



Evaluation Utilization at USAID



FEBRUARY 23, 2016

This publication was produced at the request of the U.S. Agency for International Development (USAID). It was prepared independently by Molly Hageboeck, Micah Frumkin, Jenna L. Heavenrich, Lala Kasimova, Management Systems International (MSI), a Tetra Tech Company, with Melvin M. Mark, Pennsylvania State University, and Dr. Aníbal Pérez-Liñán, University of Pittsburgh, under the Feed the Future Knowledge-Driven Agricultural Development (KDAD) project, implemented by Insight Systems Corp.

[This page is left intentionally blank.]

Evaluation Utilization at USAID



200 12th Street, South
Arlington, VA 22202 USA

Tel: + 1 703 979 7100

Sub-contracted under AID-OAA-C-13-00137

DISCLAIMER

This document was produced for review with the support of the U.S. Agency for International Development (USAID) under the Feed the Future Knowledge-Driven Agricultural Development (KDAD) project, implemented by Insight Systems Corp. and Management Systems International Inc., a Tetra Tech Company. The views expressed are those of the author and do not represent the views of the United States Agency for International Development or the United States Government.

ACKNOWLEDGEMENTS

A study of this magnitude could not have been successfully completed without the hard work and dedication of a great many people. The team would like to thank USAID/PPL/LER, specifically Tania Alfonso and Jerome Gallagher for their continuous support and guidance throughout the process. Outside of USAID, Jennifer Dahnke at the Feed the Future Knowledge-Driven Agriculture Development Project (KDAD) provided valuable substantive and procedural support. In addition, USAID's Development Experience Clearinghouse (DEC) staff provided invaluable assistance in locating evaluations and deserves credit for improvements in this collection over the past few years.

MSI would also like to acknowledge our expert and incredibly hard working team of researchers, without whom we never would have been able to achieve the ambitious task set before us. Among the MSI team we include evaluation and meta-analysis specialist, Micah Frumkin, quantitative analyst, Jenna Lindeke Heavenrich, and project manager, Lala Kasimova. Rajan Kapoor, Jasmine Yu, Jessica Ngo, Katherine Wikrent, Bethlehem Haile, and Amanda Kitanga provided additional support. In addition to these individuals, the team for this study is grateful for the guidance and support of Dr. Melvin M. Mark, Pennsylvania State University, the academic advisor for the study and Dr. Aníbal Pérez-Liñán, University of Pittsburgh, who conducted the study's logistical regressions and qualitative comparative analyses.

Lastly, and most importantly, we would like to thank the hundreds of USAID staff from across the globe and representatives from a dozen USAID implementing partners who took time out of their busy schedules to provide invaluable comments and perspectives which helped us to fully understand the ways in which USAID evaluations are used, as well as the end results of that use.

Molly Hageboeck
Team Leader/MSI Technical Director

CONTENTS

- Acronyms.....iv**
- Executive Summary.....vi**
- Introduction I**
- Study Findings.....8**
 - Question 1: How and when in the Program Cycle are evaluations used or not? 8
 - Question 2: What changes/decisions are made because of evaluations? 14
 - Question 3: To what degree and under what conditions does learning occur from evaluation findings that was not anticipated by the intended purpose of the evaluation? 18
 - Question 4: What particular business processes or enabling conditions appear to encourage or discourage the utilization of evaluations? 23
- Conclusions38**
- Recommendations40**
- Annex 1 – Evaluation Utilization Study Statement of Work.....46**
- Annex 2 – Study Methodology50**
- Annex 3 – List of Interviewees71**
- Annex 4 – Data Annexes73**
 - 4.1 – Survey Results73**
 - 4.2 – Evaluation Utilization Study Sub-studies.....110**
 - PPR Evaluation Registry Use Statements111
 - Country Development Cooperation Strategies.....116
 - USAID Agency-Wide Policies, Strategies, Frameworks and Visions.....122
 - USAID Evidence Summits124
 - Evaluation Purposes & Questions.....127
 - Evaluation Recommendations.....130
 - Standard Mission Orders on Evaluation.....134
 - Evaluation Utilization Monitoring Systems139
 - 4.3 – Logistical Regression Analysis.....143**
- Annex 5 – Study Instruments175**
 - 5.1 – Policy Document Review Template.....176**
 - 5.2 – CDCS Review Template.....177**
 - 5.3 – PPR Entry Review Template180**
 - 5.4 – Evaluation Mission Order Review Template181**
 - 5.5 – Evaluation Purposes, Questions, and Recommendations Review Template183**
 - 5.6 – Interview Guide for Operating Unit Group Discussions.....185**
 - 5.7 – Evaluation Report Quality Review Checklist, Rater’s Guide and Overall Evaluation Quality Review Score188**
- Annex 6 – Bibliography.....199**
- Annex 7 – Stories of Evaluation Use Results203**

Table of Tables

Table 1. Utilization Literature-based Factors that May Affect Evaluation Utilization	3
Table 2. Summary Statistics for the Survey of Completed Evaluations	6
Table 3. Distribution of Evaluation Survey Returns by Type, Timing, Region, and Sector	6
Table 4. Post-Evaluation Actions According to Survey and Interview Responses.....	26
Table 5. Evaluation Factors for Which a Linkage to Evaluation Use Was/Was Not Found	39

Table of Figures

Figure 1. Timeline of Recent Evaluation Policy-related Events	2
Figure 2. Documents Coded and Verified as Evaluations in the DEC, by Year.....	5
Figure 3. USAID Program Cycle	8
Figure 4. Evaluation Uses in the Project & Activity Design and Implementation (N=118).....	9
Figure 5. Number of CDCSs by Region and Percentage that Cited USAID Evaluations	11
Figure 6. Perceptions of Effects of Taking Action on Evaluation Results (N=118).....	16
Figure 7. Time Period over which Learning from an Evaluation Occurs (N=118).....	18
Figure 8. How Staff Learned about Projects/Activities from USAID Evaluations (N=118).....	21
Figure 12. USAID Involvement of Partners in Evaluation Processes (N=118)	32
Figure 13. Presence of Factors that may Foster or Impede Evaluation Utilization in USAID (N=118 USAID Evaluations for which Survey Responses Exist).....	37

ACRONYMS

ADS	USAID's Automated Directives System
Af/Pak	Afghanistan and Pakistan
AWDP	Afghanistan Workforce Development Program
CDCS	Country Development Cooperation Strategy
DEC	Development Experience Clearinghouse
DO	Development Objective
E3	USAID Bureau for Economic Growth, Education and Environment
E&E	Europe & Eurasia
GAO	Government Accountability Office
HVC	Highly Vulnerable Children
IDIQ	Indefinite Delivery, Indefinite Quantity Contracts
IP	Implementing Partner
IR	Intermediate Result
LAC	Latin America and the Caribbean
LER	USAID Office of Learning, Evaluation and Research
M&E	Monitoring and Evaluation
MO	Mission Order
MSI	Management Systems International
NGO	Nongovernmental organization
OECD/DAC	Organization for Economic Cooperation and Development/Development Assistance Committee
OMB	Office of Management and Budget
OU	Operating Unit
PAD	Project Appraisal Document
PMP	Performance Management Plan
POC	Point of Contact
PPL	USAID Bureau for Policy, Planning, and Learning
PPR	USAID's Annual Performance Plan and Report
RDCS	Regional Development Cooperation Strategy
RDMA	USAID Regional Development Mission for Asia
SOW	Statement of Work
SPP	USAID Office of Strategic and Program Planning
USAID	United States Agency for International Development
USG	U.S. Government

[This page is left intentionally blank.]

EXECUTIVE SUMMARY

Introduction

Study Purpose, Audience, and Intended Use

This study was undertaken to help the U.S. Agency for International Development (USAID) determine the extent to which its evaluations are being used and what guidance, tools, or Agency practices might be improved to enhance evaluation utilization. It was commissioned by USAID's Bureau for Policy, Planning, and Learning's Office of Learning, Evaluation, and Research (PPL/LER). Internal and external audiences for this study include USAID management, program, technical, and regional staff whose work can be informed by evaluations, as well the Department of State and other U.S. Government (USG) colleagues on country teams; Congress; the Office of Management and Budget (OMB); partner country governments; and other donors with whom USAID collaborates overseas.

Study Questions

Four study questions guided the work of the study team:

1. How and when in the Program Cycle are evaluations used or not used?
2. What changes/decisions are made because of evaluations?
3. To what degree and under what conditions does learning occur from evaluation findings that were not anticipated by the intended purpose of the evaluation?
4. What particular business processes or enabling conditions appear to encourage or discourage the utilization of evaluations?

Study Methods

The design for this study included a thorough document review, a survey of USAID staff knowledgeable about specific evaluations, and group interviews with USAID staff in 24 Agency Operating Units (OUs) in Washington and overseas, in which 250 individuals participated. Interviews were also conducted with key informants in USAID and nine partner organizations that conduct evaluation for USAID directly or under USAID-funded activities. A survey that complemented this range of interviews yielded 118 responses on a sample of 206 out of the 609 evaluations in the study universe for 2011-2014, or a 57 percent response rate. To help ensure that survey findings were representative of the study universe, in light of this response rate, survey findings are reported on a weighted basis. This combination of data sources includes both self-reported descriptions of evaluation utilization from Agency staff and, where possible, confirming data from other sources, with the most extensive set of documents of this type being 30 USAID country strategies that cited evaluation findings. Stories about the effect of actions taken based on evaluations on broader development outcomes were also collected, and for a portion of these stories, further research provided confirming data from other sources. All of these data sources provided descriptive information on evaluation utilization that was used to answer Questions 1-3. These data were then analyzed further, under Question 4, to explore relationships between evaluation characteristics, including characteristics of USAID processes, and evaluation utilization outcomes, including learning, actions taken by USAID and partner staff and, to some extent, the effects of such actions on broader development outcomes.

Study Limitations

Limitations of this study include the self-reported nature of survey and interview data, a survey response rate below the ideal level, and the extraction of information about utilization from documents that were not designed for that purpose.

Study Findings

QUESTION 1: HOW AND WHEN IN THE PROGRAM CYCLE ARE EVALUATIONS USED OR NOT?

Study findings showed that for 93 percent of a sample of USAID commissioned evaluations, survey respondents indicated that the specific evaluations with which they were familiar had stimulated learning in USAID and, to a degree, among its partners as well. Ninety percent of these evaluations were reported to have resulted in decisions being made and actions being taken at appropriate stages in USAID's Program Cycle. Group interviews with USAID OU staff yielded qualitative findings along these same lines.

Project Design and Implementation Study data from interview and survey responses indicated that the Project Design and Implementation stage of the Program Cycle is where evaluation utilization occurs most frequently. Findings showed that 71 percent of the evaluations on which survey responses reported had been used to design or modify a USAID activity or project, and 47 percent were used exclusively for project or activity design purposes, while 11 percent were used exclusively for project or activity modification, and 13 percent were used for both of these purposes. Other study data sources confirmed survey findings.

Strategy and Policy Formulation In the USAID Program Cycle, an important step is the development of a multiyear Country Development Cooperation Strategy (CDCS). The study team's review of 51 approved CDCSs showed that 30 (59%) cited completed USAID evaluations. Eighty-two percent of the evaluations on which these CDCSs drew were evaluations undertaken at the project or activity level, while 41 (7% of the universe of 609) were undertaken at the sector level or for multiple activities or projects in a single Mission.

USAID evaluations rarely focus on the policy level, and Agency requirements for drawing on evaluation evidence do not apply to policy formulation. At this level, USAID evidence-based policy papers drawn on syntheses of state-of-the-art research and Agency experience consolidated through evidence and experience summits it organizes. Not surprisingly, the study team did not find either relevant evaluations at the policy level or their active use in 19 topical policy papers it reviewed.

QUESTION 2: WHAT CHANGES/DECISIONS ARE MADE BECAUSE OF EVALUATIONS?

Reporting on evaluation utilization, survey respondents indicated that evaluation results were used for new project and activity design at least twice as often as they were used to modify existing activities, projects, and strategies, while a somewhat higher level of evaluation use to refocus existing projects and activities was documented in Annual Performance Plan and Reports (PPRs). Other common changes include revising delivery mechanism work plans, extending activity timelines, or expanding activity geographic areas. Survey responses also showed that USAID staff noted that their implementing partners' actions complemented USAID post-evaluation decisions and changes in connection with 19 percent of the sampled evaluations. Eight percent of 118 surveyed evaluations indicated that country government strategies and/or policies had changed to reflect learning from the evaluation in question.

When survey respondents reported that action was taken based on evaluation findings and recommendations, they also, in some instances, reported on the results of this action. For 27 percent of the evaluations on which surveys reported, respondents indicated that the effectiveness of the programs, projects, and activities increased. Survey results also showed that evaluation findings informed the development of new strategies for 20 percent of the 118 evaluations on which the survey reported.

In addition, action taken as a result of evaluation findings reportedly enhanced sustainability for 20 percent of this set of evaluations, and improved cost-effectiveness for programs associated with 7 percent of these evaluations. In the course of this study, the team also collected 58 stories about the effects of evaluation utilization. For 10 of these stories which the studies followed up on through online research, Evaluation Utilization Briefers are presented under Question 2 in the report and in Annex 7.

QUESTION 3: TO WHAT DEGREE AND UNDER WHAT CONDITIONS DOES LEARNING OCCUR FROM EVALUATION FINDINGS THAT WAS NOT ANTICIPATED BY THE INTENDED PURPOSE OF THE EVALUATION?

Evidence of learning is difficult to capture directly in the absence of some type of “before” and “after” comparison. Nevertheless, responses to survey questions provided an indication of what USAID staff learns from the evaluations to which they are exposed. Notably, their responses showed that while most learning from evaluations occurs within the first three months after an evaluation is completed, in some cases the learning period has been up to two years. With respect to what survey respondents learned about specific programs, projects, or activities as a result of evaluations, 52 percent indicated that these evaluations affected their opinions on the merits of the project or activities evaluated, and 54 percent said evaluations provided insights about the soundness or adequacy of the activity design. Still other respondents said that they learned about how to make projects more effective in the future, or why objectives had not been achieved. Study data, including discussions in group interviews, also indicated that much of what USAID staff learned from evaluations confirmed what they felt they already knew (74% of the survey responses). New information that contributed to learning was also reported by respondents to be present in evaluation reports (52% of the evaluations covered by survey responses), which participants in group interviews indicated was generally useful, even when it revealed problems USAID was not aware of.

As the foregoing suggests, most of what USAID staff reported they learned from evaluations did not surprise them. Where the study team did detect unanticipated learning was from answers USAID staff provided to questions about what they had learned about evaluation processes and purposes from being involved in the planning, implementation, or review of the evaluation. Among respondents, 77% indicated that participation in the evaluation had improved their understanding of the purpose and role of evaluation in the USAID Program Cycle; 70% indicated that they had a better understanding of evaluation types and methods, and 76% noted that they felt better prepared to critique evaluation products.

QUESTION 4: WHAT PARTICULAR BUSINESS PROCESSES OR ENABLING CONDITIONS APPEAR TO ENCOURAGE OR DISCOURAGE THE UTILIZATION OF EVALUATIONS?

The study team examined a range of business practices and evaluation characteristics to determine which ones appeared to encourage or discourage the utilization of evaluations. On the business process side, two sets of practices recommended by USAID guidance received considerable attention, namely:

- USAID’s dissemination of its evaluation results. USAID’s dissemination of evaluation results through briefings, report dissemination, and other events was reported to be strong to USAID staff (97% of the survey responses), but weaker in reaching USAID implementing partners (76% of responses reported this type of dissemination) and country partners (43% reportedly disseminated to country partners). Notably, dissemination to country partners was found to be statistically significant in relation evaluation utilization by those partners. The study also found that simply delivering

evaluation reports to USAID’s online library, the Development Experience Clearinghouse (DEC), was not statistically linked to utilization by USAID or its partners.

- Post-evaluation action planning meetings. Fifty-three percent of respondents indicated that post-evaluation action planning meetings were held. Furthermore, 46 percent of these designated who was to take action on specific findings or recommendations, and 33 percent reported that timelines were established for implementing those actions. Proactive post-evaluation action planning meetings were found to be statistically significant in relation to study evidence of evaluation utilization.

Other factors that were found to be statistically associated with evaluation utilization included whether evaluation reports were considered to be timely for decision making, a finding supported by group interviews with USAID staff. The study also found a significant relationship between evaluation utilization and average evaluation report quality scores at the OU level (calculated using a method developed for an earlier evaluation quality study commissioned by PPL/LER, as described in the study methods annex), although evaluation report quality scores for individual evaluation reports, one at a time, were not found to be significant in relation to evaluation use.

Conclusions

Broad conclusions reached by the study team include:

- USAID evaluation utilization practices are already strong and compare well to those of other USG agencies examined in parallel studies conducted by the GAO.
- Broad responses indicating an overall high rate of evaluation use in USAID are notable, but somewhat mask the fact that there has been only partial uptake by OUs of USAID business processes designed to foster evaluation utilization, including both evaluation dissemination and post-evaluation reviews and action planning. Study findings also suggest that USAID staff are more knowledgeable about evaluation results and their effects than USAID’s annual PPR Evaluation Registry statements suggest. Survey responses and stories shared in interviews demonstrate the range of this knowledge, and compare favorably to the 12 percent of PPRs that claim some form of utilization. Staff knowledge about evaluation utilization that is not currently captured in PPRs is potentially valuable for communicating USAID’s achievements.
- Important opportunities for improving evaluation utilization exist where recommended practices are not yet fully implemented (see Figure 13 in the body of the report), and where findings show a statistically significant relationship between evaluation utilization and specific practices or evaluation characteristics.
- With respect to taking actions to improve evaluation utilization, the study team found that USAID OUs that commission evaluations are the prime intended users of USAID evaluations and are thus best positioned to make needed changes to enhance evaluation use. Technical and Regional Bureaus in the Agency, along with PPL/LER, have important roles in supporting improvement initiatives.

Recommendations

The following recommendations are organized by, and addressed to, specific USAID units the study team feels are best suited to take action. By taking the recommended actions, USAID and its partners can enhance evaluation utilization to improve program performance and positive development outcomes.

It is recommended that USAID Operating Units (OUs) (Field Missions and Washington Offices that directly administer the delivery of foreign assistance through projects, activities, or other mechanisms):

- I. Conduct a brief, collaborative, OU-level internal assessment, next quarter, of the current level of evaluation use within the OU.

2. Create and obtain OU management approval for an OU-specific plan for enhancing the utilization of USAID evaluations, including by OU partners and staff.

It is recommended that USAID Technical Bureaus:

3. Inventory what compendiums, syntheses, or meta-analyses already exist for USAID evaluations in technical fields that Bureaus support. Ensure that existing analyses of evaluation results are available to Missions, Regional Bureaus and to other USAID staff working in relevant technical areas.
4. Collaborate with PPL/LER to develop more comprehensive and improved approaches for ensuring that evidence from future USAID evaluations are widely available to Agency staff. Technical Bureaus should work with Regional Bureaus to ensure that evaluation evidence products accurately reflect geographic considerations.

It is recommended that USAID Regional Bureaus:

5. Review with their Missions, and other Washington-based Offices delivering foreign assistance in their regions, the adequacy of their M&E staffing and budget resources relative their needs and evaluation commitments made in CDCSs, Project Appraisal Documents (PADs), and PPRs. Such reviews should ensure that no Mission or Office is unduly disadvantaged in its ability to access and utilize evaluation results based on its size, geographic location, or the nature of its portfolio.
6. Further develop, implement, and share information Agency-wide about regionally appropriate mechanisms for collaboratively drawing on existing M&E resources. Arrangements should built on productive experiences such as the peer exchanges—which help Missions in need expand their capacity to take on a short-term evaluation tasks—or the Regional Mission evaluation hub model that is emerging in the Asia Bureau, through which evaluation services are provided to smaller Missions and Country Offices.

It is recommended that PPL/LER:

7. Minimally refine official Agency guidance, in line with findings from this study, to draw attention to and enhance evaluation utilization. Exhibit I includes a list of illustrative modifications to this end.
8. Continue to develop “How-To” guides and “Technical Notes,” or expand its Evaluation Toolkit, in response to expressed needs. Consider USAID staff suggestions with respect to the development of new “How-To” guidance on (a) writing good evaluation questions, (b) dissemination planning, and (c) well-constructed evaluation recommendations. Incorporate into this last topic a recommended range for number of evaluation recommendations, akin to the Agency’s recommendation on the number of evaluation questions.
9. Examine very short evaluation summaries or briefers used by other development agencies that USAID might adopt over time and reduce its reliance on complete evaluation reports as the primary means of disseminating information about findings and evidence. Such products include, for example, standalone evaluation briefs (1-2 pages) or abstracts (300-500 words), which more succinctly convey evaluation findings and evidence for future programming.
10. Improve USAID’s ongoing monitoring of evaluation utilization through existing reporting mechanisms if possible, or create a new reporting mechanism if existing mechanisms cannot be improved. More specifically, consider ways in which the PPR reporting mechanism could be modified to permit more comprehensive reporting and consider adopting the use of post-evaluation Action Plan tracking systems being piloted in some Missions. Encourage OUs to quantify their performance in this regard by calculating the percent fully implemented across all evaluations as an OU performance measure, and after piloting this indicator, test the value of its use for reporting at the Agency level.

INTRODUCTION

Study Purpose, Audience, and Intended Use

The U.S. Agency for International Development's (USAID) Bureau for Policy, Planning, and Learning's Office of Learning, Evaluation, and Research (PPL/LER) commissioned this study to help USAID determine the extent to which its evaluations are being used and what guidance, tools, or Agency practices might be improved to enhance evaluation utilization. Evidence from this study is expected to help Agency leadership and staff foster and support evaluation utilization. The study's internal audience includes USAID evaluation points of contact (POCs) and staff responsible for monitoring and evaluation (M&E) functions within strategy teams, in the United States and overseas, as well as program, technical, and regional office staff whose work can be informed by evaluations. Beyond USAID, key stakeholders for this study include the Department of State and other U.S. Government (USG) colleagues on country teams; Congress; the Office of Management and Budget (OMB); partner country governments; and other donors with whom USAID collaborates overseas. A Statement of Work (SOW) for this study is available in Annex I.

Defining Evaluation Success

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

—USAID Evaluation Policy

Study Questions

Four study questions guided the work of the study team:

1. How and when in the Program Cycle are evaluations used or not used?
2. What changes/decisions are made because of evaluations?
3. To what degree and under what conditions does learning occur from evaluation findings that were not anticipated by the intended purpose of the evaluation?
4. What particular business processes or enabling conditions appear to encourage or discourage the utilization of evaluations?



FTS Evaluation Techniques Presentation at USAID 2014

The USAID Program Cycle, referenced in Question 1, is the Agency's conceptual framework for strengthening its policy, planning, and evaluation capacities in line with priorities established in the 2010 Presidential Directive on Global Development and in USAID Forward, the Agency's internal reform agenda. The Program Cycle establishes a common set of practices through which these reforms are being implemented. It encompasses Agency policy formulation and the development of multi-year Country Development Cooperation Strategies (CDCSs) under which projects and the activities are implemented. Monitoring and evaluation are integrated into all stages of this cycle. These Program Cycle elements are, in turn, linked to and supported by learning and adaptive management practices, and strategic and evidence-based budgetary resources.

