provincial and national levels, clarifying definitions of "engaging" and "most relevant" elsewhere, and enhancing feasibility by pointing to population set of "teachers, school boards, and government officials."