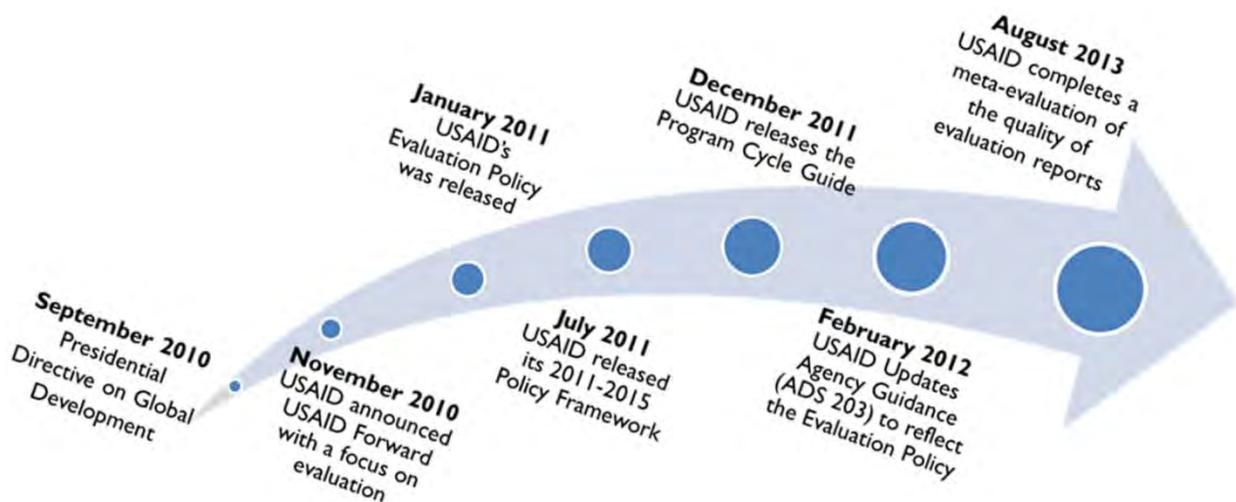


Figure 1. Timeline of Recent Evaluation Policy-related Events

Study Background

USAID's 2011 Evaluation Policy builds on the Agency's long history of conducting evaluations to support its program management processes and challenges its staff and partners to enhance the quality and utilization of the evaluations they undertake. The Evaluation Policy was introduced after a period of decline in the number of USAID evaluations and external studies had raised questions about the quality of evidence presented in evaluations undertaken in the international development.¹ USAID's Evaluation Policy addressed these challenges. It raised expectations concerning both the range and quality of evaluations USAID intended to undertake, and it stimulated the development of supporting guidance materials and training programs to reinforce this policy, as Figure 1 illustrates. Consistent with its policy and oversight role in the Agency's evaluation process, PPL/LER provides support and quality assurance for evaluation design, quality, dissemination, and synthesis of evaluation system products. It also occasionally organizes external technical audits of the quality and utilization of evaluation results. PPL/LER's 2013 meta-evaluation of evaluation report quality (2009-2012) responded to this mandate, as does this study of evaluation utilization.²

Study Methods and Limitations

This section describes the study design, methods, limitations, and characteristics of study populations.

¹ Figure 2 documents the decline in USAID evaluations; external challenges focused on the quality of evidence gathered in USAID and other donor evaluations included the Center for Global Development's 2006 report entitled: *When Will We Ever Learn?* (2006), and a second report that examined the quality of a sample of USAID evaluations in the democracy/governance portfolio (Bollen, Kenneth, et. al. *Assessing International Evaluations: An Example From USAID's Democracy and Governance Program* (2005).

² The meta-evaluation that PPL/LER commissioned, entitled *Meta-Evaluation of Quality and Coverage of USAID Evaluations* (2009-2012), was conducted by Management Systems International (MSI).

Study Design and Methods

This study was structured around the four questions provided in USAID’s SOW. Because the first three questions similarly sought descriptive information on evaluation uses, the design and methods used to answer them were relatively similar. The design focused on a thorough document review, a survey of USAID staff knowledgeable about particular evaluations, and group interviews with USAID staff and partners. Question 4 sought information about the linkages between USAID businesses processes and conditions, and the resulting utilization outcomes. This different type of question required a different approach. As explained further below and in Annex 2, the study team’s approach to Question 4 began with a review of the literature on evaluation utilization. This review identified factors that other studies, including U.S. Government Accountability Office (GAO) reviews, indicated either encourage or discourage evaluation use.³ The factors identified through this process are summarized in Table 1.

Table 1. Utilization Literature-based Factors that May Affect Evaluation Utilization

Factors Associated in Literature with Evaluation Processes & Utilization	
Need for, or purpose of, the evaluation	Sponsor organization’s quality control practices
Country or regional context	Relevance to client; client ability to act on results
Commissioning organization evaluation culture	Timeliness (in time for decisions; on schedule)
Credibility and relevant knowledge of evaluators	Presence of new information
Participation of stakeholders in the evaluation	Recommendations (specific, actionable, practical)
Quality (methods, data, analysis)	Post-evaluation dissemination
Quality (overall credibility for learning & action)	Post-evaluation review, action planning, & tracking

Data collection and analysis for all four study questions used a mix of methods, as summarized in the text box below and described in greater detail in Annex 2.

Both the survey and group interviews involved self-reporting by USAID staff. In this sense, the study methods somewhat resembles the GAO’s model for evaluation utilization reviews of U.S. federal agencies. This study, however, went beyond the GAO norm by drawing on other data sources to confirm or contest self-reported answers, and by comparing the answers obtained from the survey to other forms of self-reporting by Agency staff, i.e., group interviews and PPRs. Findings from PPRs used to make these comparisons are described in Annex 4.2: Sub-Study 1.

Data sources that supplemented self-reported data, and facilitated team efforts to triangulate data from several sources when responding to the study questions, were generally linked to the Program Cycle. They included (a) USAID country strategies, which the study team examined extensively and reports on in Annex 4.2, Sub-Study 2; (b) Project Appraisal Documents, which USAID itself is examining for this same purpose; (c) recent policy papers and Evidence Summit records, all of which the study team examined, as reported in Annex 4.2: Sub-Study 3; and (d) solicitations (RFPs and RFAs) that indicate evaluation use, which are difficult to locate, even though they are in the public domain. The evidence chain for utilization in one such solicitation is described in this study.

³GAO. “Program Evaluation: Strategies to Facilitate Agencies’ Use of Evaluation in Program Management and Policy Making.” United States Government Accountability Office, Report to Congressional Committees, GAO-13-570 (June 2013). Accessed June 9, 2015. <http://www.gao.gov/products/GAO-13-570>.

Summary of Data Collection and Analysis Methods

1. **Group interviews** with USAID Operating Units (OUs) were used to address **all study questions**; a list of interviewees is provided in Annex 3.
 - OUs were selected considering the number of evaluations they undertook, size of their portfolio, and geographic distribution. The 24 OUs interviewed represent all **six regional** and **four Technical Bureaus** in Washington, **12 Country Missions** and **two Regional Missions** based on size, numbers of evaluations, and geographic distribution.
 - **Two** interviews were held per selected OU, with mid- and senior-level staff. Additionally **two** individual key informant interviews were conducted with PPL staff. Interviews were also conducted with staff from six firms and three nongovernmental organizations (NGOs) that conduct evaluations of USAID activities. In total, more than 250 individuals participated in study interviews.
2. A **survey** of a sample from the study universe of 609 evaluations completed between 2011 and 2014 was used gather information on **all study questions**.
 - The sampling procedure involved categorizing evaluations by scope and timing and then sampling them within each cluster based on **random sampling principles**. For the largest cluster along the scope dimension, which consisted of 503 single country/single project or activity evaluations, a simple random sample was drawn after a discussion of options with USAID; this discussion also considered stratifying evaluations in this cluster by region and sector, but stratification was not selected by USAID which favored the clarity and simplicity of a simple random sample for this large cluster. For smaller clusters by scope, ranging from 10 to 43 evaluations, a quota of 10 evaluations was established, and specific evaluations up to the quota were randomly sampled; for the smallest clusters, i.e., nine or fewer, a census of all the evaluations was included in the sample used to solicit responses to the study survey.
 - Out of **206 surveys sent, 118 surveys were returned** yielding a **57% response rate**.
 - To ensure that reporting for the 118 completed survey responses accurately represents tendencies among the evaluation that made up the study universe, given this relatively low response rate and the multiple sampling techniques used to create the survey sample, the study's survey findings are reported as weighted percentages in the body of this report. Annex 4.1 provides a full set of responses on both a weighted and unweighted basis.
3. **Desk Reviews** covering USAID CDCSs, PPRs, Mission Orders, Policy Papers, and Evidence Summits were also conducted; all sub-study reports can be found in Annex 4.2. In addition, stories about the effects of actions taken based on evaluations were collected from a range of sources, as discussed in Annex 7, which includes eight Evaluation Utilization Briefers; another two can be found under Question 2 in the body of the report.
4. **Data Analysis** used qualitative techniques for **open-ended questions** (content and pattern analyses) and quantitative techniques for **close-ended questions** (including frequency distributions, averages, ranges, chi squares, and logistical regressions); Annex 4.3 provides a detailed description of the logistical regression analysis.

Other data sources included USAID evaluation Mission Orders, some of which include guidance that goes beyond USAID's policy and Agency-wide Automated Directives System (ADS) guidance, as described in Annex 4.2, Sub-Study 7, and data on evaluation purpose statements, questions, and recommendations, drawn from evaluation report quality rating exercises for 45 evaluations from the study universe of 609 and analyzed before the survey was underway, as described in Annex 4.2, Sub-Studies 5&6. In addition, the study team collected USAID's own compilation of stories about the effects of evaluation utilization on broader development outcomes. To validate a portion of these stories the team also sought information from non-USAID sources through online searches, as illustrated in the 10 Evaluation Utilization Briefers in this volume under Question 2 in the report and in Annex 7. In general,

these various efforts confirmed the self-reported evidence of USAID staff. A final set of sources focused on how other donor agencies monitor evaluation utilization. These sources and findings are summarized in Annex 4.2, Sub-Study 8.

Data collection instruments used in this study are provided in Annex 5. The logistical regression that supports the response to Question 4 is included in Annex 4.3, as noted in the text box above, and the study's bibliography is in Annex 6.

USAID Evaluations in the Study Sample

Since 1993, USAID, MSI, and others in the evaluation community have monitored the number of evaluations USAID has funded on an annual basis. All documents entered into USAID's Development Experience Clearinghouse (DEC) are periodically examined to determine which of are actual evaluations, as opposed to being audit reports or other kinds of analyses. Figure 2 displays the count of evaluations in the DEC for each year through 2014, as of July 2015.⁴

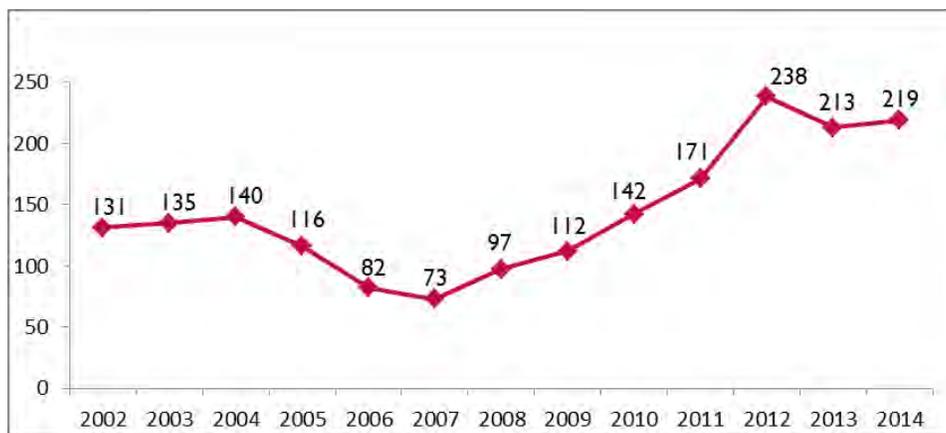


Figure 2. Documents Coded and Verified as Evaluations in the DEC, by Year

To construct the utilization study universe, the study team used a list provided by the DEC in February 2015 that contained 758 evaluations (83 fewer than were in the July 2015 list used to create Figure 2). From these 758, the team removed duplicates, reports in languages other than English, and partner self-evaluations funded by USAID, which resulted in a study universe of 609 evaluations carried out by teams led by external team leaders. Among these 609 evaluations, 503 focused on single projects or activities in a single country while the remaining 106 represented several other focus clusters. Table 2 shows how evaluations were selected for the study survey using a simple random sample for the large cluster and a selection process that involved either a census of all evaluations for very small clusters or a quota sample, with randomized selection, for the other clusters. Procedures ensured that all of the ex-post and impact evaluations were included in the set to which surveys were sent, as Annex 2 explains. The survey return rate was 57 percent with 118 survey responses received. To better understand this response rate, MSI explored non-response patterns in relation to such possibly explanatory factors as year, sector, region, and evaluation type and scope. In all instances, non-response patterns were reasonably similar to the sample distributions along these variables, meaning that there were no unusual non-response patterns detected to explain why surveys linked to specific evaluations were or were not returned. The results of the team's non-response analysis are included in Annex 2.⁵

⁴ Annex 2 includes a more detailed version of this graph.

⁵ There is one notable exception to these conclusions. Early in the survey process the study team received a communication from one USAID Mission, which it shared with PPL/LER, indicating that the Mission in question

Table 3 shows how these 118 evaluations were distributed by type of USAID evaluation, performance or impact, evaluation timing (mid-term, final, or ex-post), region, and sector. Of the 118 evaluations for which the study team received survey data, 97 percent were performance evaluations with the remaining being impact or ex-post evaluations. In describing why these evaluations were undertaken, 44 percent were cited as being required while another 44 percent were described as being elective. Among the elective evaluations, 23 percent were undertaken to support forward programming; 10 percent were to support key decisions about ongoing activities; and 8 percent were undertaken in response to questions that came up in performance reviews. As Table 3 shows, the study universe was skewed with respect to some variables, including region and sector, with Africa region and Global Health having the largest share, respectively. During the study's design stage, the team provided USAID with two options for sampling the large cluster of single country/single project or activity evaluations. One was a simple random sample of the universe of 609 evaluations and the other was a stratified sampling option that would have involved samples from 30 cells of a matrix that divided the universe by region and sector. USAID selected the first of these options. Thus chance largely determined how many evaluations represented each region and sector at the point where surveys were sent out, and non-response rates further modified the numbers of evaluations on which data were obtained.

Table 2. Summary Statistics for the Survey of Completed Evaluations

	Universe 2011-2014	Sample Surveys Sent	Response Surveys Returned	Sampling Method
Single Activity, Single Country	503	155	94	Simple Random Sample
Other Smaller Clusters	106	51	24	Quota, Randomly Selected
Total	609	206	118	

Table 3. Distribution of Evaluation Survey Returns by Type, Timing, Region, and Sector

Evaluation Type and Timing		Region		Sector	
Performance (Mid-Term)	55	Africa	37	Economic Growth	35
		Europe and Eurasia	23	Democracy/Governance	32
Performance (Final)	52	Asia	21	Global Health	29
Performance (Ex-Post)	8	Latin America/Caribbean	14	Education	12
Impact	3	Af/Pak	12	Agriculture	9
		Middle East	6	Multi-Sector	1
		Global	5		

As indicated in the summary table on study methods, survey response findings are presented on a weighted basis in the body of this report, while Annex 4 presents responses on each survey question on both a weighted and non-weighted basis. The purpose of weighting is to ensure that reported results are as representative as possible of the underlying population, or study universe of 609 evaluations, given the sampling procedures used and the relatively low overall response rate of 57 percent. The effect of weighting was tested on roughly half a dozen survey questions that USAID identified before it was applied to the full survey. That test did not identify differences between weighted and unweighted responses in percentage terms. Nevertheless, USAID requested that weighted percentages be reported in this study, and that is what was done. As noted elsewhere, readers will find both weighted and unweighted percentages presented in Annex 4.1.

declined to participate in the utilization study. As this situation only involved two survey responses, it did not stand out in MSI's non-response pattern analysis.

Evaluation Focus on USAID Projects and Activities

In the Program Cycle, USAID makes a clear distinction between projects, which are designed and managed by USAID's staff, and supporting activities, funded through mechanisms such as contracts and grants. In the evaluation reports that this study examined, USAID's distinction between projects and activities was not always honored. In practice, 83 percent of the evaluations examined focused on a single effort implemented by one partner, though whether these were called projects or activities varied. Therefore, this study refers to evaluations using the phrase "project or activity."

Survey Respondent Characteristics

Of the USAID staff that provided survey data on the 118 evaluations, 86 percent were based overseas in 46 country and Regional Missions, while 14 percent worked in four Technical Bureaus in USAID's Washington headquarters. The majority (73%) have worked for USAID for more than five years. Just under one-quarter (23%) have worked for the Agency between two and five years, and 4 percent have been with USAID for a year or less. All of these individuals were preselected with the assistance of USAID M&E POCs because of their awareness of specific evaluations selected, randomly or as part of a census, for examination.

Operating Unit Group Interviews

Over 250 individuals participated in group interviews conducted with 24 USAID OUs overseas and in Washington. Interviews were carried out in split sessions; one involved mid-level staff managing USAID assistance projects and activities, and the other involved the OU's management team. All Regional and Technical Bureaus in Washington were represented as were two Regional Missions and 12 Bilateral Missions. The process for selecting Missions is further explained in Annex 2.

Study Limitations

Limitations of this study include the self-reported nature of survey and interview data, a survey response rate below the ideal level, and the extraction of information about utilization from documents that were not designed for that purpose. A more detailed statement of study limitations is provided in Annex 2.

STUDY FINDINGS

This section presents findings for each of the four study questions. Answers to these questions build upon each other. Question 2, on the changes that result from evaluation use, builds on Question 1 descriptions of when use occurs, which is complemented by answers to Question 3 about unanticipated learning from evaluations. Question 4, in turn, looks behind evidence presented for previous questions to identify the processes and factors that seem to precipitate evaluation utilization.

QUESTION 1: HOW AND WHEN IN THE PROGRAM CYCLE ARE EVALUATIONS USED OR NOT?

Study findings show that for 93 percent of a sample of USAID commissioned evaluations, survey respondents perceived these studies as having stimulated learning within the Agency and among its partners.



For 90 percent of these evaluations, respondents reported that it appeared that evaluation findings and recommendations had led to decisions being made and actions being taken at appropriate stages in USAID’s Program Cycle.⁶

Both group interview and survey respondents reported utilization of evaluation results not only after the receipt of a report from an evaluation team but also during the evaluation process.

MSI’s presentation of study findings for Question 1 moves counterclockwise through the Program Cycle graphic in Figure 3, starting with Project Design and Implementation, then moving to CDCS development and finally to Agency Policies.

Figure 3. USAID Program Cycle

Evaluation Use in Project and Activity Design and Implementation

Study data from interview and survey responses indicate that evaluation utilization occurs most frequently in the Project Design and Implementation stage of the Program Cycle. In the survey, 71 percent of the respondents indicated that the evaluations on which they reported had been used to design or modify a USAID activity or project.

Further analysis of the survey responses indicates that 47 percent of the evaluations were reported to have been used exclusively for project or activity design purposes, 11 percent were used exclusively for project or activity modification, and 13 percent were used for both of these purposes, as illustrated in Figure 4 below. Within the cluster of evaluations that were used exclusively to support project or activity design, slightly more than half were used to support new project designs, while the remainder supported the design of follow-on projects or activities.

⁶ USAID’s Program Cycle is described on page 1 and can be found at: http://pdf.usaid.gov/pdf_docs/Pdacs774.pdf

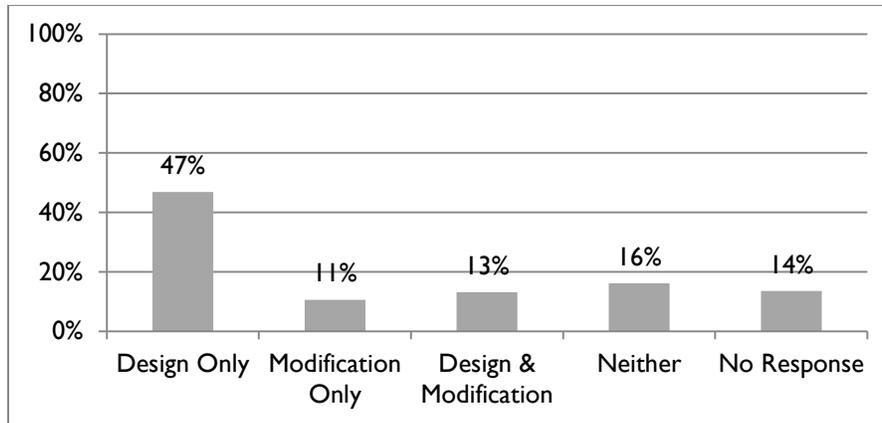


Figure 4. Evaluation Uses in the Project & Activity Design and Implementation (N=118)

“The Project Appraisal Document (PAD) for the Intermediate Result in the Mission's new CDCS, entitled “Political and Electoral Processes are more Competitive, Deliberate, and Transparent,” is a priority PAD for the Mission and has been developed during the FY2013. The Mission has evaluated all three activities ongoing in this area and the results have been used during the PAD development process. At the same time, evaluation results have been used by the AOR for course corrections/focus shifts during the remaining years of implementation of these activities. Results have also been important in light of the changed political environment and the 2014 elections.”

–PPR Evaluation Utilization Entry

Study data further indicate that mid-term evaluations are more likely to be used for project and activity design and modifications than final evaluations. USAID staff, in interviews, suggested this is because final evaluations are not necessarily completed by the time evidence is needed to support decisions. Because findings about evaluation use for project and activity design and modification are largely self-reported by Agency staff, USAID expressed an interest in understanding other ways these reports could be verified. One approach is through PPL/LER's ongoing review of PADs, which are closely held Agency documents due to their procurement sensitivity. Another, which this study team tested on a limited basis, involved searching USAID solicitations for new or follow-on projects or activities. An example of how this follow-up process can validate self-reported information is provided in the text box below.⁷



Photo: DAI. Learning from experience supports the design of a community mobilization program on water in Indonesia.

⁷ While the study team found searches of USAID solicitations to be a feasible method for verifying group interview and survey findings, it was also labor-intensive and thus was not used extensively.

Tracing the Use Evaluations in Activity Design

While reviewing PPR descriptions of evaluation utilization, the study team found a reference to an evaluation of a 2010-2013 legislative strengthening activity in Somalia. The PPR entry claimed the evaluation was used to design a follow-on activity. Included in the referenced evaluation was a finding suggesting that while “USAID’s partners had established good will and strong relationships with Somali counterparts, the lack of constant presence of the [implementing] partners in Somalia was reported to...challenge the[ir] ability to understand the nuances of the [local] partners to the extent possible.” A new procurement was awarded and the nationwide governance project it envisioned is now underway. Tracing forward, MSI located USAID’s May 2014 RFP # SOL-623-14-000015 for a new governance activity in Somalia and found that the solicitation had reframed this evaluation finding as a procurement requirement: “The Contractor must be based in Somalia and must utilize local presence in Mogadishu.” A follow-up email exchange with USAID staff in the region confirmed that this change “is making a huge difference in their operations.”

—Follow up on a USAID PPR Evaluation Utilization Entry

Evaluation Use in Country Development Cooperation Strategies (CDCSs)

In the USAID Program Cycle, an important step is the development of a multiyear Country Development Cooperation Strategy, or CDCS. These strategies identify the main results on which USAID assistance will generally focus for the next five years. In most CDCSs, or their regional equivalent, the RDCS, up to four significant development objectives (DOs) are identified in topical areas, such as health, economic growth, or agriculture, or on a cross-cutting concern, such as youth.

When developing a CDCS, USAID staff are expected to examine and cite existing evidence that support



USAID/Malawi launches its 2013-2018 CDCS with the country’s Minister of Finance and the U.S. Ambassador.

its proposed strategy and action, which will be largely achieved through projects and the activities used to implement them. Such evidence is likely to come from USAID evaluations, published sources, partner government documents, and evaluations conducted by other parties.

Data collected from 45 CDCSs and 6 RDCSs indicate that evaluation utilization in these strategy documents is widespread, but not universal. Comments in group interviews and PPRs also cited evaluation use in CDCSs, with one Mission indicating that while its use of evaluations for its first CDCS had been limited, its second CDCS, which was being

planned at the time of the interview, will benefit from evaluations the Mission has initiated with the development of their new

CDCS in mind.

Summary data from the completed CDCSs reviewed by the study team show that:

- Thirty (59%) of 51 approved strategies the team reviewed cited completed USAID evaluations; notably, the percentage of CDCSs citing evaluations was fairly stable over the four study years.⁸

⁸ A summary of the study team’s review of approved CDCSs and RDCSs is provided in Annex 4.2.

- Seven out of twelve USAID Missions that participated in group interviews described how they had used, or are preparing to use, evaluations as a source of evidence for their CDCSs.
- Seven of one hundred thirty-six PPR statements on evaluation use indicated they had been used to develop CDCSs.
- Twenty-two (20%) survey respondents identified their evaluation as having been used in “revising or developing a strategy for the country, region, or office” in which they worked.

Eighty-two percent of the evaluations on which these CDCSs drew were undertaken at the project or activity level, as the team’s examination of the focus on evaluations in the study universe identified only 41 (7% of the universe of 609) that were undertaken at the sector level or for multiple activities or projects in a single Mission.

“Based on a performance evaluation of a recently completed justice reform project, USAID will continue to work closely to build the capacity of the Superior Council of Magistrates (CSM- the lead body for the independent judicial branch), the new Constitutional Court, and the national Court of Appeal, by helping them to develop and implement procedures and internal court management regulations.” –USAID/DRC CDCS

Regarding CDCS use of evidence from USAID evaluations, the team’s review showed that 47 percent used the information to support sector strategies at the DO level, while 60 percent used the information to support their choice of Intermediate Results (IRs) under the DOs they presented. A smaller fraction (17%) cited USAID evaluations only for the contextual information they provided.

While the overall level of evaluation use in CDCSs was 59 percent, there were differences noted at a regional level. CDCSs from Africa utilized USAID evaluations more frequently (72% of its 18 published CDCSs and RDCSs) than other regions, as Figure 5 shows.

“I am sure you have not read our strategy, our CDCS, because it is not yet approved. What you would see is that a lot of the findings from one of our projects informed how the strategy for our DO team is now different. Yes! Our evaluations actually affect our strategy.” –USAID Staff Member

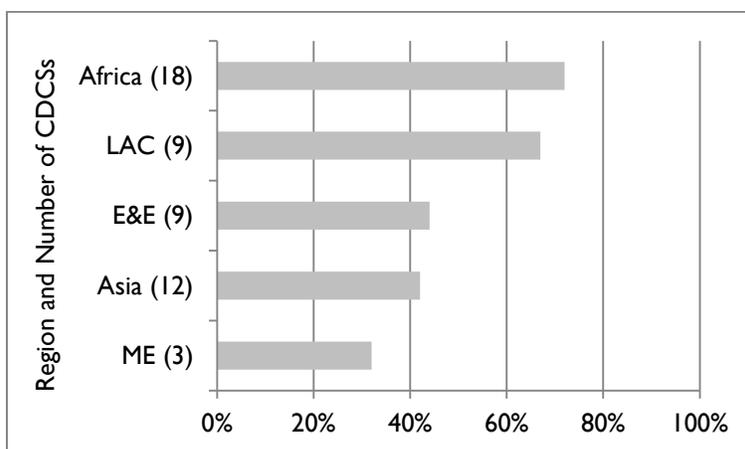


Figure 5. Number of CDCSs by Region and Percentage that Cited USAID Evaluations

The extent of evaluation use in support of strategy development was consistent with the study team’s review of a sample of 33 USAID evaluations which included clear management purpose statements. This

review showed that for 12 (36%) of these 33 evaluations, at least one intended use was to help formulate or refine a strategy.⁹

Interviews with USAID staff and a review of resource lists from CDCSs (where available) made it clear that Missions draw on evidence and experience from a wide array of sources beyond USAID evaluations. Of the 51 CDCSs and RDCSs reviewed:

- Forty-five (89%) referenced other types of USAID documents
- Forty-nine (96%) referenced other non-USAID documents
- Twenty-one (41%) referenced cited published research
- Eight (16%) cited non-USAID evaluations

Evaluation evidence, the study team was told, often confirms the soundness of a Mission's existing strategy, but it can also catalyze a change in strategy or direction, as the text box below illustrates.

Melding Evidence from Several Sources to Inform CDCS Development

"A mid-term project evaluation took place almost two and a half or three years ago, at virtually the same time or overlapping with the time that we were doing our sectoral assessment to inform our CDCS. I think that mid-term evaluation combined with the sectoral assessment helped us make the decision to shift course as it relates to our relationship with parliamentary strengthening. We had quite large, expansive program focused primarily on parliamentary strengthening and on enhancing the relationship between parliament and policy influence on the part of civil society and think tanks. The evaluation helped us focus on that second part. In fact, programmatically we made a very dramatic shift to get out of our explicit parliamentary strengthening focus and for the balance of the program, which is still ongoing, to set up a series of policy clusters. Our inputs now focus on the civil society and think tank influence of policy rather than the actual strength or capacity of staffers in the legislature."

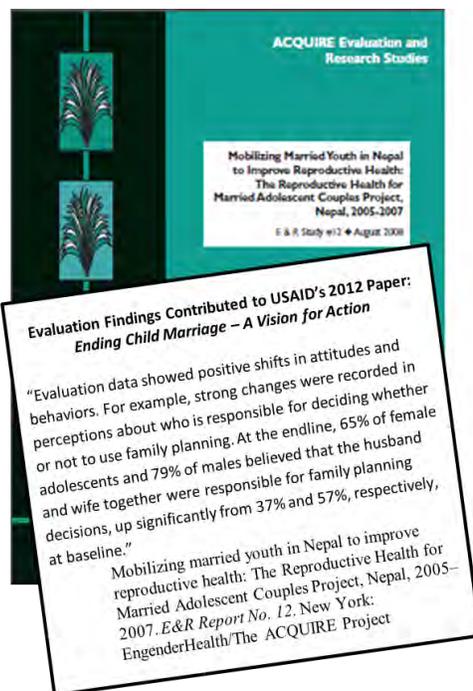
—Excerpted from a group interview with a USAID Mission

Evaluation Use in Policy Formulation

Consistent with the Agency's USAID Forward agenda, interviews with PPL staff highlighted the importance of evidence as the foundation for USAID's policy efforts, and in particular the 2011-2015 Policy Framework and 19 topical policy papers issued since 2010. In support of some of these policy papers and other programming decisions, USAID mounted Evidence or Experience Summits that brought together academics and development practitioners to examine the range of existing evidence and its relevance for development programming.

It should be noted that while USAID requires drawing on its evaluations for project design and CDCS development, this is not required when developing policies.

On the supply side, the study team found very few policy-level evaluations in the study universe. Similarly, on the utilization side, examples were limited. In USAID's Policy Framework, 19 topical policy papers, and nine Evidence Summits, only a few references to USAID evaluations were



⁹ Additional information on this review of evaluation purposes is provided in Annex 4.2, Sub-Study 5.

found, including:¹⁰

- USAID’s 2012 policy paper, *Ending Child Marriage & Meeting the Needs of Married Children: The USAID Vision for Action*, cited evidence from a 2008 USAID evaluation.
- A 2011 evaluation, entitled *DGP Global Program Evaluation (A mid-term, real time, learning-oriented evaluation)*, was listed among the papers for a 2012 USAID Experience Summit on *Strengthening Country Systems*. This summit was later identified in USAID’s 2014 policy paper, *Local Systems: A Framework for Supporting Sustained Development*, as a contributing resource, but no direct line was drawn in the policy paper to the evaluation.
- Materials as part of USAID’s 2012 USAID microenterprise summit, *From Microenterprise to Inclusive Market Development*, included a suggested reading, a paper on the evaluation of USAID’s PROFIT Zambia Project.

Other sources of data, including the study’s survey of completed evaluations and group interviews with USAID Mission and Washington Bureau staff, also provided a limited amount of evidence linking evaluations to policy decisions. In the evaluation survey, 14 percent of respondents indicated, through their responses to a multiple choice question about evaluation use, that the specific evaluation they represented had been “used to revise or develop a USAID policy for this sector, particular problem or topic.”

In the course of interviews with USAID Mission and Bureau staff, the most specific example of evaluation use contributing to policy development came from USAID/Afghanistan (see box below).

Using Evaluation to Improve Existing Policies and Guidance

In 2011, a USAID/Afghanistan evaluation of a provincial roads project reported on the implementing mechanism used, stating that “USAID selected the Cooperative Agreement as the appropriate implementing mechanism...stating that [these agreements] carry certain disadvantages, such as (critically) less accountability. The alternative implementing mechanism—Contract—has a different set of advantages including contractor responsiveness to USAID, high accountability, strong technical skills and a high USAID design input. But as with Cooperative Agreements, Contracts also have disadvantages such as mixed success with capacity building, minimal NGO innovation, greater USAID management and the possibility that higher overhead will equal higher program cost.” Based on these differences and what it learned about the project’s experiences, the evaluation team concluded that “the greater responsiveness found in contracts (typically through detailed reporting requirement) and accountability would have better served the Strategic Provincial Roads Project (SPR).” According to a USAID staff member interviewed during this study, when this evaluation report was completed and circulated in the Mission and in Washington, it garnered a good deal of attention and contributed to deliberations that led to USAID’s 2013 update of its construction project guidelines, which now mandate the use of contracts rather than other mechanisms in USAID-funded construction activities.

—Excerpted from a group interview with a USAID Mission

In summary, MSI’s findings on how and when in the Program Cycle evaluations are or are not used show that while USAID’s Evaluation Policy states that the two purposes of evaluation are learning and accountability, utilization practice suggests that the main way evaluation results are used is to support future actions. The overwhelming majority of evaluations are conducted at the project or activity level. The most common use of findings is to design future project activities, though modifications to existing projects and activities were also cited as an evaluation outcome. In a few cases, project- and activity-

¹⁰ The study team’s review of USAID policy papers is provided in Annex 4.2, followed by its review of evidence summits.

level evaluations, as well as an occasional sector or country program-wide evaluation, were found to be used in support of proposed country strategies. Interview data suggest that USAID requirements for the inclusion of evaluation evidence in CDCSs and PADs have encouraged their use in these types of design documents. Evaluations of USAID policies are almost non-existent, and evaluations are rarely used to inform the formulation of new policies.

“Policies, strategies, frameworks and visions are designed in a collaborative, evidence-based process that engages expertise from throughout the agency and from external resources.”
—USAID website

QUESTION 2: WHAT CHANGES/DECISIONS ARE MADE BECAUSE OF EVALUATIONS?

Study findings indicate that changes that result from decisions to act on evaluation findings fall into several clusters, including:

- **Direct actions or changes** that USAID or one of its partners makes based on evaluation findings or recommendations, such as adjustments to priorities or work plans of ongoing activities;
- **Discernible effects of direct action or changes**, such as improvements in the efficiency or effectiveness of an ongoing activity; and
- **Changes in development outcomes**, such as increased numbers of beneficiaries whose lives are improved because activities and projects became more effective, efficient, or sustainable.

This section draws on study data from the survey, group interviews, and documentary sources.

Direct Actions and Changes Linked to Evaluation Results

In the previous section, survey data were presented in Figure 4 that showed that evaluation results were used for new project and activity design at least twice as often as they were used to modify existing activities, projects, and strategies. This section takes a more detailed look at the types of changes that result from decisions made on evaluations results. On this topic, USAID’s annual PPR describes actions taken in more detail than was found in other study sources.

For a relatively small number of evaluations, entries in USAID’s annual PPR provide detailed information on changes that had been made based on evaluations. In the PPR it was possible for entries to identify more than one use. Among these entries, modifications described included:

- Refocusing or prioritizing of interventions within a project or activity (45 instances)
- Revising a project or activity’s work plan (15 instances)
- Revising an activity M&E plan or Performance Management Plan (PMP) (11 instances)
- Extending a project or activity (4 instances)
- Expanding a project to cover additional geographic or technical areas (2 instances)
- Terminating a project or activity (1 instance)
- Developing an exit plan and sustainability strategy (1 instance)

The PPR narratives often lacked specificity as to the exact actions taken, and many were too vague in their statements of what changed to permit categorization. An example of how the PPR was used to describe the results of an evaluation is provided in a one-page briefer on Afghanistan AWDP on p. 17.

USAID staff, in group interviews, largely agreed with PPR reports about the types of changes in on-going activities and project that are made by USAID based on evaluations. According to participants in group interviews, adjustments to implementing partner (IP) work plans and realignment of project or activity priorities are among the most frequent changes made based on evaluations. Survey responses similarly found that direct action or changes are a common response to evaluation findings, with 90 percent of respondents stating that actions were taken based on at least some of the evaluation's recommendations. Survey data further indicate that beyond USAID, partners also make changes in the activities they implement. More specifically:

“The Mission is using the report to inform its phase-out plan by prioritizing aspects of the program according to performance and potential for sustainability. The Mission shared the report with the implementing partner to enable them to learn about their strengths and weaknesses and adjust their programs accordingly; the Government so it can learn more about what is working and what is not to improve maternal, newborn, and infant health; and other relevant development partners so they can use the findings to identify priority interventions for scale-up.”

–PPR Evaluation Utilization Entry

- Nineteen percent of survey respondents indicated that they believed IPs had taken actions based on these evaluations which complemented USAID post-evaluation decisions and led to positive changes.
- Eight percent of respondents reported that partner country government strategies and/or policies reflected learning from the evaluation that was reviewed, and
- Seven percent of respondents indicated that it appeared that other donor programs/projects had applied lessons from the evaluation that was reviewed.

In interviews with representatives of organizations that implement USAID activities overseas and have undertaken internal or “self-evaluations” of their work, the study team asked similar questions about the utilization of these internal evaluations, many of which can be found in the DEC. These organizations indicated that they had acted upon their internal evaluations and had provided USAID with copies, but they were unaware of whether or how USAID had used the results of their evaluations.

Discernible Changes and Development Outcomes

In five group interviews, participants mentioned discussions in portfolio reviews about whether progress was being made on implementing actions that stemmed from evaluations. The study team also asked participants what changes had resulted from taking actions based on evaluations.

The study survey also probed USAID respondents' awareness of whether the evaluations on which they reported resulted in improvements, e.g., increased effectiveness or sustainability. As Figure 6 shows, over half of the survey respondents indicated, through selections they made on a multiple choice question, that they were aware of improvements in the effectiveness of USAID efforts where action had been taken based on evaluation findings or recommendations. With reference to Figure 6, the study team noted that the frequency with which survey respondents selected any possible answer was lower than the frequency with which at least one answer was chosen from most of the other multiple choice questions in the survey. This suggested to the study team that awareness of the effects of post-evaluation actions may be relatively low among USAID staff. Discussions in group interviews at the OU level tended to confirm this impression, as participants in those interviews were quite knowledgeable about what actions USAID had taken, but few were comfortable with questions about what happened because of USAID's post-evaluation actions, and one or two individuals in these interviews volunteered that despite their best intentions, they often did not have time to follow-up after evaluations were completed.

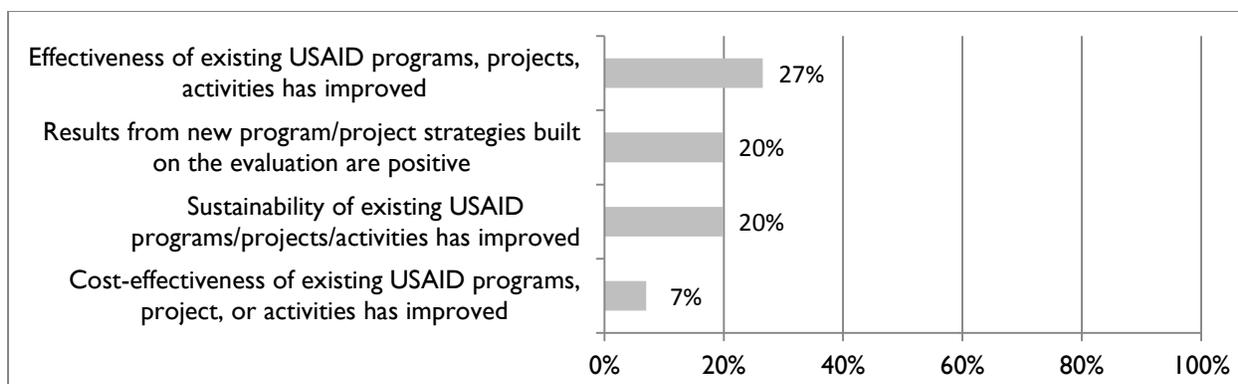


Figure 6. Perceptions of Effects of Taking Action on Evaluation Results (N=118).

The study team consistently asked USAID staff and partners for examples of evaluations that had made a difference at the level of a development outcome, or what might be the level of an IR or DO in a USAID CDCS. In all, 57 stories emerged that provide insights into the ways in which projects and activities had improved. Occasionally these stories described important development outcomes that resulted from actions taken on evaluations. Two of these stories are included as one-page briefers in this section. Annex 7 in this report includes eight additional examples of the most interesting and verifiable of these stories.

Evaluation stories at this level go beyond claims or evidence of improved effectiveness or efficiency. Their results can be demonstrated to have had an observable impact on an important development problem. It was anticipated that verifiable examples of the effects of evaluation utilization at this level might be scarce, based on the results of a similar investigation conducted a decade earlier by the World Bank, which identified and documented eight evaluations they considered to be influential in this sense.¹¹

In summary, this study found that evaluations not only result in immediate decisions about whether projects and activities need to be modified, or new ones planned, they also have downstream effects. More specifically, while little effort is made to trace the development effects of decisions made and actions taken based on USAID evaluations, downstream outcomes of evaluation can be important. USAID country partners often appear to play a critical role in stimulating downstream effects, e.g., by changing country policies and practices.



USAID Rebuild Project, on Twitter, in a Strategic Planning session, which is a natural forum in which to discuss what has been learned from evaluations.

¹¹ The World Bank. "Influential Evaluations: Evaluations that Improved Performance and Impacts of Development Programs." Washington D.C.: International Bank for Reconstruction and Development, 2004 <http://documents.worldbank.org/curated/en/2004/01/6051317/influential-evaluations-evaluations-improved-performance-impacts-development-programs>.

ACTION ON EVALUATION RECOMMENDATIONS IN AFGHANISTAN

USAID/Afghanistan’s Work Force Development Program (AWDP) is working to increase job placements and wages for 25,000 Afghans – at least 25 percent of them women – through a four-year, \$72 million program that seeks to strengthen the labor pool in major economic areas of the country, in close collaboration with local ministries, through the project’s on-budget and off-budget components.

USAID’s mid-term evaluation of AWDP credited the project, stating that “high satisfaction ratings from employers and participants are a reflection of the program’s success,” noting that AWDP has introduced “sound methods of determining labor market needs and rigorous procedures for monitoring and evaluation (M&E) and quality assurance (QA), [which] are largely new to all stakeholders...” The evaluation also showed that “AWDP has documented much greater success facilitating promotions with raises for currently employed workers (86%) than it has in facilitating employment for new job seekers (26%),” noting that the project had trained many more women than originally anticipated. Recommendations largely on the project’s training practice in suggesting that AWDP “agree with grantees on a set of minimum requirements that relate to training systems and processes,” require that “grantees...demonstrate to AWDP that they have aligned their curricula and the modes of delivery and assessment to a competence-based training approach,” and improve “M&E processes and practices to capture more feedback on trainer performance.” (Performance Evaluation: Afghanistan Workforce Development Program (2014) at http://pdf.usaid.gov/pdf_docs/pa00k48w.pdf.)

For Immediate Release, Thursday, May 16, 2013

KABUL, AFGHANISTAN- Kardan University held their second annual job fair on Thursday with 50 private sector companies and over 2,000 job seekers. The event gave skilled Afghan job seekers the opportunity to meet with employers, while giving employers access to highly qualified candidates. Kardan University has received two competitively awarded grants from **USAID’s Afghanistan Workforce**



Photo: Kardan University, Kabul, Afghanistan.

Development Program (AWDP). To date, AWDP has awarded 26 grants to private professional training organizations. Through these grants, over 8,500 mid-level job seekers are gaining skills in high-demand areas. All training providers that are part of the program agreed to place at least 70 percent of the job seekers participating in their training programs. Thursday’s event helped increase job placements, offered job placement support services, and facilitated access to other training opportunities for many Afghans. By the end of the event, more than 300 meetings had taken place between employers and AWDP graduates.

Following up on its mid-term evaluation, USAID/Afghanistan reported in the following year’s annual PPR that “evaluation recommendations that are being implemented include (i) refining M&E processes for the off budget component; (ii) grantees aligning their curricula and the modes of delivery and assessment to a competence-based training approach with an appropriate balance between theory and practical components; and (iii) grantees ensuring that the language of instruction is always appropriate to the needs of the participants; where, for example, English language skills are underdeveloped or where Pashto rather than Dari is the predominant language.”



Photo: Creative Associates, USAID AWDP project.

QUESTION 3: TO WHAT DEGREE AND UNDER WHAT CONDITIONS DOES LEARNING OCCUR FROM EVALUATION FINDINGS THAT WAS NOT ANTICIPATED BY THE INTENDED PURPOSE OF THE EVALUATION?

Under Question 2, the study team reported on actions and changes that flow from decisions based on evaluation findings and recommendations. In this section, USAID’s question is about the influence of evaluations on learning, which does not necessarily translate directly into actions.

Evidence of learning is difficult to capture directly in the absence of some type of before and after comparison. Even self-reported learning is difficult to obtain in group interviews where the main focus is on actions taken. Accordingly, the study team drew heavily on questions that gathered perceptions of what had been learned from evaluation reports and processes. Reporting on Question 3 also benefits from PPR evaluation utilization entries that document where learning from evaluations appears to have occurred, and from interview and survey responses about the degree to which “new” information was provided to USAID staff.

Self-Reported Learning from Evaluations

Responses to a survey question about the period over which learning from evaluations occurs indicated that while most learning from evaluations occurs within the first three months after an evaluation is completed, for some evaluations the learning period has been longer, as shown in Figure 7.

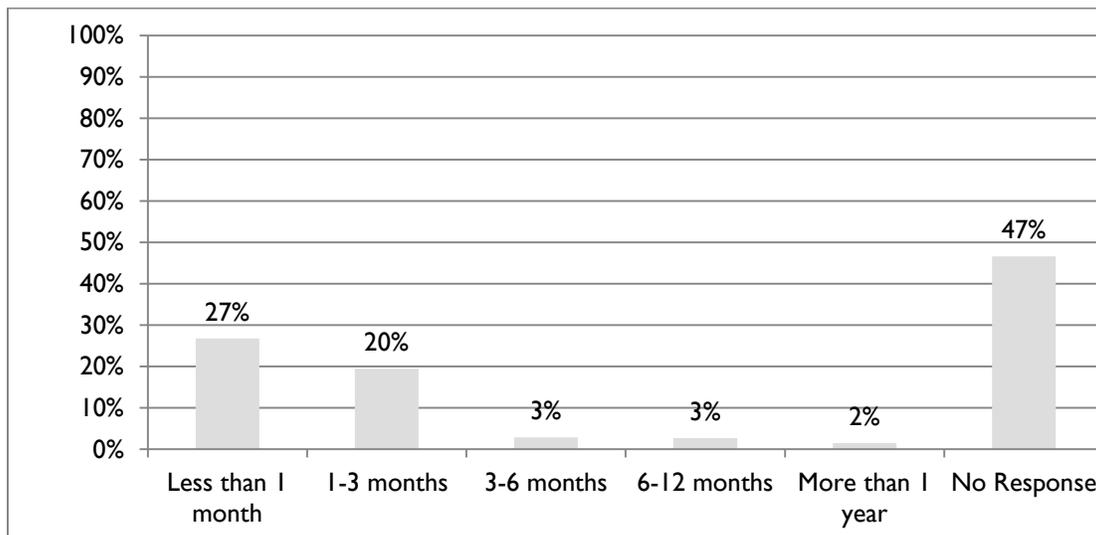


Figure 7. Time Period over which Learning from an Evaluation Occurs (N=118)

To determine what USAID staff felt they had learned from the evaluations on which they reported, the study team used multiple choice, multiple possible response questions. The first asked what respondents had learned about the activity or project that had been evaluated, while the second asked what respondents had learned about the evaluation process from the exposure to a specific evaluation. Each

DEVELOPMENT EFFECTS OF EVALUATION UTILIZATION IN ETHIOPIA

In Ethiopia, a mid-term evaluation commissioned by USAID’s implementing partner for the Mission’s \$92 million Yekokeb Berhan Program for Highly Vulnerable Children revealed that while the percentage of highly vulnerable children (HVC) who were tested for HIV since the activity’s baseline study was conducted in 2009 had increased by approximately 7 percent, and many of those who tested positive were receiving ART and further monitoring through USAID’s project, still only 46 percent of HVC in Ethiopia had been tested and knew their HIV status.

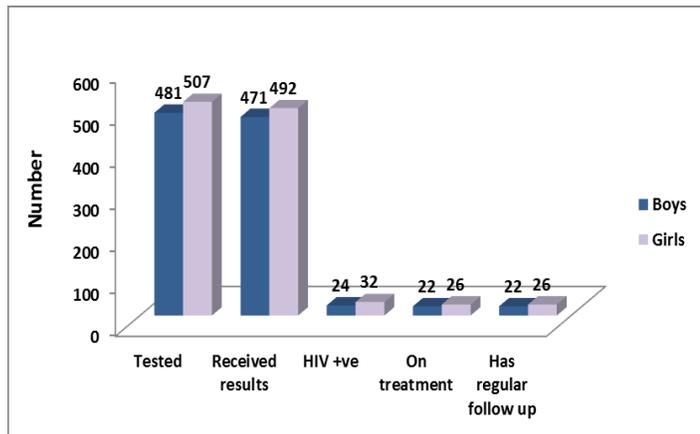


<https://www.youtube.com/watch?v=WrtOgC80ikM#t=56>

Excerpt from: Mid-Term Evaluation Report, Yekokeb Berhan Program for Highly Vulnerable Children in Ethiopia

http://www.pactworld.org/sites/default/files/field-pdf/RFP_Yekokeb%20Berhan%20Mid-term%20Eval%208.16.13.pdf)

Of 963 HVC (481 boys and 507 girls) who received HIV test results by the mid-term, 56 (5.8%) were positive for HIV. When disaggregated by sex, 24 (5.1%) boys and 32 (6.5%) girls reported being HIV-positive and the difference was not statistically significant. Baseline data show that 14.2% of HVC who were tested and received their results were positive for HIV. Altogether, 48, or 85.7 percent, of HIV-positive HVC reported being on ART during the mid-term and of these, 48 (85.7%) reported having regular follow up in a health facility. In the baseline, only 61% of those HIV positive HVC were receiving ART treatment.



Following receipt of this evaluation, USAID staff expressed to government representatives their concern about the low priority HVCs have received in terms of resources for HIV testing compared to



Photo: Boston University, Baseline Evaluation of USAID/Pac Program for Highly Vulnerable Children: Yekokeb Berhan.

populations at lower risk. The State Minister for Health agreed with USAID’s position that HVCs should become one of Ethiopia’s priority populations for HIV testing. Later in 2014, the National Guidelines for Comprehensive HIV Prevention were revised. As a result, HIV testing for HVCs is now a priority in work plans for USAID

question included pre-determined responses with the option for explaining a response of “other.” Potential issues with responses to both of these questions stem from the fact that learning is self-reported, and respondents may have felt obligated to give at least one answer to each of these two questions.

With respect to the first question, 94 percent of respondents claimed to have learned about projects and activities in some manner. Detailed responses on learning about projects and activities that were evaluated indicate greater learning about the specific project or activity involved, rather than learning about one activity in relation to a broader similar group of projects or activities. Lower response rates for some topics, such as unplanned consequences and problems beyond an IP’s control, may suggest that there was no information on these topics in evaluations on which the survey collected data. Overall responses to the first of two survey questions about learning indicate that the evaluation affected:

- Opinions on the merits of the project or activities (52%)
- Views about the soundness or adequacy of the design (54%)
- Understanding of how to make this type of project more effective in the future (49%)
- How respondents think about other activities in the same sector or topic (44%)
- Understanding of why some results were or were not achieved (30%)
- How respondents think about partner collaboration in the Program Cycle (36%)
- Understanding of the unplanned consequences of projects like this (29%)
- Understanding of schedule and budget problems beyond the IP’s control (27%)

Questions about the responses listed above were worded to reveal not only whether a particular type of learning had occurred, but also whether that learning had come from reading the evaluation report, participating in the evaluation process, or both. Figure 8, which includes only the first three response options described above, depicts the “both” as a white line and shows that this was the most frequent way in which participants reported they learned. At the same time, Figure 8 shows that among those that chose between learning from an evaluation report and learning from participation in the evaluation process, learning from the report dominated. This pattern was consistent across all of the response options for which a total percentage that learned about each option is shown above.

In group interviews with USAID staff at the OU level, discussions about what staff had learned from evaluations indicated that being surprised by findings in an evaluation was not that common. Even problems that evaluations surfaced were described as being things that staff already knew or had a sense might be an issue. New information, which the study survey indicated was included in 52 percent of the evaluations examined, was generally perceived by group interview participants as being a positive feature of evaluations even when that new information included negative findings USAID was not aware of.

“The results of this evaluation were used to evaluate activity progress and provide lessons learned for other USAID missions in designing and implementing current and new programs in energy infrastructure construction and oversight services. This activity is very different from USAID’s traditional assistance programs in the country. It utilizes an innovative mix of both private sector and government-controlled organizations as implementers, and has direct contracts as well as host-country contracts which is unique. This evaluation has highlighted important lessons learned not only for the Mission but for the Agency as a whole. Lessons learned from the evaluation focus on sustainability considerations of large-scale infrastructure programs, implementation modality and other important aspects of this type of assistance.”

–PPR Evaluation Utilization Entry

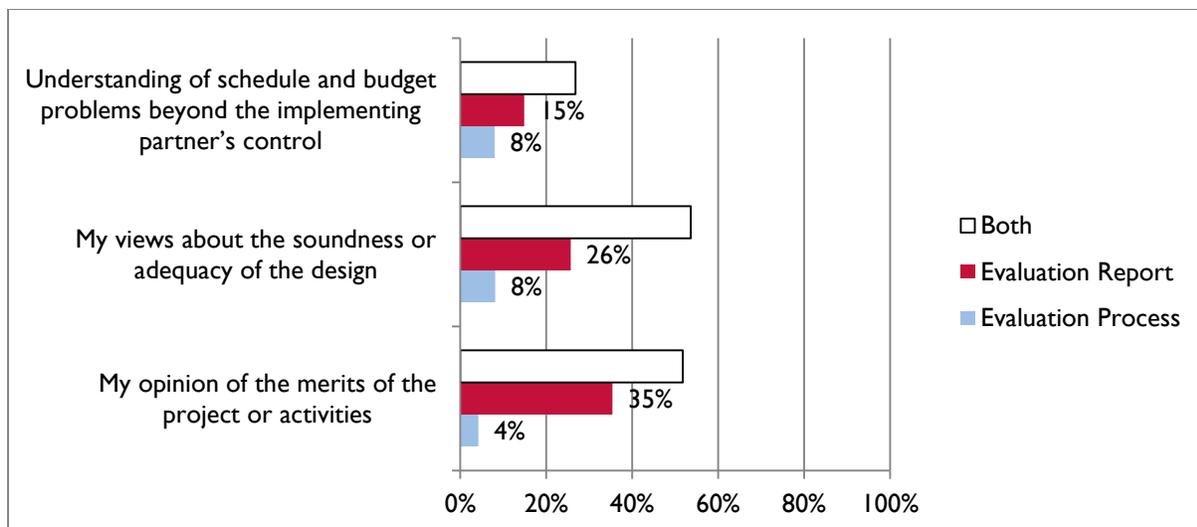


Figure 8. How Staff Learned about Projects/Activities from USAID Evaluations (N=118)

On the second question about types of learning that the survey examined, 97 percent of respondents claimed to have learned more than they had previously known about evaluation methods and processes, even though learning about evaluations was not an explicit aim of their participation. While the presence of this question in the survey indicated that the study team thought such learning was possible, it had not found a large body of empirical studies to support that hypothesis. On the other hand, evaluation scholars, most notably Michael Quinn Patton, had suggested over a number of years that such learning would be a side benefit of participation on teams or in evaluation management processes. In response to questions about their learning about evaluation, respondents reported that both participation in the process and reading the evaluation on which they answered survey questions, had affected their:

- Understanding of the purpose and role of evaluation (77%)
- Ability to review or critique evaluation products (76%)
- Ability to engage stakeholders in an evaluation processes (73%)
- Ability to lead or support post-evaluation action planning and follow-up processes (71%)
- Understanding of evaluation types and methods (70%)
- Interest in monitoring and evaluation (69%)
- Understanding of how to oversee an evaluation team (65%)

One other answer volunteered on this question indicated that through exposure to this evaluation, the respondent had come to realize the importance of “our office's standing team that reviews evaluation SOWs and helps with the design. It was a stronger evaluation because of this body.”

New Information

The evaluation utilization literature suggests that novel or new information may contribute to learning and other forms of evaluation utilization. Accordingly, the study team raised questions in interviews and the survey about whether evaluations included information that was “new” to USAID staff.

“We have general beliefs and feelings about the likely answers but sometimes it's helpful to have them confirmed. An evaluation may not give us new information that we didn't already suspect, but the fact that it confirms it is useful in and of itself. Now some may argue that that wasn't money well spent if it doesn't tell us something that we didn't already know but I think it's also how you use the information.”

—USAID Mission Staff Member

On this topic, survey respondents indicated whether, based on their personal knowledge, at least some information was presented in an evaluation that was previously unknown to USAID staff. This was reported to be the case for 52 percent of the evaluations on which the study team has survey data. Group interview responses concurred, noting that some but not all evaluations provide them with new information. At the same time, group interview participants indicated they appreciated that evaluations often provide evidence confirming what they already thought they knew. New information was generally perceived by group interview participants as being a positive feature of evaluations even when that new information included negative findings USAID was not aware of.

Survey responses mirrored this sentiment, with 74 percent of respondents indicating evaluation results confirmed things USAID already knew. In one interview, a Mission Director said that while she had a negative impression of one of the Mission’s activities, she only acted to close it down when an evaluation confirmed what she had suspected.

Promoting Learning from Evaluations

USAID staff pointed out in group interviews that when their Missions or offices have a robust learning agenda and a variety of learning activities underway, learning from evaluations is often a focus. One Mission reported that it has an evaluation “book club” that meets regularly; another described how staff in that Mission post summaries of what evaluations found in prominent locations to foster learning. In another Mission, learning activities are sometimes convened at project sites, where participating staff engage in direct observation and discussions of performance monitoring and evaluation findings.

Two OUs mentioned in group interviews that in addition to learning from individual evaluations, they are attempting to learn from clusters of evaluations. Illustrative examples of USAID efforts to synthesize findings from clusters of evaluations include the “State of The Field Report: Examining The Evidence In Youth Workforce Development” meta-analysis under the USAID Youth Research, Evaluation, and Learning Project by the Aguirre Division of JBS International, Inc. in 2013, and two synthesis reports prepared by the Bureau for Economic Growth, Education, and Environment (E3), covering the findings from evaluations worldwide in each of these sectors and topical arenas in which E3 works.



Photo: MIT. USAID program releases evaluation of water filters (2015).

To summarize, MSI’s answer to USAID’s question about what degree and under what conditions does learning occur from evaluation findings that was not anticipated by the intended purpose of the evaluation shows that learning from evaluation by USAID includes learning about conducting evaluations as well as learning what evaluations found out from examinations of programs, projects, and activities. Further, evaluations were found to stimulate new thinking when they presented new information, but equally important for action are evaluations that confirm what USAID staff already suspect to be true. In addition, the answer to this question demonstrated that it is not only USAID that learns from its evaluations; its country and implementing partners also learn and apply the results of USAID’s evaluations.

QUESTION 4: WHAT PARTICULAR BUSINESS PROCESSES OR ENABLING CONDITIONS APPEAR TO ENCOURAGE OR DISCOURAGE THE UTILIZATION OF EVALUATIONS?

In this section, the focus of the study shifts to the mechanisms and other factors in USAID's environment that lie behind its current evaluation utilization practices and status. The section is divided into two distinct discussions:

- USAID processes that directly support evaluation utilization, including building awareness of evaluation results through dissemination; post-evaluation reviews and associated action planning; monitoring the implementation of evaluation-based decisions; and utilizing evaluation evidence to develop CDCSs and PADs, including the aggregation of evaluation findings for this purpose.
- Characteristics of evaluations and the evaluation process that appear to facilitate or impede evaluation utilization, including evaluation planning; evaluation relevance, as related to evaluation timing; partner participation in the evaluation process; evaluation report quality; and human and financial resources for evaluation.

Following these two discussions, a graphic ranks the frequency with which the factors discussed under Question 4 are present according to survey responses.

Processes that Directly Support Evaluation Utilization

USAID directly supports evaluation utilization through processes called for in Agency-wide guidance, including the 2011 Evaluation Policy and the Automated Directives System (ADS). The section focuses on the dissemination of evaluation results to USAID staff and other stakeholders and on USAID's implementation of post-evaluation review and action planning steps that the Agency recommends to support evaluation utilization.

Dissemination of Evaluation Results

This section focuses on deliberate efforts that USAID staff make to generate an awareness and understanding of evaluation findings and recommendations among Agency staff, USAID partners and other stakeholders. Dissemination, in this sense, does not include action on USAID's requirement that a copy of every evaluation be sent to USAID's DEC, which stores and supports searches for Agency evaluations.¹²

Three general approaches to dissemination were examined through the study survey in interviews, namely:

- Briefings in which the study team provides USAID, and in some cases other stakeholders as well, an overview of the evaluations purpose, methods, findings, conclusions, and recommendations, and then entertains questions about their study.



Briefing on the findings of a USAID/Kazakhstan evaluation of its Business Connections project.

¹² The study analysis examined whether the delivery of evaluations to the DEC is associated with evaluation utilization at a statistically significant level and found that it is not.

- Deliberate distribution of copies of the evaluation report, or its executive summary, in draft and/or final form.
- Supplementary dissemination events such as workshops, webinars, or conference presentations.

The first dissemination approach utilized with most evaluations is a briefing with the evaluation team. Briefings with the evaluation team, listed as the first of six steps to take in responding to an evaluation in ADS 203, were reportedly held for USAID staff for 88 percent of the sample of evaluations examined through the survey, as Figure 9 shows.

Group interviews suggest a slightly lower use of this approach, with 6 of 12 USAID Missions indicating that they routinely hold such briefings. Both interview responses and survey data indicated that USAID implementing partners and country partners are sometimes present at these evaluation team briefings, but country partners, in particular, attend only a third of such briefings organized for USAID staff. Interviews indicated that in overseas Missions, senior staff were generally represented at these debriefings, and the Mission Director frequently attended.

Depending on how inclusive post-evaluation briefings are, awareness of a particular evaluation’s findings across DO teams and Offices within Missions may vary. The same is true for awareness of evaluation findings between field posts and Washington, with some USAID/Washington staff indicating that their OU is routinely briefed, whereas others reported the opposite. In the same vein, field staff in interviews indicated that they were not aware of findings from evaluations in other OUs in their sectors, and found it difficult to access that kind of information. USAID staff also noted that their awareness of evaluation findings within and beyond their OUs was often linked to conversations about USAID’s guidance on using evaluation evidence in developing PADs and CDCSs, and some noted difficulties when using the DEC to try to locate this type of evidence. One interviewee indicated they had directly contacted individuals they knew in other Missions for information about their evaluations.

“When I was trying to design a potential domestic violence activity, I went into the DEC to find out what has been happening in other Missions. It’s a somewhat laborious process to find how evaluations and other documents informed our design process. It would be great if I could call into PPL/LER and say, “I’m designing this. Can you give me a summary of what has already been done?”

–USAID Mission Staff Member

Given the injunction in USAID’s Evaluation Policy and ADS place to “share and openly discuss evaluation findings, conclusions, and recommendations with relevant customers, partners, other donors, and stakeholders,” the study team asked USAID staff why *evaluation* dissemination efforts for specific evaluations were much less frequently directed at partners than staff, as shown in Figure 9. USAID staff explained some of this difference, particularly with regard to country partners, by noting that not all of the projects and activities USAID evaluates directly involve government ministries. One staff member said that sometimes USAID plans to reach out to others concerning evaluation results, but end up being too busy with other obligations to do so.

The quantitative analysis showed a statistically significant relationship between USAID’s dissemination of evaluation results to country partners and survey respondents’ perceptions about whether (a) the project or activity became more effective or sustainable when evaluation results were utilized, or (b) partner government strategies or policies reflected learning from the evaluation.¹³

¹³ This finding was significant at the .05 level.

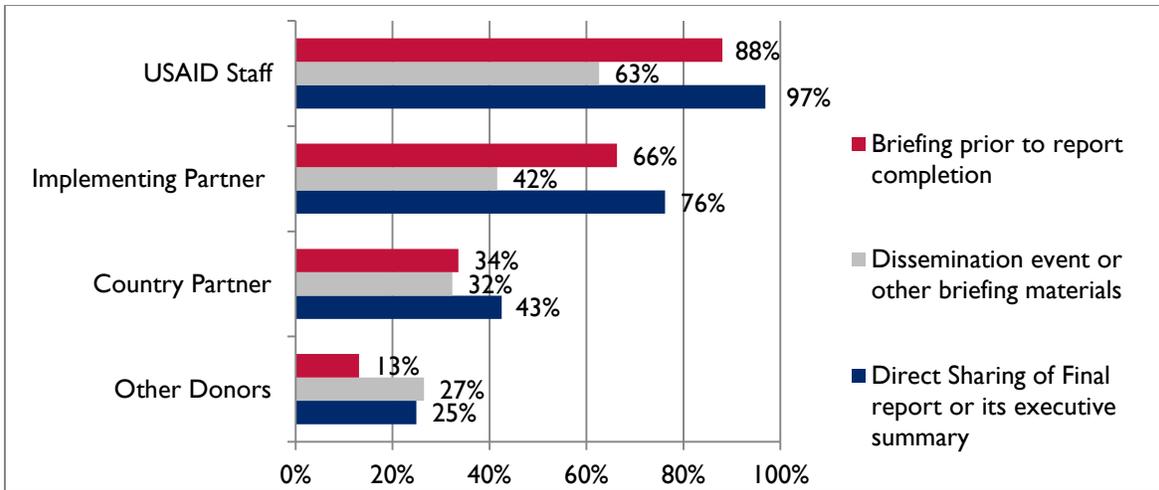


Figure 9. Audiences Benefiting from USAID Evaluation Dissemination

Thus, while dissemination to country partners was not found to “cause” increased effectiveness or sustainability in USAID activities, country partner awareness of evaluation findings, together with USAID and implementing partner efforts, may have helped achieve those results. Similarly, dissemination to country partners was not found to “cause” country partner changes in government policies; but government partner awareness of evaluation findings may have played a role in their decisions. An example of this latter relationship is described for Ethiopia in an evaluation use “briefer” on page 19.

Additionally, study data indicate that there could be a relationship between USAID dissemination of evaluation results to implementing partners and implementing partners making changes that complement USAID post-evaluation decisions and follow-up actions, but evidence on this fell just short of statistical significance.

“We all intend to share that with the broader stakeholder group but I think because of many reasons, one practical reason is getting busy with many other schedules, often we don’t do it perhaps. If the evaluation is ex post, perhaps we don’t share much with other broad stakeholder groups. But if it is of an activity evaluation for a handoff, like when we evaluated a family planning activity, we always start with a pre-evaluation meeting with the government.”

–USAID Mission Staff Member

USAID Staff Suggestions

- Raise the status of evaluation briefings. Having too many “brown bags” rather than meetings during regular hours says to staff that evaluation is not important (1 OU)
- Pilot new approaches that Missions might be able to copy and utilize – help develop a culture of dissemination, make reasons for sharing with stakeholders clearer and raise its priority (5 OUs)
- Develop guidance on how to write evaluation summaries and provide samples of two page briefings on evaluations (2 OUs)
- Build a culture where people want to disseminate the evaluations they commission (1 OU)
- Develop knowledge syntheses on topics and make them widely available (14 OUs)
- Create an online platform for sharing evaluation results in various ways (2 OUs)
- Create a centralized database of evaluations; an easier approach to accessing them than the DEC is needed (4 OUs)

Notably, while USAID’s Evaluation Policy calls for dissemination plans for every evaluation, USAID participants in group interviews made no suggestions for improving evaluations that involved such plans. Across all of the group interviews, only one staff member mentioned dissemination plans, saying that their OU is starting to add a few new elements into its SOW development process, and dissemination plans are one of them.

Post-Evaluation Action on Evaluation Results

USAID guidance intended to foster evaluation utilization is found primarily in USAID’s ADS section, “Responding to Evaluation Findings,” shown in the text box below. Steps 1 and 6 focus on the dissemination of evaluation results and were discussed in the previous section. This section focuses on the decision making and action planning processes outlined in Steps 2-5.

Responding to Evaluation Findings: Post-Evaluation Decisions and Actions

“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 the 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, 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.”

USAID ADS, 203.3.1.9
Effective date: 1/31/2003

The study team examined what happens in OUs following the receipt of an evaluation. Table 4 shows survey and interview data side by side for as many of the ADS post-evaluation steps as those data sets allow. While data from these two sources are not directly comparable, both of them indicate that, in practice, less than half of USAID evaluations appear to receive the kind of post-evaluation attention the ADS envisions.

Table 4. Post-Evaluation Actions According to Survey and Interview Responses

Post-Evaluation Actions	Survey Responses (by evaluation) N=118	Group Interviews (by OU) N=24
Held post-evaluation meetings	53%	63%
Accepted/rejected recommendations	34%	21%
Decided on actions to be taken	34%	46%
Set timeframes	18%	--
Prepared a written Action Plan	38%	--
Tacking implementation of actions	--	21%

In addition to examining whether and to what degree USAID takes post-evaluation steps in support of learning and utilization, the study team also examined whether accepted evaluation recommendations were actually implemented. Only two evaluations, according to the survey results, provided no recommendations, and thus were not considered in this part of the analysis. Several facts about the implementation of recommendations found in USAID evaluation reports are summarized below:¹⁴

- Of 45 surveyed evaluations, the number of recommendations ranged from zero to more than seventy in any given report, with an average of eighteen.
- On average, survey respondents indicate that their preferred number of recommendations for evaluations would be between three and ten.
- In one-third of the evaluations that included recommendations, respondents indicated that 75 percent or more of the recommendations were reportedly accepted by the OU.
- For 79 percent of the evaluations where some recommendations were accepted, more than 50 percent have been implemented.
- For 22 percent of the evaluations where some recommendations were accepted, 100 percent of the recommendations have been implemented, according to survey respondents.

“He was, of all the Mission Directors I've ever worked for, the most involved and interested in learning from evaluations and making sure that the teams apply what we learn to correct our programming. He also held the team accountable for reporting out and in fact we've started a process here where we use a tracker. I think that PPL/LER recommended this and we've already instituted it. It's a tracker of the recommended actions and what we've done about them.”

–Regional Mission Staff Member

The study team examined whether there is a relationship between the OU's implementation of the post-evaluation action planning steps from the ADS and the successful implementation of accepted evaluation recommendations. The study found that the relationship between them is statistically significant. In other words, OUs that implemented 75 percent or more of the evaluation recommendations had also undertaken at least some of the post-evaluation steps listed in the ADS.

Another factor may be helping a few USAID OUs ensure that the evaluation recommendations they elect to act on are fully implemented. This factor, or tool, is the action trackers that five of OUs interviewed indicated they had developed and use to monitor the status of post-evaluation actions on recommendations they have adopted. Figures 10 and 11 below display two of the seven different action tracking tools the study team received from USAID staff, including staff from OUs that were not interviewed.¹⁵

Figure 10. Example Template for Summarizing Post-Evaluation Decisions and Tracking Action Status

No.	Findings and Recommendations	Management and Program Actions Needed	Individual Responsible for Completing Actions	Proposed Date for Completion of Actions	Status of Actions
1					
2					
3					

¹⁴ The study team notes that, according to survey responses, actions were generally taken on recommendations rather than on evaluation findings.

¹⁵ Midway through this study, PPL/LER released its Evaluation Toolkit, first for Mission comment, and subsequently as a resource available through USAID's Learning Lab. This toolkit includes sample templates along the lines shown above. The toolkit can be found at: <http://usaidlearninglab.org/evaluation>

Figure 11. Example Template for Summarizing Post-Evaluation Decisions and Tracking Action Status

Evaluation/Study Recommendation	Mission Decision Accept/Reject	Responsible Party/Person	Action Recommended	Deadline for Action	Evaluation/Study Recommendation

When discussing post-evaluation meetings and tracking actions on recommendations, one staff member from a Technical Bureau and two from PPL/SPP reminded the study team that USAID did these same sorts of things in earlier eras. These individuals more specifically recalled a form that used to be the cover sheet used by OUs when sending evaluations to Washington.¹⁶ This form included information on post-evaluation actions the OU was taking as well as an abstract of the evaluation. These abstracts, in a pre-internet era, were compiled and shared back with OUs as a quick reference meant to help staff worldwide identify evaluations that might be relevant for their future planning purposes.¹⁷

“This agency used to have a form when I joined. You had to give a summary of the evaluation on the first page then you had to list chief recommendations and then you had to say whether you accepted them. Then you had to give a date when you would address them and you had to give a responsible party. This got signed off by the Mission director or the head of the operating unit. We don’t have that anymore but if USAID were serious about doing this...it is worth a couple of pieces of paper to document intent to follow up on the major recommendations. Right now that doesn’t exist as a requirement and it is something that could make our system a little bit more formal.”

–USAID Bureau Staff Member

Suggestions from USAID staff on ways to improve the likelihood that post-evaluation meetings would be held and would foster action on evaluation findings and recommendations are displayed below.

USAID Staff Suggestions
<ul style="list-style-type: none"> • Make OUs more accountable for evaluation utilization (1 OU) • Develop a more detailed step-by-step guide for post-evaluation action planning and tracking (4 OUs)¹⁸ • Share Missions’ post-evaluation trackers and experiences with other Missions (8 OUs)¹⁹ • Change policy to require portfolio reviews to more formally examine evaluation results and actions taken based on them (3 OUs) • Create incentives for using evaluation results that will recognize and reward staff efforts (4 OUs)

¹⁶ A sample of the form these staff members recalled (AID 1330-5) can be found in Annex B, p. 47, of a report prepared for USAID entitled *Trends in International Development Evaluation Theory, Policy and Practices* (2009) at: http://pdf.usaid.gov/pdf_docs/PNADQ464.pdf

¹⁷ An example of USAID’s earlier practice of using evaluation abstracts draw from copies of FORM AID 1330-5 to inform staff about the existence of potentially useful evaluations is provided at http://pdf.usaid.gov/pdf_docs/PNABK292.pdf. A more recent example of the use of evaluation abstracts as a quick guide to relevant evidence is a volume created for USAID’s mobile education technologies initiative in the E3 Education Office that coded 58 evaluation abstracts by their focus and by the their evaluation design and methods, as a proxy for evidence strength, at: http://www.meducationalliance.org/sites/default/files/meducation_evaluation_abstracts_6-30-15.pdf

¹⁸ After these interviews were conducted, PPL/LER released its Evaluation Toolkit, which includes resources that support post-evaluation action planning.

¹⁹ Ibid.

Evaluation Support for Utilization for Planning Requirements

As discussed under study Question 1, ADS 201 indicates that CDCSs and PADs are to be evidence-based, drawing on published literature, evaluations and other sources. This ADS section goes on to state that “the Mission must reference the assessments and evaluations used to reach significant conclusions in the CDCS.” Evidence that more than 50 percent of approved CDCSs include this element was already discussed under Question 1. In OU interviews, one participant reported that his Mission intends to implement several evaluations specifically to support its next CDCS development round. In group interviews, other USAID staff indicated that procedures for identifying and obtaining evidence from USAID evaluations need to be simpler and less time-consuming.

USAID Staff Suggestions

- Provide summaries of collections of lessons from evaluations by topic, broken down by region (9 OUs)
- Create an online platform or hold group conference calls among Missions to share evaluation findings and lessons (5 OUs)
- In place of the DEC, construct some sort of utilization-focused database (4 OUs)
- Create a centralized location for impact evaluations in particular (1 OU)
- Create a newsletter to disseminate summaries of evaluation evidence (2 OUs)
- Technical Bureaus could do more syntheses (1 OU)

Evaluation Utilization Monitoring at the Agency Level

Monitoring evaluation utilization at the Agency level is a matter of concern to all development agencies, some of which have developed robust systems for this purpose. As discussed in USAID’s 2009 *Trends in International Development Evaluation Theory, Policy and Practice*, donors have strengthened their guidelines for post-evaluation follow-up. Several organizations (SIDA, World Bank, DFID, and UNDP) are monitoring the frequency with which evaluation recommendations are accepted and, once accepted, whether and how thoroughly they are implemented.²⁰

Two other studies released by the Organization for Economic Cooperation and Development/Development Assistance Committee (OECD/DAC) took a more in-depth look at evaluation practices in a wide variety of organizations. These studies referred to “management responses,” which typically consist of a written, formal response identifying the agreed-upon follow-up actions on recommendations made by evaluators. According to one study, 20 of 24 agencies had a mechanism in place to ensure such management responses are drafted in response to evaluation findings, and that follow-up actions take place, even if the mechanisms did not always work well; 17 percent claimed that they did not work well. Some DAC members, such as the Asian Development Bank, even went so far as to publish these management responses alongside the evaluations on their websites.²¹ Another DAC study, published more recently, claimed that about two-thirds of DAC members have such mechanisms in place.²²

At the Agency level, USAID’s current mechanism for collecting data on evaluation utilization is a section of the annual PPR template set aside for that purpose. On this template, USAID staff identify evaluations their OUs have undertaken during the current fiscal year and describe how those evaluations may have

²⁰ Op. cit.

²¹ OECD, “Evaluation in Development Agencies,” http://www.oecd-ilibrary.org/development/evaluation-in-development-agencies_9789264094857-en

²² Ibid.

been utilized in that time period. By its nature this reporting mechanism excludes information about evaluations conducted in previous years that were utilized in important ways after the fiscal year when they were undertaken. According to USAID staff who participated in this study's validation workshops, PPRs also under-report evaluation utilization because they are not viewed as being serious reporting mechanisms, and Agency staff do not always provide thoughtful entries for them, even for utilization that occurred during the reporting year.

Evaluation Characteristics and Utilization

This section examines aspects of USAID evaluations that, according to study data, may either support or impede evaluation utilization.

Evaluation Planning and Statements of Work

In group interviews, USAID staff from seven OUs cited weaknesses in evaluation planning as an impediment to producing useful evaluations. Six of these OUs highlighted poorly crafted evaluation questions as a problem. This is the only data source that identified evaluation questions, or other aspects of a SOW, as a potential impediment to evaluation utilization.

USAID Staff Suggestions
<ul style="list-style-type: none">• Improve guidance on what should be evaluated, using what type of evaluation, and when. (2 OUs)• Update SOW guidance that includes good examples of what is needed when drafting SOWS (3 OUs)• Create new guidance on writing good evaluation questions that actually shows how to write them (3 OUs)²³• The Evaluation Policy should be changed to encourage broader evaluations that focus on more than just one project or activity, to foster deeper learning (1 OU)• PPL should encourage more ex-post evaluations (1 OU)

Evaluation Timing and Relevance

Timeliness and relevance emerged as inter-related factors that can affect utilization. USAID staff from 11 of 24 OUs interviewed cited problems with the timeliness of evaluations as an impediment to evaluation utilization. Specific problems included the receipt of an evaluation too late in a project or activity cycle to act on its findings within that cycle as well as evaluations that are completed too late to inform follow-on project or activity designs and related solicitations. According to USAID staff, while the timeliness of an evaluation is conceptually separate from its relevance, in practice the relevance of evaluations, particularly their recommendations, suffers when the timing of evaluations is less than optimal.

Survey data on the timeliness and relevance of evaluation reports confirm that USAID receives both well timed and poorly timed evaluations. On balance, however, positive experiences are much more common, as highlighted in group interviews:

- 94 percent of survey respondents indicated that the evaluation on which they reported had been designed to be relevant for upcoming decisions of the OU.
- 88 percent of these respondents indicated that the evaluation was considered to still be relevant for that purpose, while 12 percent were not.

²³ After interviews were completed for this study, PPL/LER released an Evaluation Toolkit that includes a reference paper on writing evaluation questions that is available at: http://usaidlearninglab.org/sites/default/files/resource/files/tips_for_developing_good_evaluation_questions_july_2015.pdf

At the same time, however:

- 26 percent of respondents, when discussing evaluation recommendations that had not been acted on, indicated that they were no longer relevant, as the situation had changed.

In group interviews, participants indicated that the mismatch between evaluation timing and programming information needs cannot necessarily be solved by conducting evaluations earlier because activities are often designed to yield most of their results toward the end of the project, and would be missed if final evaluations were conducted too early. USAID staff also identified the length of time needed to procure the services of an evaluation team as a contributing factor to the timeliness problem. They also stated that the problems are compounded when contracting mechanisms that OUs use, such as Indefinite Delivery, Indefinite Quantity Contracts (IDIQs), hit their contractual ceiling and cannot be used without modification, or come to the end of their lifecycle without being rapidly replaced.

Survey data, in contrast, show that 91 percent of evaluations examined were described as being received “in time” to support decisions. A somewhat smaller percentage (85 percent) were found to be “on time,” meaning on schedule regardless of decision timelines. These data suggest that while Mission experience with poor timing can be problematic, as interview data indicated, the absolute number of USAID evaluations that are underutilized due to timing and associated relevance issues may be relatively low. Nevertheless, the study’s quantitative analysis supported Missions’ concerns about timing as an impediment to evaluation utilization. It found a statistically significant relationship between the timeliness of evaluations and how much respondents learned from those evaluations.²⁴ Comments from staff interviews indicate that less attention is paid to evaluations when they are not timely.

USAID Staff Suggestions

- Shorten the procurement cycle for evaluations under established mechanisms (2 OUs)
- Within OUs, plan evaluations and other Program Cycle actions in a more integrated way (3 OUs)

Partner Participation in the Evaluation Process

USAID, in ADS 203 – Assessing and Learning, strongly encourages Missions and Offices to include partners when planning and conducting evaluations, and further calls for widely sharing evaluations with these and other stakeholders. This precept echoes voices in the evaluation literature that assert a linkage between participation and utilization. Thus, the study survey included a question about aspects or stages of the USAID evaluation process in which implementing and country partners participated. Responses showed that implementing partners in USAID evaluation process at one or more stages more than 50 percent of the time, while country partners, generally meaning government or civil society, participated less than half the time, as Figure 12 shows.



Photo: CIFOR, Evaluation of community and government Co-managed forestry resources

²⁴ This finding was significant at the .01 level.

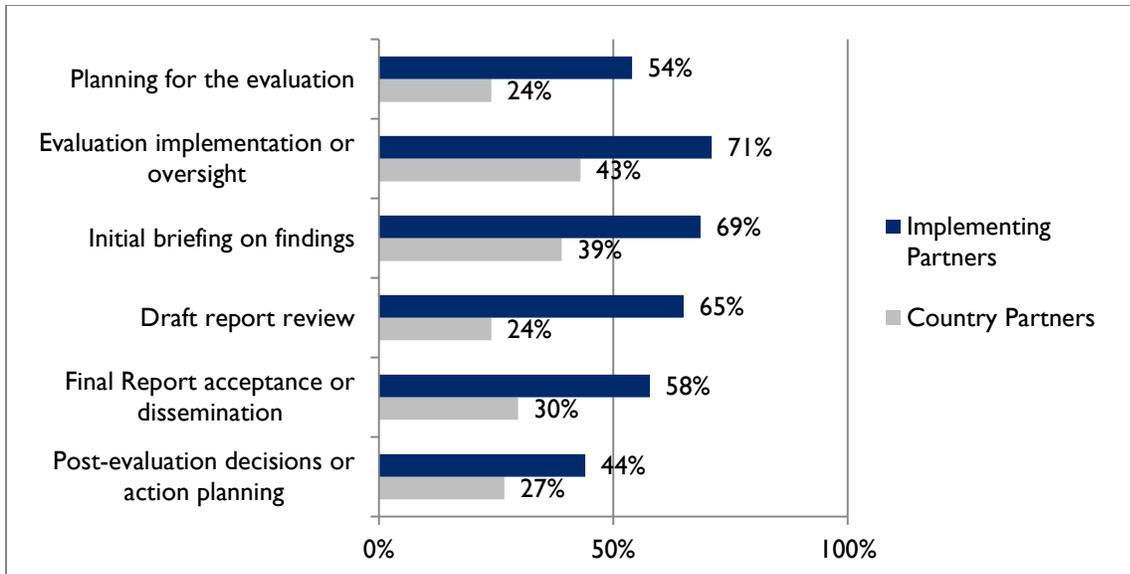


Figure 12. USAID Involvement of Partners in Evaluation Processes (N=118)

As Figure 12 indicates country partner participation in evaluations was evident in during the implementation of evaluations 43 percent of the time. In some cases this simply means that someone from a Ministry participates as a full-time evaluation team member. Less frequently, the methodology for an evaluation is highly participatory and engages country partner personnel at several levels. An example of this type of intense participation is described in the photo below, taken from a methodology description prepared by Management Sciences for Health (MSH) for a 2010 participatory evaluation of a health program in Malawi which it led.

The study's quantitative analysis did not find a direct relationship between implementing or country partner participation at some stage in the evaluation process and whether actions were taken based on an evaluation, but this analysis did indicate that a statistically significant relationship does exist between



Photo: Management Sciences for Health/Natalie Campbell.

whether USAID staff learned from an evaluation, with learning being a form of utilization, and the involvement or participation of country partners in some stage of the evaluation process.²⁵ This finding does not explain why learning is higher for USAID when country partners participate in the evaluation processes, but it does suggest hypotheses or propositions that could be investigated further, e.g., whether USAID staff view country partners as providing insights about evaluation findings that USAID might not otherwise be exposed to, and thus learns from.

²⁵ This finding was significant at the .05 level.

Country Partner Participation in USAID Evaluations

“To evaluate the mobile phone intervention, a participatory evaluation method called Net-Map was used” along with other study methods. For the Net-Map analysis, health workers (CHWs) and “district personnel discussed information needs and gaps and the roles of different actors in their information networks. They then used drawings and 3-dimensional objects to create baseline and endline maps showing the linkages and levels of influence among members of the information network. Net-Map provided them with... evidence of differences before and after the mobile phone initiative.” Then, “at workshops, project evaluators entered data from hand-drawn maps into Visualyzer™ software to create computer-generated maps. On the maps, the size of each node corresponded to the height of the ‘influence tower’ for that actor, as seen in the baseline and endline maps. This allowed evaluators to quantify centrality of the roles... At baseline, CHWs were not mentioned as actors in the information network, while at endline they were seen to have significant connections with colleagues, beneficiaries, supervisors, and district health facilities, as both recipients and providers of information.”

– Management Sciences for Health (MSH), 2010
online synopsis of methods used in an evaluation of a
USAID mobile phones for health project in Malawi

Evaluation Report Quality

In group interviews, 10 of 24 OUs identified inconsistent quality or low quality evaluations as an impediment to utilization. Staff in eight of these OUs linked poor quality evaluation reports to the qualifications of the evaluation teams that produced them. Survey results provided a more balanced view of how evaluations as a whole are perceived, with respondents rating overall credibility as excellent for 28 percent of them, good for 59 percent, and inadequate for only 6 percent. Similarly, 79 percent of evaluations were rated as having excellent (36%) or good (43%) levels of evaluation expertise on their teams, while fewer than 12 percent were rated as having inadequate evaluation expertise.

Commenting on an evaluation report the Operating Unit had received, one staff member said “It was a doozy. I ended up talking to some of my colleagues, and they mentioned a history of not getting evaluation reports that were up to snuff. The evaluation process is a larger issue that I think needs some attention to make sure that we get quality products.”

–USAID Mission Staff

Evaluation methods were also rated highly, with 82 percent of evaluations described as having excellent (20%) or good (62%) designs and methods, while 5 percent were rated inadequate on this factor. Similarly on the overall credibility of evaluations for learning and decision making, 87 percent were rated as being good or excellent, compared to 6 percent that were coded as being having inadequate credibility. Further, data available on the evaluation report quality scores for 241 of the evaluations in the study universe shows that the average scores for these evaluations has risen from

an average of 5.93, for those scored under USAID’s 2009-2011 meta evaluation to 7.34 out of maximum score of 10 for those scored under a 2013-2014 sector synthesis prepared by the E3 Bureau using the same methodology.²⁶ Thus, while OU interviews indicate that poor quality of an evaluation impedes evaluation utilization, the number of such evaluations may be fairly small, which appears to contradict survey data.

The study analyzed the relationship between an individual evaluation report’s quality and its ultimate utilization, but found no significant relationship. Given the sense of conviction in group interviews that such a relationship exists, the study team looked at this question from a more aggregate level, namely the average evaluation report quality score at the OU level. At this level the relationship between

²⁶ Op. cit. – Meta-Evaluation (2013).

evaluation quality and utilization was visible. Average OU quality scores are statistically associated with learning from evaluations (at the .05 level) and, separately, at the same level of significance, with actions taken based on evaluations. This is not to say that evaluation quality scores cause utilization. A more likely explanation, but one which the study team did not actually explore, might be that OUs that have a strong “evaluation culture” prioritize and invest in both high quality evaluation and their utilization.

“A critical point for utilization of the evaluations is the quality of evaluations and increasing it. We have identified, and PPL has as well, that there are deficiencies within even the infrastructure for evaluation and the resources available. The evaluation policy has done a lot to make how we define an evaluation stricter, what we are not supposed to select for the evaluation team. It is getting better but it is still not where it needs to be. We need to strengthen the availability of good organizations that can deliver quality evaluations. There just aren't enough organizations that are out there yet to do good high-quality evaluations because it was sort of pushed aside for a couple of decades. We run into this problem all the time so we end up getting less than ideal or sub-par evaluations....”

–Technical Bureau Staff Member

Capacity of Organizations that Provide Evaluation Services to USAID

A related problem, cited by eight OUs, was the capacity of the firms procured to provide evaluation services. One Mission called the problem “weak bench strength,” meaning that some of the organizations have too few highly qualified evaluators on which to call. Two other OUs cited the capacity of the small firms through which USAID procures evaluation services, saying that while they occasionally field strong teams, they seem to lack the capacity to assemble teams that include a strong evaluator, as well as highly qualified local team members, on a consistent basis.

USAID Staff Suggestions

- Institute a more rigorous evaluator selection process so only qualified evaluators are hired (1 OU)
- Train the current pool of evaluators for USAID on USAID expectations (2 OUs)
- Supervise evaluator quality (1 OU)
- Keep a record of quality evaluators in a database (1 OU)
- Have an internal M&E expert review evaluator bids (1 OU)

Evaluation Recommendations

Recommendations, while part of an evaluation report, play a critical role in transforming the lessons of evaluations into meaningful action. Survey data show that all but 2 of the 118 evaluations examined included recommendations, and that USAID adopted at least some recommendations from all but 10 percent of them. For those recommendations that USAID rejects, the reasons are similar to those found in a study that the World Bank conducted of its own evaluation work in 2008.²⁷ In that study, the World Bank reported that the quality of its recommendations were a significant impediment to the utilization, as many of them were too vague and impractical to be adopted by managers.

Survey findings on common reasons for not accepting and acting on recommendations are listed in a text box on this page. Following up on these survey findings, the study team examined recommendations in 45 of the evaluations surveyed. A key finding was that the structure, content, quality, and number of recommendations are highly varied across evaluations, with the number of

²⁷ World Bank. 2008a. *Annual Report 2008*. Washington, D.C.: World Bank.

<http://siteresources.worldbank.org/EXTANNREP2K8/Resources/YR00 Year in Review English.pdf>

recommendations ranging from zero to seventy, with an average of eighteen formal recommendations. Recommendations in this sub-set did not always meet USAID quality expectations for being specific, action-oriented, practical, and directed to specific parties. Survey data concurred, indicating among 18 evaluations surveyed, respondents found that recommendations in 65 percent of them were action-oriented, 61 percent were practical or feasible, 58 percent were clear about who should act on them, and 55 percent were well supported by evidence, while only 16 percent of these evaluations were coded as having recommendations that were cost-conscious.

Reasons Why Recommendations Were Not Adopted for Action

- Change in context (26%)
- Not timely—no longer relevant (20%)
- Too costly (18%)
- Not action-oriented (16%)
- Insufficient supporting evidence (15%)
- Too vague (11%)
- Too complicated (10%)
- Too many recommendations (8%)

In addition to probing characteristics of evaluation recommendations that are not being accepted, the survey asked whether accepted recommendations are always fully implemented, and if not, why not. Responses revealed that regardless of whether half or all of an evaluation’s recommendations were accepted, those that were fully implemented ranged from 9 percent to 30 percent. Reasons cited for failure to implement accepted recommendations include their lack of relevance, insufficient funds, inappropriate timing, lack of a champion, and implementation capacity gaps. Some of these are the same reasons that recommendations in other evaluations,

perhaps in other Missions, were not accepted in the first place. Notably, only a small number of surveys and interviews identified the absence of “political will,” which respondents described as being either insufficient interest to ensure that evaluation recommendations are acted upon, or resistance to particular recommendations by USAID or one of its partners.

USAID Staff Suggestions

- Work with USAID’s pool of evaluators to give them a better understanding of USAID’s context, so that their recommendations are more realistic (2 OUs)
- Evaluators should be required to include a section in the recommendations that gives recommendations to the country partner government (1 OU)

Political Will/Evaluation Culture

Some survey responses and interviews identified “political will,” or broad evaluation culture considerations, as occasionally constraining evaluation utilization. In this vein, 18 percent of respondents indicated that there was insufficient interest or political will to pursue action recommended in an evaluation, and for 5 percent of evaluations where recommendations had been accepted, respondents indicated that USAID or its partners resisted implementing them. This issue emerged in some form in 10 of the 24 OU group interviews, with interview participants identifying three instances that involved political will with USAID, and five instances that involved resistance among USAID partners.

Human and Financial Resources

In group interviews, three OUs indicated that they needed more M&E staff to effectively follow up on evaluation utilization. Five OUs identified funding as a constraint on evaluation in their units, including for the resources needed to more proactively engage in evaluation utilization. Smaller Missions were particularly concerned with monetary resources, and one said that their current evaluation budget was too small to allow them to use USAID’s evaluation IDIQs given the relatively high cost of services provided through that mechanism.

USAID Staff Suggestions

- OUs don't all have sufficient financial resources to conduct evaluations of all of the required "large" projects, so the number of required evaluations should be decreased (1 OU)
- One central Bureau suggested creating a centrally located pool of additional resources that OUs could draw on when they needed extra funding for evaluations (1 Bureau level OU interview)
- Send M&E staff on TDYs to assist with evaluations; have technical staff in Washington carry out statistical analyses using field data; or encourage peer-to-peer exchanges in regions for these purposes (2 OU)
- PPL staff should be more directly involved with field M&E staff by mentoring/coaching in Missions and participating on evaluation teams (1 OU)

Presence of Factors that Appear to Affect Evaluation Utilization

Figure 13 shows factors that either the literature or study findings indicate are associated with evaluation utilization. They are ranked in terms of how frequently they were found to have occurred in relation to the USAID evaluations in the survey sample.

In Figure 13, factors which either the literature or study findings indicate are associated with evaluation utilization as documented for USAID by this study. They are rank ordered in terms of how frequently they were found to have occurred in the relation to the USAID evaluations in the survey sample. Moving from left to right, the frequency of occurrence increases, with "dissemination of evaluations to USAID staff" appearing in the furthest right column as the most frequently present factor.

Factors are further organized into four clusters based on rankings, i.e., factor found in more than 75% of the 118 evaluations covered by the study survey, 50% to 74% of evaluations, 25% to 49% of evaluations, and less than 25% of evaluations. To distinguish between performance clusters in the figure, colors are used to identify where each factor lies, for example those factors in blue are present for more than 75% of the 118 evaluations examined through the survey.

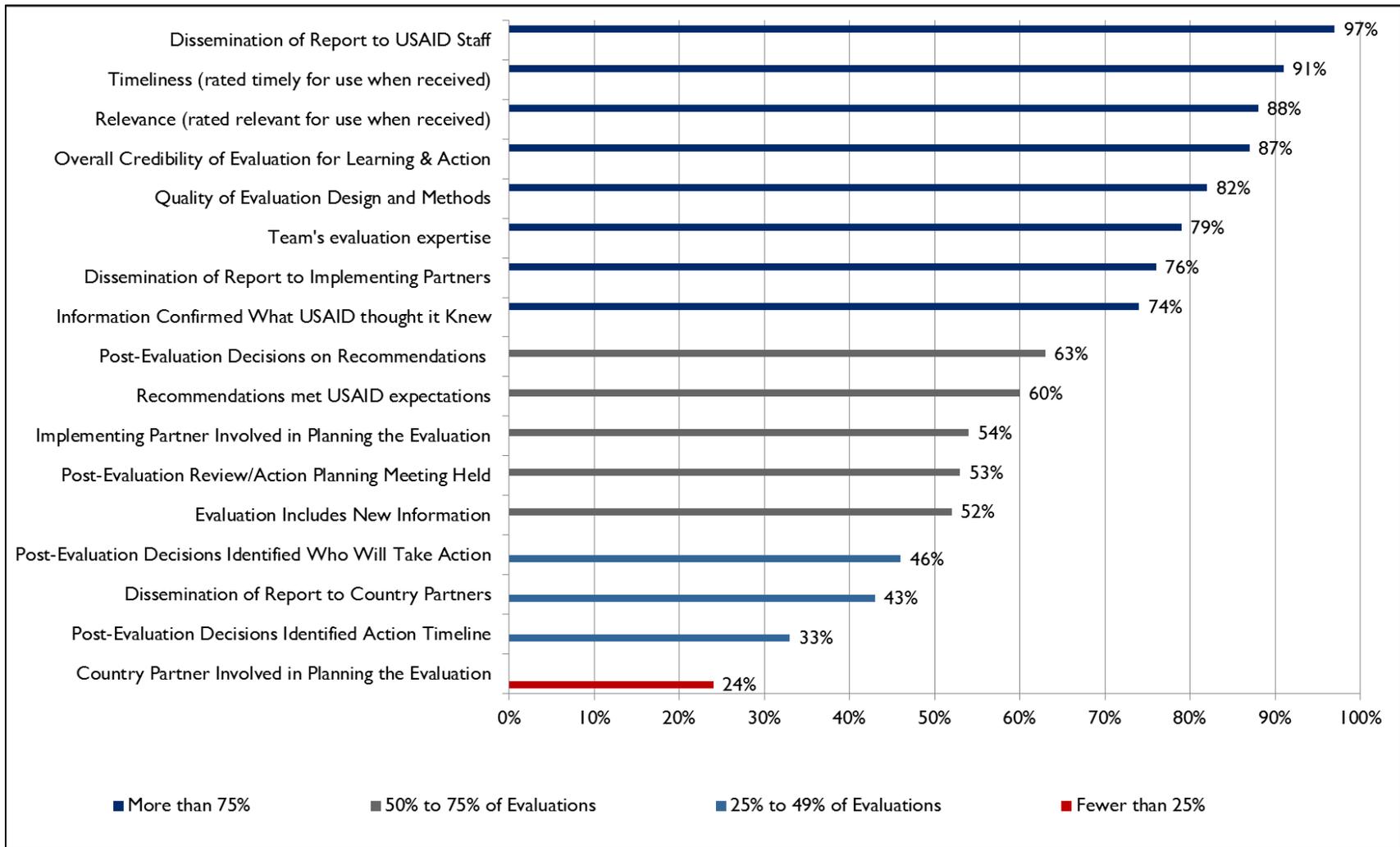


Figure 13. Presence of Factors that may Foster or Impede Evaluation Utilization in USAID (N=118 USAID Evaluations for which Survey Responses Exist)

CONCLUSIONS

USAID guidance, over the decades, has always encouraged evaluation utilization to some degree, but never as specifically as in recent years through the 2011 Evaluation Policy and guidance associated with the Program Cycle. The Evaluation Policy explicitly states that evaluations are not an end in themselves, but rather their value is established by their use, not their existence or even their quality. Program Cycle-related guidance introduced new requirements for conducting evaluations, drawing upon their evidence, and citing them when developing CDCSs and PADs.

While these changes indicated the heightened importance of learning from experience and evidence based decision making, USAID did not establish specific targets for any aspect of evaluation utilization, nor was a baseline created for comparing progress over time. Accordingly, this study focused on establishing the status of USAID's utilization of its evaluations and on identifying impediments and opportunities for improvement. It does not make retrospective comparisons or judgments. In that light, based on findings, the broad conclusions reached by the study team are that:

- USAID evaluation utilization practices are already strong and compare well to those of other USG agencies examined in parallel studies conducted by the GAO. USAID's broad claims of 90 percent use for learning purposes and to support decisions and related actions are higher than the 80 percent range of equivalent self-reported use in other agencies.²⁸
- USAID's utilization of evaluation results is more prevalent in the planning and implementation stage of its Program Cycle than in other stages. This is not unexpected given that most of the Agency's work is focused on designing and implementing the projects and activities through which foreign assistance is delivered.
- Responses indicating an overall high rate of evaluation use at the Agency somewhat mask the fact that there has been only partial uptake by OUs of USAID business processes designed to foster evaluation utilization. Several examples from the study findings illustrate this point:
 - Through structured reviews recommended in USAID's ADS, USAID made decisions about which evaluation recommendations to accept/reject for 63 percent of sampled evaluations; identified who should be responsible for implementing "accepted" recommendations for 42 percent of the evaluations; and established timelines for their completion for 33 percent.
 - Pursuant to USAID guidance that calls for broadly sharing evaluation results, USAID disseminated copies of its evaluation reports to its own staff for 97 percent of sampled evaluations; to IPs for 76 percent of those evaluations; and to country partners for 43 percent.
 - USAID's involved its IPs in evaluation planning in 54 percent of sampled evaluations and country partners in 24 percent of them, in relation to guidance that encourages this participation and recommends "linking evaluation questions to specific future decisions made by USAID leadership, partner governments, and/or other key stakeholders."
 - In line with USAID guidance, evaluation evidence was cited and used to support DOs and strategies proposed in 59 percent of USAID CDCSs approved to date.
- USAID staff are more knowledgeable about the ways in which evaluation results have been used and the effects of that use than statements the PPR's Evaluation Registry suggest. Survey responses and stories told in interviews demonstrate the range of this knowledge, and compare favorably to the 12 percent of PPRs that claim some form of utilization. This knowledge is valuable for explaining USAID's achievements.

²⁸ Op, cit. GAO-13-570 (2013).

- Some, but not all, of the factors examined in this study were found to be linked, either statistically or through interview content analysis, with some type of evaluation utilization or its effects. Table 5 below shows which factors—initially identified in Table 1 and discussed in Question 4—were found to be associated with evaluation use and others for which no compelling evidence of such a relationship was found, even if one does actually exist.

Table 5. Evaluation Factors for Which a Linkage to Evaluation Use Was/Was Not Found

Appear to be Linked to Evaluation Utilization²⁹	No Direct Evidence of Link to Utilization Found
Post-evaluation dissemination*	Participation of stakeholders in the evaluation
Post-evaluation review, action planning, & tracking*	Credibility and relevant knowledge of evaluators
Timeliness (in time for decisions; on schedule)*	Need for, or purpose of, the evaluation
Quality (methods, data, analysis)* ³⁰	Country or regional context
Quality (overall credibility for learning & action)*	Sponsor organization’s quality control practices
Recommendations (specific, actionable, practical)*	Commissioning organization evaluation culture, including the political will to act on recommendations (on which a few USAID staff survey responses and interview comments remarked)
Presence of new information+	
Relevance to client; client ability to act on results (which USAID staff indicated is tied to timeliness)^	

The best opportunities for improving evaluation utilization at USAID are to ensure at least one of the factors or processes depicted in the left column of Table 5 are present for any given evaluation. Prior to determining which factors to focus on, however, the study team calls attention to Figure 13 and the current status of USAID’s evaluation practices, particularly for each of the factors in the left hand column. In that figure the factors are disaggregated, where possible, by stakeholder groups including USAID staff, IPs, and country partners. What that figure shows, for example is that USAID already does very well at disseminating to its staff, but not very well at disseminating to country partners. These distinctions are important for determining exactly what types of investments are needed to improve the status of factors affecting utilization. Factors where there is already a high frequency of occurrence may not produce the same level of results as factors with lower frequencies.

The study team also looked at opportunities to improve evaluation utilization within the USAID structure. Study data suggest that the “front line” role of OUs, as commissioners and primary users of evaluations, makes them the most logical actors to initiate a variety of actions that would likely increase evaluation utilization and enhance its effects on development outcomes. The study team also concluded that Technical and Regional Bureaus in Washington, as well as PPL/LER, can play important supporting roles to help ensure the success of such efforts initiated in OUs.

²⁹ The symbol* indicates a statistically significant relationship. The symbol + indicates statistical evidence of an association but at a slightly lower than significant level. The ^ symbol indicates evidence of an association based on content analyses of interview data.

³⁰ Findings about the relationship between evaluation quality (subsuming three factors: methods, recommendations, and overall credibility) were found at the OU level rather than the individual evaluation level. This indicates that improvements in the average evaluation report score is what is linked to evaluation use, even if individual evaluations vary from that average. OUs with higher average evaluation report scores also demonstrated greater evaluation utilization in survey responses.

RECOMMENDATIONS

The study team sees a need for broad-based Agency participation in any effort to improve evaluation utilization, particularly given the front line role played by USAID Missions and Washington Offices in commissioning and using evaluations and the important supporting roles of Regional Bureaus, Technical Bureaus, and PPL/LER. For this reason, recommendations are organized by, and addressed to, specific USAID units the study team feels is best suited to take action. Such actions are intended to translate the results of this study into enhanced evaluation utilization by USAID and its partners, resulting in improved program performance and positive development outcomes.

It is recommended that USAID Operating Units (Field Missions and Washington Offices that directly administer the delivery of foreign assistance through projects, activities, or other mechanisms):

1. Conduct a brief collaborative internal assessment, next quarter, of the current level of evaluation use within the OU. Consider study findings about USAID business practices and evaluation characteristics (under Study Question 4) that may be affecting the OU's evaluation utilization. Information reported on the Program Cycle (under Study Question 1), along with Figure 13, may be helpful to the OU's M&E POC and others involved in guiding this assessment.

2. Create and obtain OU management approval for an OU-specific plan for enhancing OU partner and staff utilization of USAID evaluations. The plan should include specific actions focused on OU business practices that the study suggests warrant attention, since their absence may limit awareness or utilization of evaluation results. Identify simple ways in which the OU can update this assessment in future years, including encouragement to staff and partners to highlight effects of evaluation utilization in stories and other communications they share with OU management. Identify needs for assistance from relevant Washington Offices and request their support as warranted. Integrate progress reports of the OU's Action Plan for improving evaluation utilization into the OU's Fiscal Year 2016 and future year portfolio reviews.

It is recommended that USAID Technical Bureaus:

3. Inventory what compendiums or meta-analyses already exist for USAID evaluations in technical fields that Bureau supports. Ensure that existing compendiums of evaluation results are available in OUs and to other USAID staff working in technical areas the Bureau supports.

4. Collaborate with PPL/LER to develop more comprehensive and improved approaches for ensuring that evidence from future USAID evaluations are widely available to Agency staff. Work with Regional Bureaus to ensure that whatever evaluation evidence products are produced accurately reflect geographic considerations.

It is recommended that USAID Regional Bureaus:

5. Review with their Missions, and Washington Offices delivering foreign assistance in their regions, the adequacy of their M&E staffing and budget resources relative to their needs and evaluation commitments made in CDCSs, PADs, and PPRs. Such reviews should ensure that no Mission or Office is unduly disadvantaged in its ability to have and utilize evaluation results based on its size, geographic location, or the nature of its portfolio.

6. Further develop, implement, and share information Agency-wide about regionally appropriate mechanisms for collaboratively drawing on existing M&E resources. Arrangements should build on productive experiences such as the E&E Bureau peer exchanges—which help Missions in need expand their capacity to take on a short-term evaluation tasks—or the Regional Mission evaluation hub model

that is emerging in the Asia Bureau, through which evaluation services are provided to smaller Missions and Country Offices.

It is recommended that PPL/LER:

7. Minimally refine official Agency guidance, in line with findings from this study, to draw attention to and enhance evaluation utilization. Exhibit I includes a list of illustrative modifications to this end.

8. Continue to develop “How-To” guides and “Technical Notes,” or expand its Evaluation Toolkit, in response to expressed needs. Consider USAID staff suggestions with respect to the development of new “How-To” guidance on (a) writing good evaluation questions, (b) dissemination planning, and (c) well-constructed evaluation recommendations. Incorporate into this last topic a suggested range for number of evaluation recommendations, akin to the Agency’s recommendation on the number of evaluation questions. Identify OUs that are already tracking the status of commitments to implement “accepted” evaluation recommendations, then work with them to develop short case studies on their experiences; these case studies will encourage implementation of the practice more broadly at that OU and can be used to introduce the practice to other OUs wishing to begin the process.

9. Collaborate with Technical and Regional Bureaus to make findings and evidence from USAID evaluations more readily available to USAID staff charged with responding to ADS 201 requirements. In this regard, play a leading role in an assessment of options that not only increase access to evaluation resources, but make it easier to quickly identify relevant evaluations. To that end, examine options listed below that are used in other development assistance agencies that might, over time, be adopted and supersede USAID’s current reliance on full evaluation reports as the primary locus of information about what findings and evidence might be useful for future programming:

- **Evaluation Briefs (1-2 pages)**, or short, standalone summaries, prepared by evaluation teams as they prepare their reports. This type of quick reference on evaluation findings is routinely produced by at least half a dozen OECD/DAC member country development agencies as well as by UNICEF and other U.N. operating agencies. They are well suited for broad in-person and online dissemination.
- **Evaluation Abstracts (300-500 words)**, which can also be prepared by evaluation teams as they complete their reports. For years these were required by USAID as part of the evaluation transmission package from the field to Washington, and in some years were assembled and issued as compendiums. Recently, USAID’s Office of Education improved upon the compendium of abstracts approach by adding to it a coding process that made individual abstracts searchable based on methodology, as a proxy for evidence strength, as well as topic.³¹
- **Evaluation Syntheses** take a number of forms, all of which involve a review of a set of evaluations by topic, region, or time period. USAID has created this type of document for OU use periodically over the years, but on an ad hoc basis. For example, USAID’s Office of Conflict Management and Mitigation (CMM) commissioned Social Impact to prepare a synthesis focused on their grant evaluations under the CMM People-to-People Reconciliation Fund, called an Evaluative Learning Review, which used meta-analysis approaches to extract key findings on a country-specific as well as topical basis.³²

³¹ Mobiles for Education Evaluation Abstracts

http://www.meducationalliance.org/sites/default/files/meducation_evaluation_abstracts_6-30-15.pdf

³² Evaluative Learning Review Synthesis Report: *USAID/CMM’s People-to-People Reconciliation Fund, Annual Program Statement (APS)*. Arlington, VA: Social Impact, 2014.

10. Improve USAID's ongoing monitoring of evaluation utilization through existing reporting mechanisms if possible, or create a new reporting mechanism if existing ones cannot be improved. More specifically:

- Consider ways in which the PPR reporting mechanism could be modified to permit more comprehensive reporting on evaluation utilization, perhaps by opening an avenue through which additional information could be added in subsequent years after evaluations were concluded. Additionally, develop a systematic process for retrieving, analyzing, and reporting back to the Agency on evaluation utilization reporting in the PPR. This process will add credibility that staff indicated it needs to incentivize more serious utilization reporting by Agency staff.
- Consider adopting as a USAID “best practice,” and possibly incorporating into policy, the use of post-evaluation Action Plan tracking systems such as those that this study identified for roughly half a dozen OUs (and possibly more). Encourage OUs to quantify their performance in this regard by calculating the percent fully implemented across all evaluations as an OU performance measure and, after an appropriate pilot period, consider aggregating data from a new OU performance measure of this sort as a reportable Agency-level indicator for evaluation utilization, to complement stories of evaluation use it collects through the PPR and by other means.

Exhibit I. Illustrative Suggested Modifications of Existing USAID Guidance

The study team for this project recognizes that USAID’s ADS guidance has been under review throughout 2015, and that the next version may be very different than its predecessor. As a result, this exhibit may touch on changes USAID has already made. Nevertheless, the suggestions included below may provide insight into ways that various USAID guidance products could enhance evaluation utilization, even beyond the ADS. While the ADS is specifically mentioned in some of the suggestions below, it is not the only mechanism for communicating ideas that will improve evaluation utilization in the Agency. Other guidance, including How-To Notes, Technical Notes, and guidance on Mission Orders all have a role to play.

Suggestions in this exhibit highlight where new or additional language in USAID’s suite of evaluation guidance materials, as of December 2015, might help the Agency operationalize concepts and approaches envisioned in USAID’s Evaluation Policy, ADS, and other documents. The language in the suggestions below is worded in a manner consistent with USAID’s ADS style. In using this style, it is not the study team’s intent to overreach, but rather to illustrate the kinds of modest changes USAID might consider making, consistent with the findings, conclusions, and recommendations provided in this study.

Guidance Document	ADS 203.3.1.4 - Planning Evaluations
Possible Modification	ADS 203.3.1.4 currently includes six elements related to planning evaluations. A seventh element introducing evaluation dissemination plans could be added.
Rationale	<ul style="list-style-type: none"> Dissemination plans are already called for in USAID’s Evaluation Policy (page 9) but are not yet mentioned in the ADS or the Standard Mission Order on Evaluation. Draft dissemination plans can help guide the development of deliverables lists in evaluation SOWs. As such, the proposed language on dissemination plans would appropriately fit into ADS 203.3.1.4 prior to the line stating, “These plans will be used to inform evaluation statements of work.”
Possible Language	<u>Prepare a draft evaluation dissemination plan identifying likely audiences and evaluation products that an evaluation team would be expected to produce beyond an evaluation report, e.g. translation of the executive summary, two-page executive briefer, etc.</u>

Guidance Document	ADS 203.3.1.8 - Documenting Evaluations
Possible Modification	ADS 203.3.1.8 currently includes eight criteria that an evaluation report must meet. A ninth element, the inclusion of an evaluation abstract, could be added.
Rationale	<ul style="list-style-type: none"> ADS 201 requires staff to use evaluations when preparing CDCSs and PADs. In interviews for this study, staff expressed interest in easier ways to access information from evaluations, indicating that searching the DEC was too difficult and labor-intensive. Evaluation abstracts—which were required as part of USAID’s evaluation transmission forms until the mid-1990s—concisely provide information from evaluations and can be used to aid staff in locating appropriate evaluations. Historically, these abstracts were compiled and circulated back to Missions. Other donors require similar abstracts, including New Zealand’s NZAID requiring an abstract of no more than 250 words, and DFID, which requires a two-page briefer.
Possible Language	<u>Evaluations should include, on a separate page at the front, an abstract of not more than 250 words briefly describing what was evaluated, questions addressed, specific methods used, key findings, and recommendations. These can be posted or circulated separately, with a link to the full report.</u>

Guidance Document	ADS 203.3.12 - Mission Portfolio Reviews
Possible Modification	Add language to the ADS section on Portfolio Reviews on follow-up actions to be taken based on evaluation findings or recommendations that matches guidance in USAID's Standard Mission Order on Evaluation. Place this new language at the end of the existing list of bullets in this section of the ADS
Rationale	Similar language to what is proposed below is included in the Standard Mission Order on Evaluation. This addition to the ADS would make these two guidance documents more similar.
Possible Language	<u>During Portfolio Reviews, the status of Action Plans for evaluation findings and their use in respective decisions will be discussed and documented.</u>

Guidance Document	ADS 203.3.1.9 - Responding to Evaluation Findings
Possible Modification	This section of the ADS provides six steps USAID should follow when responding to evaluation findings. The wording of Steps 3 & 4 should be modified to provide greater clarity and guidance. An additional step, Step 7, should be added to encourage the capturing of evaluation utilization experiences and effects.
Rationale	<ul style="list-style-type: none"> In Step 3 the term "team" should be replaced with "USAID," as team has been misunderstood in EES trainings to mean the evaluation team. In Step 4 language should be added regarding the monitoring of post-evaluation Action Plans over time. This would be consistent with what eight Missions are already doing to track accepted recommendations and is in line with the sample trackers provided in USAID's Evaluation Toolkit Step 7, on capturing evaluation utilization experiences and effects, should be added to reinforce existing guidance in the PPR Evaluation Registry and other mechanisms.
Possible Language	<p>(3) Determine whether <u>USAID</u> accepts/supports...</p> <p>(4) Identify any management or program actions needed and assign responsibility and the timelines for completion of each set of actions; <u>monitor their implementation;</u></p> <p>(7) <u>Document known/verifiable stories of the effects of evaluation utilization on program performance and development outcomes using PPRs, CBJs, and other existing evaluation reporting/planning mechanisms.</u></p>

Guidance Document	Standard Mission Order on Evaluation
Possible Modification	In the section titled <i>Sharing Evaluations with Stakeholders</i> , consider incorporating more specific language on the specific audiences with whom findings should be shared.
Rationale	Language in the Standard Mission Order on Evaluation is not as clear and specific as language in the ADS and USAID's Evaluation Policy regarding the sharing of evaluation results with partners and other stakeholders. This clarification will also remedy some uncertainty indicated by staff in interviews regarding whether dissemination of evaluations is meant to be only within USAID or to partners as well.
Possible Language	Findings from evaluations will be shared as widely as possible <u>with USAID implementing partners, country partners, and other stakeholders</u> , with a commitment to full and active disclosure...

Guidance Document	Standard Mission Order on Evaluation
Possible Modification	In the section titled <i>Responding to Evaluation Findings</i> , consider including language on monitoring the implementation of commitments made during post-evaluation action planning meetings.
Rationale	Eight Missions have already begun implementing tracking tools for this purpose, and USAID has similarly added such tools to its Evaluation Toolkit. Adding the proposed language would align the Standard Mission Order on Evaluation with existing activities as well as the proposed modification to ADS 203.3.1.9 mentioned above.
Possible Language	Follow up on the status of actions the Operating Unit decided to take based on evaluation findings or recommendations to ensure that they are implemented.

Guidance Document	USAID How To Note on Preparing Evaluation Reports
Possible Modification	In Table 1, page 2 of this guidance document, on the row that discusses quality control, encourage USAID staff and evaluators to use the Evaluation Report Review Checklist from the PPL/LER <i>Meta-Evaluation of Quality and Coverage of USAID Evaluations 2009-2012</i> , conducted by MSI, together with the Rater's Guide, and Scoring System instructions to self-check their evaluations before finalizing them as a routine quality control process. Include in the How-To's additional resource list, p. 7, references to these three quality control aids.
Rationale	Findings from this study of evaluation utilization identified the average evaluation quality scores of Operating Units as one of the factors that influences evaluation utilization. This echoes PPL/LER's Meta-Evaluation findings and recommendation on improving staff access to evaluation report quality control aids ("Recommendation 2. Intervene with appropriate guidance, tools, and self-training materials to dramatically increase the effectiveness of existing USAID evaluation management and quality control processes"). The Meta-Evaluation checklist, rater's guide and scoring system were replicated in 2015 when the E3 Bureau used them for its 2013-2014 Sector Synthesis of Evaluation Findings. The study team's suggestion is to highlight and make these quality control aids more widely accessible through a minor modification to USAID's How-To Note: Preparing Evaluation Reports and encourage their use by evaluation teams and evaluation peer-reviewers in USAID.
Possible Language	p.2, add a sentence to the existing text: Assess reports for quality by including an in-house peer technical review with comments provided to evaluation teams. Encourage evaluation teams and peer reviewers to familiarize themselves with USAID evaluation report quality control aids, such as the evaluation report quality checklist, rater's guide, and the scoring system used in the Agency's 2009-2012 meta-evaluation and a subsequent E3 Bureau evaluation synthesis. p.7, add an additional bulleted reference: USAID Meta-Evaluation of Quality and Coverage of USAID Evaluations 2009-2012, Management Systems International (MSI) 2013: https://pdf.usaid.gov/pdf_docs/PDACX771.pdf (2013) and the Sectoral Synthesis of 2014-14 Evaluation Findings, Bureau for Economic Growth, Education and Environment https://www.usaid.gov/sites/default/files/documents/1865/E3_Sectoral_Synthesis_Report.pdf (2015). These studies used a common set of evaluation quality control aids, which are most succinctly documented in the second of these two volumes: (a) an Evaluation Report Quality Review Checklist and linked scoring system, pp. 158-160 of the Sectoral Synthesis report and (b) a companion Evaluation Report Quality Review Rater's Guide with detailed notes on rating factors, pp. 165-171 of the Sectoral Synthesis.

ANNEX I – EVALUATION UTILIZATION STUDY

STATEMENT OF WORK

Background

USAID has been placing increasing emphasis on the quality of evaluations and evaluation processes at the Agency over recent years, largely starting with an Evaluation Policy released in early 2011. This policy has been complemented by numerous How-To documents and other guidance for USAID staff on best practices in evaluation. The Office of Learning, Evaluation, and Research (LER) within the Bureau for Policy, Planning, and Learning (PPL) has further set forth an agenda for conducting studies to better understand evaluation at USAID, including a recent 2013 study looking at the quality of evaluation reports produced by and for the Agency. In line with that agenda, USAID is looking to conduct another study which will investigate the extent to which USAID evaluations are being used by staff and partners and how those evaluations are being used. Information from the study will inform USAID's policies and practices moving forward.

Existing Information Sources

It is expected that the evaluation utilization study team will conduct a thorough literature review so as to be aware of current state of thinking in regards to evaluation in the academic and development communities. The team should also be aware of previous evaluation-focused studies conducted by LER, such as the 2013 meta-evaluation looking at the quality of USAID evaluation reports and available on the DEC. LER will make available to the study team internal documents relating to evaluation utilization as requested, such as policy documents, PPRs, Mission Orders, or any other documents deemed relevant.

Study Purpose, Audience, and Intended Uses

The purpose of this activity is to learn how USAID evaluations are being used and how application of evaluation findings may be improved through the following:

1. Learning about the utilization of USAID evaluations
2. Identifying to what extent, how, and when evaluations are being used
3. Identifying what changes are caused, prompted, or allowed by evaluation findings
4. Examining incentives, barriers, and business processes that affect the utilization of evaluation findings, conclusions, and recommendations

The primary audience for the study is USAID/PPL/LER for use in improving systems for evaluation use and learning as well as how to set up potential systems for monitoring the use of evaluation within USAID.

Other relevant stakeholders are monitoring and evaluation (M&E) points of contact (POCs) around the Agency, both within Washington Offices and Missions overseas. In addition, there is a broader audience for this information including implementing partners, Congress, the Office of Management and Budget (OMB), Agency leadership, technical officers, and other staff whose work is (or should be) informed by evaluations.

Study Questions

1. How and when in the Program Cycle are evaluations used?
2. What changes/decisions are made because of evaluations?
3. To what degree and under what conditions does learning occur from evaluation findings that was not anticipated by the intended purpose of the evaluation, or when an evaluation is

described to the study team as being undertaken primarily because it was understood to be “required”?

4. What particular business processes or enabling conditions appear to increase the utilization of evaluations?

Gender Disaggregation and Gender Differential Effects

While the nature of this study is not focused on person-level data, and therefore gender disaggregation would likely be neither appropriate nor possible, the study team should pay particular attention to any instances where gender-related issues arise in the course of the study.

Evaluation Design/Methods

A collaborative, participatory approach is desired in the design and execution of the Evaluation Utilization study. The study will take place in three phases.

The goal of Phase I of the evaluation is for the contractor to develop a solid understanding of the evaluation process within USAID and consult with relevant stakeholders to design the evaluation. Phase I may include but not be limited to background research, document reviews, questionnaires, and initial data gathering. This work will include convening a group of professional evaluators to provide input and ideas regarding the best ways to define and assess evaluation utilization within USAID.

Phase 2 will consist of conducting the evaluation, analyzing data, and presenting actionable recommendations for PPL/LER and the Agency. Data collection may include but is not limited to surveys, key informant interviews, and focus groups, among others.

Phase 3 will include dissemination of the report and presentations to PPL/LER, interested USAID staff, and outside stakeholders on the evaluation results and a proposed process for implementing the recommendations.

Data Analysis

Data analysis methods are to be proposed by the evaluation utilization study team and approved by USAID. Proposed analysis methods should be of the highest reasonable level of rigor and appropriate to the associated data collection methods outlined in the design document.

Evaluation Design Strengths and Limitations

The evaluation utilization study team will be responsible for clearly articulating the methodological strengths and weaknesses of the proposed data collection and analysis processes.

Evaluation Deliverables

The main deliverable of Phase I will be the evaluation design and Phase I report. The design will provide a detailed plan to gather and assess data and provide an outline for the final report that will propose potential templates or other similar tools that USAID can use, and include a plan for supporting USAID to implement recommendations.

Phase II will culminate in a final report providing clear findings and recommendations on how LER should move forward to improve the utilization of evaluations at USAID. This report should include tools and templates to be shared throughout the Agency as well as a dissemination plan for how study results should be shared within USAID as well as with partners and other donors.

Phase III deliverables will include dissemination of the report and presentations to PPL/LER, interested USAID staff, and outside stakeholders on the evaluation results and a proposed process for implementing the recommendations.

The study team will provide LER with electronic versions of study data, code books, transcripts and any other relevant materials upon request in formats compatible with Microsoft Office.

Team Composition

The evaluation utilization study team shall have experience working with a federal government agency and an international organization in the following:

- Conducting evaluations of international development projects or efforts
- Designing and implementing a methodology for assessing utilization of evaluation findings
- Assisting in creating learning approaches for development organizations
- Providing recommendations and designing systems to promote and increase utilization of evaluation results and/or utilization of new information
- Supporting implementation of an effective knowledge and learning cycle within a complex organization
- Providing expertise and contributing to the state of the art in the international evaluation field and to advancing knowledge on how evaluation contributes to effective international development programs
- Previous experience working in collaboration with the contracting agency during project design

Phase I will require the efforts of at least two people. The team leader will be an Evaluation Methods Specialist, with at least a Master's degree (a Ph.D. is preferred) in international development or a related field such as evaluation, economics, sociology, anthropology, or public administration. Expertise in evaluation methods, including the appropriate use of quantitative, qualitative and mixed methods is required. At least ten (10) years of professional experience in international development is required including at least five (5) years of work designing, executing, and reviewing evaluations or monitoring systems of development programs. The team leader specified above is considered a key personnel and essential to the work being performed. The team leader should be available to lead all three phases of this study. The Specialist (team leader) demonstrates a track record in building the capacity of client and partner staff in performance monitoring and evaluation planning, management, and utilization. The Specialist must also have facilitation skills both online and in-person and excellent communication skills.

The second member of the study team will be an Evaluation Analyst with a Bachelor's degree or greater in a relevant field, with a minimum of five (5) years of experience working in measurement and evaluation for USAID or other donor-funded development programs.

Participation of USAID Staff and Partners

USAID staff and partners will not be directly involved in data collection or analysis, but will be collaboratively involved in discussions regarding the design of the study as well as in the creation of recommendations once data have been analyzed and findings have been presented.

Scheduling and Logistics

The study team will be responsible for all logistical aspects involved in conducting the study, however USAID staff will assist in facilitating communications within the Agency and with some partner organizations to increase likelihood of cooperation and response rates.

All three phases of the study must be completed no later than June 30, 2015 to ensure USAID/PPL/LER is provided the necessary information in a timely enough manner to take action on key management decisions. It is anticipated that Phase I will be completed no later than January 15, 2015 with Phase II coming to an end by May 31, 2015, with one month of dissemination for Phase III in June 2015.³³

Reporting Requirements

The study team is required to submit a final report to LER which meets all USAID guidance for evaluation reports as outlined in the ADS, How-To Notes, and the Evaluation Policy, particularly the following criteria to ensure the quality of the report:

- 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 scope of work.
- The evaluation report should include the scope of work as an annex. All modifications to the scope of work, whether in technical requirements, evaluation questions, evaluation team composition, methodology or timeline need to be agreed upon in writing by the technical officer.
- Evaluation methodology shall be explained in detail and 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.

³³ The end date of the three phases was changed to September 30, 2015 via contract modification No. 3, effective September 30, 2015. The timing of phases 2 and 3 was changed accordingly: Phase 2 completed by July 31, 2015, and Phase 3 completed by November 30, 2015.

ANNEX 2 – STUDY METHODOLOGY

Study Questions

Four study questions guided the work of the study team:

1. How and when in the Program Cycle are evaluations used or not used?
2. What changes/decisions are made because of evaluations?
3. To what degree and under what conditions does learning occur from evaluation findings that were not anticipated by the intended purpose of the evaluation?
4. What particular business processes or enabling conditions appear to encourage or discourage the utilization of evaluations?

This annex explains how the study team developed and implemented its approach to answering these questions.

Literature Review

Work on this study began with a review of the literature on evaluation utilization which was undertaken to identify ways in which this topic had been examined previously and to learn what was already known from other research efforts about the types of business processes and enabling conditions that seem to encourage or deter the utilization of evaluations.

Published literature on evaluation influence and utilization is well documented in several broad reviews of conceptual models and research (Cousins (1986), Johnson (2009) and Herbert, (2014). The most recent of these reviews divides research on evaluation use and influence into three periods. The first period (1970-1986) introduced an enduring taxonomy that distinguishes decisions or actions based on evaluation findings and recommendations, from learning, or enlightenment, which affects perceptions and attitudes, but does not lead to discernible changes in the programs and projects examined.³⁴ Studies that built on this early work expanded the time horizon over which decisions and actions might occur, and added of symbolic uses of evaluations to the utilization range, including evaluations conducted primarily to demonstrate an organizational commitment to accountability, evaluations undertaken to establish legitimacy for decisions already taken.

A second period (1986-2000) broadened thinking about evaluation utilization by focusing as much on the evaluation process as on the products it produces, namely, a report and associated briefings. Studies during this period also examined the organization and societal contexts in which evaluations are undertaken. On the process side, ideas about utilization expanded to examine learning that occurs through participation in the evaluation process, including by beneficiaries, government officials, funding organizations, and implementing partners, and opportunities for enhancing evaluation utilization by focusing on evaluation utilization at the start of the process when evaluation questions are formulated and team composition are considered (Patton, 2008). Towards the start of this second period, USAID conducted its first and only other review of evaluation utilization, in the Agency.³⁵ This initial study of the utilization of USAID evaluations indicated that evaluations were widely used to support decisions

³⁴ In this literature, learning is referred to as conceptual use, while actions based on evaluations fall under the term instrumental use. In this study, however, the more commonly understood terms, learning and action are used throughout.

³⁵ Yin, Robert K.; and Carol H. Weiss. 1988. Preliminary Study of the Utilization of AID's Evaluation Reports. Washington DC: USAID.
http://usaidprojectstarter.org/sites/default/files/resources/pdfs/PNABB688.Weiss_.utilization.pdf.

about existing and future projects, as well as to foster a better understanding of what works and what doesn't in terms of international development programs.

The final period, (2000 through the present) shifted attention taxonomies to the processes and factors by which evaluations foster or influence action, and, in turn, whether actions taken based on evaluations produce societal benefits (Kirkhart, 2000), (Mark & Henry, 2004). During this period, thinking about evaluation utilization moved beyond whether studies were read and used to support decision making to ask about the effects of evaluation utilization, i.e., whether evaluation use resulted in what Mark and Henry termed "social betterment" and whether such effects could be captured. In a landmark publication for the international development field, the World Bank attempted to do precisely that in its 2004 volume entitled *Influential Evaluations*.

Across these three periods, academic researchers and international development practitioners identified a range of factors which were viewed as being closely linked to evaluation utilization and which, depending on their status, were expected to facilitate or hinder utilization. Timeliness, quality and relevance are all factors of this type. In Table I below, a list of factors the study team identified from its literature review and subsequently used to develop study instruments and shape the study's data analysis, particularly under Question 4, are listed.

Table I. Utilization Literature-Based Factors that May Affect Evaluation Utilization

Factors Associated in Literature with Evaluation Products & Utilization	Factors Associated in Literature with Evaluation Processes & Utilization
Relevance to evaluation client/audience needs and ability to act on evaluation results	Commissioning organization evaluation culture
Timeliness (in time for decisions, on schedule; mid-activity or final)	Country or regional context (political situation, evaluation logistics)
Credibility (evaluation expertise, independence; relevant knowledge (local context, study client)	Need for/purpose of the evaluation
Quality (methods, data, analysis, communication)	Participation of various types of stakeholders in one or more stages of the evaluation process
Message content (nature of findings (positive/negative); new information; convergence with existing knowledge/views	Commissioning organization quality control practices (guidance, templates, intermediate and final design and report reviews)
Recommendations (specificity, actionable, practical, cost conscious, supported by evidence)	Post-evaluation processes – dissemination; decision-making; action planning and follow-up

Study Approach

MSI's approach to this study was structured around the four questions shown above. An initial review of the questions, during the design phase, indicated that the first three questions were similar to each other in that they sought descriptive information on evaluation uses. Question 4 is different in nature, as it asks about relationship between factors (i.e., business practices, enabling conditions). Accordingly, the study team's approach to answering the first three questions involved determining what data sources were pertinent, what methods were most appropriate for obtaining data from those sources, and how that data would be analyzed. Table 2 is the study team's "Getting to Answers" matrix of questions-by-methods that summarizes the results of this process. To answer Question 4, the study team used the same step-by-step process for identifying data sources and methods, but found that it also needed a conceptual framework for hypothesizing and testing relationships between various factors that might affect evaluation utilization. To this end, the team created a causal linkage diagram that helped lay out the sequence of process steps involved in the start-up, implementation, and post-evaluation processing of evaluation findings and recommendations.

This diagram, shown in Figure 1, was not a static framework. Over the course of the study, boxes (results) and arrows (hypotheses) were added several times to reflect what was being learned, and assumptions about directionality were challenged (e.g., does dissemination lead to learning or does learning lead to dissemination, or both). The dynamic nature of this graphic is what made this a valuable tool. Among the features added to it during the study process was an overlay consisting of three large arrows that represent the Mark & Henry (2004) conceptualization of evaluation utilization as influence pathways rather than as a static result. All of the elements on this diagram plus the factors listed in Table 1 helped to suggest what relationships might be worth examining as the correlations and logistical regression aspects of the analysis plan for Question 4, shown in Table 2, was implemented.

In the paragraphs below, information from these graphics, particularly the Getting to Answers Matrix, is summarized in a narrative form.

Question 1: How and when in the Program Cycle are evaluations used or not used?

To answer this question, the study team needed descriptive data of evaluation utilization disaggregated by stages of the Program Cycle. The team gathered primary data on utilization through a survey of USAID staff; group and key informant interviews with USAID staff and implementing partners; and extracting evidence of use from USAID documents (CDCSs, USAID Policy Documents, and Evidence Summits). Secondary data came from existing USAID reports including the PPR Evaluation Registry. All data collected were analyzed to associate them with specific Program Cycle stages. Interviews and PPR statements underwent content analysis while survey data and document evidence underwent quantitative analyses, including calculations of frequencies and crosstabs.

Question 2: What changes/decisions are made because of evaluations?

For this question, the study team used the same data collection and analysis methods stated under Question 1, but analyses ignored the stages of the Program Cycle while focusing on questions related to decision making and actions taken. Comparative frequencies for each of these decision types were run. To illustrate the decisions and changes that resulted from evaluations, the team collected short narratives of evaluation use from CDCS evaluation references, USAID group interviews, crowdsourcing responses, PPR use statements, and survey free-response questions. Narratives in the report were chosen as examples of how evaluation use may result in improved development effectiveness.

Question 3: To what degree and under what conditions does learning occur from evaluation findings that were not anticipated by the intended purpose of the evaluation?

This question was answered using data on learning from individual evaluations that was taken from existing survey answers provided by USAID staff, as well as by group interview responses. Qualitative content analyses were used for interview responses while survey responses used to determine frequencies and percentages of different types of learning.

Question 4: What particular business processes or enabling conditions appear to encourage or discourage the utilization of evaluations?

This question required a wide range of data on types of use in addition to data on processes and conditions around evaluations. Data came primarily from the survey and interviews. For quantitative analysis, beyond frequencies and cross-tabs, the team looked for correlations and statistical relationships using chi-square tests and a logistic regression to determine which combinations of factors most encouraging utilization. Qualitative data on conditions that discourage use was also drawn from group interviews with USAID staff. Statements discussing impediments to utilization underwent content analysis for frequency of themes. Data collection, analysis, and sampling methods used to answer these four questions are described below in detail.

Table 2. “Getting to Answers” Matrix

Study Question	Data Sources	Data Collection Methods	Sampling or Selection Plan	Data Analysis Methods
1. How and when in the Program Cycle are evaluations used or not used?	<ul style="list-style-type: none"> • USAID staff • USAID Partners • PPRs • CDCSs • Policy Documents • Evidence Summits 	<ul style="list-style-type: none"> • Document review of PPRs, CDCSs, policy documents, and Evidence Summit materials • Online crowdsourcing survey • Key Informant Interviews with USAID staff on policies and Evidence Summits • Group interviews with Operating Units and USAID partners • Survey of USAID staff familiar with identified evaluations 	<ul style="list-style-type: none"> • 2011-2014 evaluations selected for surveying using random sampling or random census sampling, depending on scope of evaluation • OUs selected with LER input and the following criteria: region, number of evaluations conducted, claimed usage in PPRs, evaluation use in CDCS, evaluation Mission Order modifications, interested M&E POCs • All available PPRs, CDCSs, policy documents, and Evidence Summit materials were included 	<ul style="list-style-type: none"> • Descriptive statistics of survey data (frequencies and cross-tabs) • Content and pattern analyses of interview transcripts • Frequencies and content analyses of <ul style="list-style-type: none"> - Use claims in PPR - References to Evaluations in CDCSs, policy documents, and Evidence Summits
2. What changes/decisions are made because of evaluations?	<ul style="list-style-type: none"> • USAID staff • PPRs 	<ul style="list-style-type: none"> • Key Informant Interviews with USAID staff on policies and Evidence Summits • Group interviews with Operating Units • Survey of USAID staff familiar with identified evaluations • Document review of PPRs • Crowdsourcing of USAID on evaluation impact stories 	<ul style="list-style-type: none"> • 2011-2014 evaluations selected for surveying using random sampling or random census sampling, depending on scope of evaluation • OUs selected with LER input and the following criteria: region, # of evaluations conducted, claimed usage in PPRs, evaluation use in CDCS, evaluation Mission Order modifications, interested M&E POCs 	<ul style="list-style-type: none"> • Descriptive statistics of survey data (frequencies and cross-tabs) • Content and pattern analyses of interview transcripts and crowdsourcing stories • Frequencies and content analyses of use claims in PPRs
3. To what degree and under what conditions does learning occur from evaluation findings that were not anticipated by the intended purpose of the evaluation?	<ul style="list-style-type: none"> • USAID staff • USAID evaluations 	<ul style="list-style-type: none"> • Survey of USAID staff familiar with identified evaluations • Review of quality factors in USAID evaluation reports • Key Informant Interviews with USAID staff on policies and Evidence Summits • Group interviews with Operating Units 	<ul style="list-style-type: none"> • 2011-2014 evaluations selected for surveying using random sampling or random census sampling, depending on scope of evaluation • Quality data and scores from evaluations with existing quality checklists including a random sample of 2011-2012 evaluations and a census of 2013-2104 E3 sector evaluations; additional evaluations selected for known cases of high and low utilization 	<ul style="list-style-type: none"> • Descriptive statistics of survey data (frequencies and cross-tabs) -

Study Question	Data Sources	Data Collection Methods	Sampling or Selection Plan	Data Analysis Methods
<p>4. What particular business processes or enabling conditions appear to encourage or discourage the utilization of evaluations?</p>	<ul style="list-style-type: none"> • USAID staff • Mission Orders • USAID Evaluations • USAID Guidance 	<ul style="list-style-type: none"> • Group interviews with Operating Units • Survey of USAID staff familiar with identified evaluations • Review of quality factors in USAID evaluation reports • Document review of USAID guidance 	<ul style="list-style-type: none"> • 2011-2014 evaluations selected for surveying using random sampling or random census sampling, depending on scope of evaluation • OUs selected with LER input and the following criteria: region, # of evaluations conducted, claimed usage in PPRs, evaluation use in CDCS, evaluation Mission Order modifications, interested M&E POCs • Quality data and scores from evaluations with existing quality checklists including a random sample of 2011-2012 evaluations and a census of 2013-2104 E3 sector evaluations; additional evaluations selected for known cases of high and low utilization 	<ul style="list-style-type: none"> • Descriptive statistics of survey data (frequencies and cross-tabs) • Content and pattern analyses of interview transcripts • Frequencies and content analyses of changes to the Standard Mission Order on Evaluation by Missions • Logistic Regression • Correlations: <ul style="list-style-type: none"> - Chi-squares - Pearson Correlations - Spearman Correlations

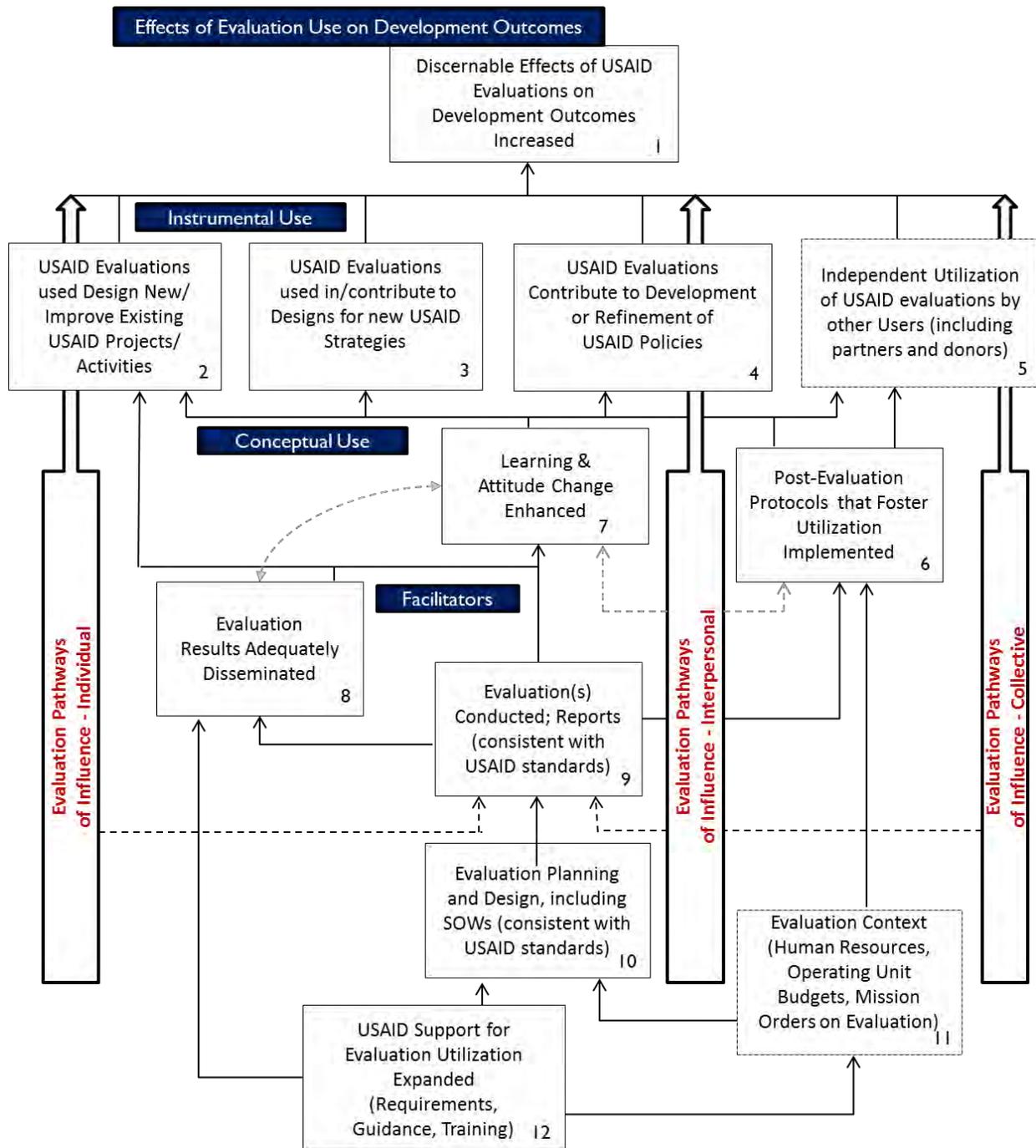


Figure 1. Evaluation Utilization Theory of Change for USAID

In the following sections, additional information is provided on the data collection, sampling, and data analysis methods used by the study team.

Data Collection

Three primary modalities for data collection were used including document reviews, survey research methods, and interviews.

Document Review

The study team reviewed a variety of USAID documents to better understand how USAID evaluations have been used. Documents reviewed include: Policy Papers, Evidence Summit documentation, PPR evaluation use statements, Country Development Cooperation Strategies (CDCSs), Evaluation Mission Orders, and evaluation reports. Additional information for each document reviewed is provided below.

- **Policy Documents** – Looking for evidence of evaluation use, the study team examined 19 policy documents and the USAID Policy Framework 2011-2015. Any references or citations included in the documents were analyzed to determine the type and source of the reference or citation, with a focus on use of USAID evaluations. The team also looked for references to Evidence Summits in policy development as a possible chain of use back to evaluations. Findings from this effort can be found in Sub-study #3 in Annex 4.2.
- **Evidence Summits** – The study team identified nine Evidence Summits and experience summits based on information posted to USAID’s website, USAID’s Learning Lab, a general internet search, and conversations with USAID staff. Resources from each summit, such as transcripts, agendas, and recommended readings were analyzed for references to USAID evaluations. Findings can be found in Sub-study #4 in Annex 4.2.
- **USAID’s Annual Performance Plan and Report (PPR)** – The study team reviewed 1,077 PPR Evaluation Registry entries from 2011-2014 and removed all entries that did not directly claim an evaluation had been used. This left 136 entries, which were then qualitatively analyzed to identify patterns in claimed use. Findings can be found in Sub-study #1 in Annex 4.2.
- **Country Development Cooperation Strategies (CDCSs)** – As of the start of the study, USAID has approved and posted 45 CDCSs and six Regional Development Cooperation Strategies (RDCCS) to its website. These were all reviewed to identify the types of evidence used to support the strategies, and where in the strategies they were used. Quantitative and qualitative analyses of evidence found, by section and type, were conducted. Findings can be found in Sub-study #2 in Annex 4.2.
- **Mission Orders on Evaluation** – Customized Mission Orders on Evaluation were analyzed for changes from the standard Mission Order template. The team was provided 41 Mission Orders available on ProgramNet. For each document, the team counted frequency of changes and used a content analysis of changes that could affect utilization. Findings can be found in Sub-study #7 in Annex 4.2.
- **Evaluation Reports** – All of the evaluations in the study universe were reviewed for basic descriptive data and were then categorized by a number of factors including country, region, sector, year, scope, and timing. Missing or incorrect data were corrected. A small sub-sample of 45 evaluations was further analyzed to understand the relationship between purpose statements, evaluation questions, and recommendations to evaluation utilization. Each evaluation’s questions and recommendations were counted while purpose statements and the first five recommendations of each report underwent content analysis. These Sub-studies, #5 and #6, can be found in Annex 4.2. The team also used USAID’s 2013 meta-evaluation evaluation report quality checklist to analyze

report quality for 241 evaluations, providing a quality score for each.³⁶ This scoring checklist instrument is included in Annex 5.7.

Survey Research

The study team used web-based surveys at two points in the data collection process. The first was to broadly gather examples of evaluation utilization while the second was meant to gather in-depth information on how specific evaluations were used.

- **Call for Utilization Stories** – Crowdsourcing was used to solicit stories of evaluation utilization and development impact from a broad range of individuals both within and outside of USAID. The study team, in collaboration with USAID, sent out emails and posted advertisements in locations such as ProgramNet and Learning Lab to encourage staff and partners to submit utilization stories using a short online survey. Responses were collected from December 2014 to February 2015. Seven stories were received with four traceable to USAID evaluations in the study universe.
- **Evaluation Utilization Staff Survey** – To obtain a well-rounded perspective of how specific evaluations were used at USAID, the study team designed an online survey geared towards USAID staff familiar with particular evaluations. Evaluations were randomly selected, at which point the team worked with M&E POCs to identify survey respondents for each evaluation who were present and familiar with what occurred when these evaluations were received. The online survey was open for approximately six weeks during May and June 2015. Of the 208 evaluations sampled, 118 surveys were returned (57% response rate). Survey results can be found in Annex 4.1.

Interviews

The study team held both group and key informant interviews with USAID staff and partner organizations implementing or evaluating activities. For a complete list of interviews held, see Annex 3.

- **Group interviews** were held with approximately 250 USAID staff members from 12 Country Missions, two Regional Missions, four Technical Bureaus, and six Regional Bureaus, as well as PPL/LER. Staff were divided based on seniority, with separate meetings held with mid-level staff and senior-level staff in nearly all Operating Units (OUs). Group interviews were also held with small groups from six partner firms and three NGOs that undertake USAID-funded evaluations. All USAID group interviews were recorded and transcripts were produced for analysis. Group interviews were held over a one-month period from June 5 to July 24, 2015.
- **Key informant interviews** were also conducted with current and former USAID staff known to be particularly knowledgeable about evaluation utilization.

Sampling and Selection Methods

For the majority of data sources and data collection efforts, the study team selected all available sources (CDCSs, PPR entries, Policy Documents, Evaluation Mission Orders, etc.). In instances where there was too large of a universe for full inclusion, a sampling protocol was required, namely for the evaluation survey and OU interviews. The sampling approach in those cases is described below.

Evaluation Utilization Staff Survey Sampling and Weighting Procedures

The study team was provided with information for 758 documents coded in the USAID Development Experience Clearinghouse (DEC) as evaluations for the years 2011-2014. Upon review, the team removed 149 documents from the universe for being duplicates, foreign language documents, and

³⁶ Hageboeck et al., 2013.

documents other than evaluations, evaluations without an identifiable sponsoring OU, or evaluations commissioned by implementing partners, resulting in a study universe of 609 evaluations.

Due to the size of the universe, it was agreed that a strategy should be developed that would best represent the variety of evaluations at the Agency. To this end, the study team examined the distribution of the 609 evaluations in the universe by year, scope, timing, region and sector. On an annual basis, the number of evaluations varied, but not dramatically, and it was agreed with USAID that this was not a main focus for the study. With respect to timing, the vast majority of the evaluations in the study universe were either mid-term (55) or final (52), with only eight ex-post evaluation and three impact evaluations in the study universe. Sorting evaluations in the universe by scope, the team found that 83 percent were single country evaluations with a single project/activity focus. Other clusters were all much smaller, as Table 3 shows.

Table 3. Number of Evaluations in the Study Universe by Scope (N=609)

Evaluation Scope	Number of Evaluations
Single Project/Activity, Single country	503
Single Project/Activity, Multiple Country	43
Sector-wide, Single Country	10
Regional Program or Project	15
Multiple Projects or Activities, Single Country	26
Multiple Projects or Activities, Multiple Countries	5
Global Program or Project	7

In addition, the study universe was found to be skewed rather than evenly distributed on both a regional and sector basis, as shown in Figure 2.

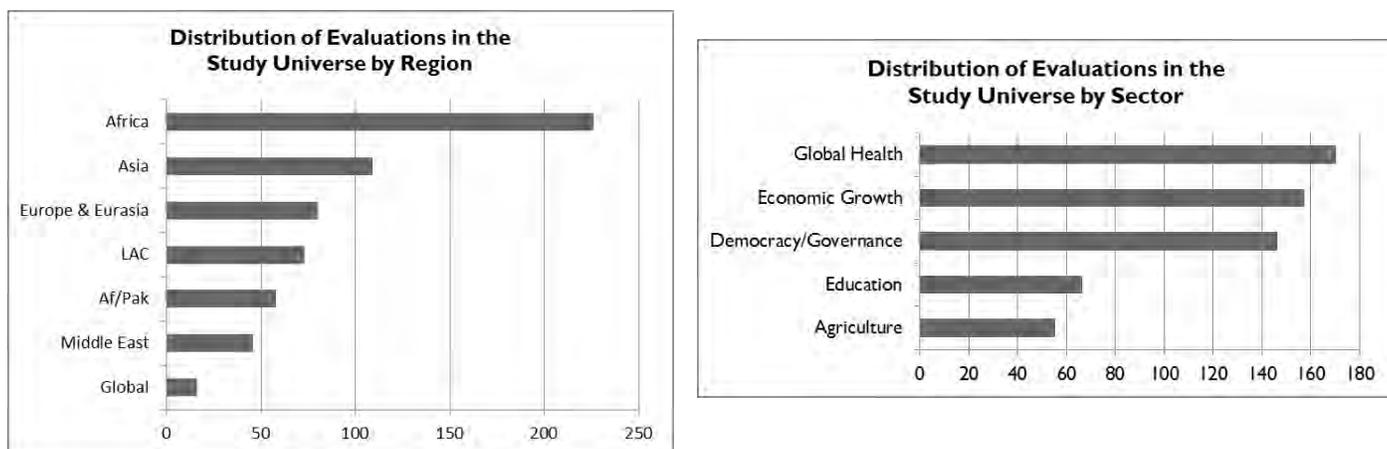


Figure 2. Distribution of Evaluations in the Study Universe Region and Sector

Through discussion with USAID, MSI prioritized these various factors to create a sampling plan in line with the spirit of the study’s commitment to represent the universe as fairly as possible. Among the factors discussed above, study scope was viewed as a characteristic on which there was a good chance

that evaluation use would vary, and the same was true for evaluation timing. Neither region nor sector were expected to play a critical role in determining evaluation use, as they had not been explanatory factors in early evaluation studies of which the team and PPL/LER were aware.

With these conclusions in hand, MSI set up a sequenced sampling plan. The first step involved sectioning off the largest cluster of evaluations by scope (single country, single project/activity) consisting of 503 evaluations, to be sampled using a random sampling technique to ensure a representative sample. For the smaller clusters by scope, it sorted between those with less than 10 evaluations, for which it recommended selecting all cases, and clusters of more than 10 evaluations, for which it recommended setting a quota (at 10 evaluations for each of these clusters), with random sampling up to the quota inside each cluster. In discussions with USAID, it was decided that two other small clusters of evaluations should also be treated in this manner, namely the eight ex-post evaluations and the three impact evaluations in the universe. These two small clusters were then assigned to the same regime as other clusters of 10 or fewer evaluations, i.e., all cases were selected for inclusion in the sample on which evaluation surveys were sent out.

The study team next consulted with USAID to determine what type of random sampling approach to apply to the large single country, single project/activity cluster of 503 evaluations. The first decision made was to not stratify that cluster by timing as it was relatively evenly split between mid-term and final evaluations and any random sampling procedure would be likely to result in a fairly equal distribution along that dimension. The next decision to be made was whether to stratify by sector or region or both. During the design phase, MSI presented options to USAID, based on a matrix it had developed that showed the distribution of the 503 evaluation in the large cluster by region and sector. The pros and cons of the two main options, a simple random sample of the 503 evaluations in this cluster that ignored region and sector and a disproportionate stratified sample, were outlined for the Agency, as shown below in Table 6. USAID selected the single random sample option for selecting evaluations from the large single country, single project/activity cluster.

Table 6. Advantages and Disadvantages of Sampling Options for Single Project/Activity, Single Country Cases

Sampling Approach	Advantages	Disadvantages
Simple Random Sample	<ul style="list-style-type: none"> • Much of the likely audience is at least somewhat familiar with this sampling approach • No weighting required during analysis 	<ul style="list-style-type: none"> • Would yield a sample with more cases of some kinds than USAID might want to focus on (e.g., Africa, health) • Could fail to represent some cases it does want to focus on (e.g., Middle East)
Disproportionate Stratified Sample	<ul style="list-style-type: none"> • Ensures that all subpopulations of interest are represented, albeit not proportional to their numbers in the study universe • Caps the number of cases examined from large subpopulations 	<ul style="list-style-type: none"> • Much of the likely audience may not be familiar with this sampling approach • Weighting during analysis required • May require that two USAID regions be combined in the sampling approach so that at least 2 evaluations show in every cell.

The resulting simple random sample, drawn at the level required for 85 percent confidence level and 5 percent margin of error, included 206 evaluations. Table 7 displays this large cluster sample size as well

as the sample sizes for other clusters which were sampled on a quota (randomly selected) or census basis.

Table 7 also shows the survey response rate by sampling cluster and the response rate of 57 percent overall on a cluster basis. To understand whether the sample survey's relatively low response rate (compared to the number needed to achieve an 85 percent confidence, 5 percent margin of error level), MSI analyzed this response rate by year, scope, timing, sector and region, as shown in Table 8. No strong patterns that would suggest a particular response rate bias were found.

Given the variety of samples used to ensure that small clusters were represented and the overall modest response rate for the study's survey on evaluation utilization, USAID asked MSI to test whether weighting the survey sample would yield different answers for a subset of the survey questions. While that test indicated that there was little difference between weighted and unweighted sample responses, USAID asked that the study report present weighted responses in the body of the report, to help ensure that the answers PPL/LER receives and shares with others are as representative of the study universe of 609 evaluations as possible. This request was honored and the main report is presented on a weighted basis. Annex 4.1, which provides answers to all survey questions, is presented on both a weighted and unweighted basis.

Table 7. Summary of the Utilization Study Universe, Sample, Response Rate and Sampling Method by Evaluation Scope and Timing

Evaluation Qualities	Total in Universe	Sent for Survey	Surveys Returned	Response Rate (%)	Sampling Method
Single Project/Activity, Single Country	503	155 (31%)	94	61%	Sampling varied by timing, see timing subsets
<i>Mid-term & Final</i>	487	143(29%)	86	60%	Simple random sample from 487 cases
<i>Continuous/Impact</i>	6	5 (83%)	3	60%	Sampled for timing. There were 8 evaluations where the timing was continuous/impact in the universe of which 6 were under this scope cluster. Five of the 6 in this scope cluster were impact evaluations and all were sampled; one was a pre-post only with no comparison group and was not sampled as the target for this timing cluster was to include all impact evaluations
<i>Ex-post</i>	10	7 (70%)	5	71%	Quota sample of 10 from a total of 13 ex-post evaluations in the universe. Ten were randomly selected, of which 7 turned out to be in this scope cluster and the other 3 fell into other scope clusters.
Single Project/Activity, Multiple Country	43	10 (23%)	4	40%	This cluster was larger than 10 and quota sampling of 10 was the rule applied.
<i>Mid-term & Final</i>	41	8	4	50%	8 evaluations were randomly selected from the 41 mid-term and final towards the quota of 10, after 2 evaluations were selected based on timing which also contributed to meeting this quota.
<i>Continuous/Impact</i>	2	2	0	--	Sampled for timing before the remainder of the quota sample was selected. Two of the 8 evaluations with this timing designation were in this scope cluster; both were sampled.
<i>Ex-post</i>	0	0	0	--	N/A
Sector-wide, Single Country	10	10 (100%)	7	70%	This cluster included only 10 cases and fell under the census rule.
<i>Mid-term & Final</i>	9	9	6		All of these evaluations were included in the sample

Evaluation Qualities	Total in Universe	Sent for Survey	Surveys Returned	Response Rate (%)	Sampling Method
<i>Continuous/Impact</i>	0	0	0	--	N/A
<i>Ex-Post</i>	1	1 (100%)	1	100%	Sampled for timing; also included in the census of all evaluations in this scope cluster
Regional Program or Project	15	10 (67%)	2	20%	This cluster was larger than 10 and quota sampling of 10 was the rule applied
<i>Mid-term & Final</i>	15	10	2	20%	Quota sample of 10, randomly selected
<i>Continuous/Impact</i>	0	0	0	--	N/A
<i>Ex-Post</i>	0	0	0	--	N/A
Multiple Projects or Activities, Single Country	26	10 (38%)	5	50%	This cluster was larger than 10 and quota sampling of 10 was the rule applied
<i>Mid-term & Final</i>	24	8	3	38%	8 cases were randomly selected from the 24 in this sub-cluster to contribute to the quota of 10
<i>Continuous/Impact</i>	0	0	0	--	N/A
<i>Ex-Post</i>	2	2 (100%)	2	100%	Sampled for timing before the quota sample was drawn for the mid-term and final evaluations. The 2 evaluations in this timing cluster were both verified as impact evaluations
Multiple Projects or Activities, Multiple Countries	5	5 (100%)	3	60%	This cluster included fewer than 10 cases and fell under the census rule.
<i>Mid-term & Final</i>	5	5	3	60%	Census; all included
<i>Continuous/Impact</i>	0	0	0	--	N/A
<i>Ex-Post</i>	0	0	0	--	N/A
Global Program or Project	7	6 (86%)	3	50%	This cluster included fewer than 10 cases and fell under the census rule, with one exception (see below)
<i>Mid-term & Final</i>	7	6	3	50%	Census, excluding one evaluation where sector = "other" and the study design had said cases of sector = other would be excluded

Evaluation Qualities	Total in Universe	Sent for Survey	Surveys Returned	Response Rate (%)	Sampling Method
<i>Continuous/Impact</i>	0	0	0	--	N/A
<i>Ex-Post</i>	0	0	0	--	N/A
TOTAL EVALUATIONS	609	206	118	57%	—

Table 8. Survey Non-Response Analysis by Scope, Region, Sector, Timing and Year

<u>Cluster by Scope</u>	Universe 2011- 2014	Sample Surveys Sent	Response Surveys Returned	Responses Not Returned	Percent Not Returned
Single Activity, Single country	503	155	94	61	39%
Other Smaller Clusters (total)	106	51	24	27	53%
Single Project/Activity, Multiple Country	43	10	4	6	60%
Sector-wide, Single Country	10	10	7	3	30%
Regional Program or Project	15	10	2	8	80%
Multiple Projects or Activities, Single Country	26	10	5	5	50%
Multiple Projects or Activities, Multiple Countries	5	5	3	2	40%
Global Program or Project	7	6	3	3	50%
TOTAL	609	206	118	88	43%

<u>Region</u>	Universe 2011- 2014	Sample Surveys Sent	Response Surveys Returned	Responses Not Returned	Percent Not Returned
Africa	226	79	37	42	53%
Europe & Eurasia	80	31	23	8	26%
Asia	109	34	21	13	38%
LAC	73	28	14	14	50%
Af/Pak	58	19	12	7	37%
Middle East	46	6	6	0	0%
Global	16	9	5	4	44%
Multi*	1	0	0	0	N/A
TOTAL	609	206	118	88	43%

*Evaluations with this designation were not sent surveys, per the study plan

<u>Sector</u>	Universe 2011- 2014	Sample Surveys Sent	Response Surveys Returned	Responses Not Returned	Percent Not Returned
EG	157	58	35	23	40%
DG	146	54	32	22	41%
Global Health	170	54	29	25	46%
ED	66	21	12	9	43%
AG	55	18	9	9	50%
Multi-Sector	1	1	1	0	0%
Other*	14	0	0	0	N/A
TOTAL	609	206	118	88	43%

*Evaluations with this designation were not sent surveys, per the study plan

<u>Evaluation Type and Timing</u>	Universe 2011- 2014	Sample Surveys Sent	Response Surveys Returned	Responses Not Returned	Percent Not Returned
Mid-term	291	99	55	44	44%
Final	297	90	52	38	42%
Ex Post	13	10	8	2	20%
Impact	8	7	3	4	57%
TOTAL	609	206	118	88	43%

<u>Year</u>	Universe 2011- 2014	Sample Surveys Sent	Response Surveys Returned	Responses Not Returned	Percent Not Returned
2011	134	45	21	24	53%
2012	196	60	39	21	35%
2013	172	53	29	24	45%
2014	107	48	29	19	40%
TOTAL	609	206	118	88	43%

Interviews

The study team conducted a large number of both individual key informant and group interviews with USAID OUs, i.e., Missions and Offices in USAID/Washington, which were conducted at two levels (mid-level, AOR/COR staff and the senior staff level). For the individual key informant interviews, it was fairly easy to identify respondents who could speak for the portfolios and topics of interest to the study team. For OU group interviews, on the other hand, a more systematic selection approach was needed.

In the study's approved Design Proposal, MSI included a table that suggested how many OUs of various types and sizes appeared to be needed to fairly represent the Agency's diversity in this regard, which is shown in Table 9 below, from the approved design report. As is evident, this table includes information on the first criteria (i.e., number of evaluations in the study universe) the study team used to segment the universe of USAID OUs.

Table 9. Design Stage Sampling Plan for Selecting OUs

Operating Unit Cluster	Number of Units or Offices	Proposed Sample of Units
Missions with 1 DEC evaluation in the 2011-2014 period	13	3-4
Missions with 2-3 evaluations in the study period	15	4
Missions with 4-7 evaluations in the study period	15	4
Missions with 8-30 evaluations in the study period	16	4-5
Regional Missions or Office ³⁷	7	2
Global Health Bureau	Spread among topical offices	4
DCHA	Spread among topical offices	4
EGAT	Spread among topical offices	4 or more ³⁸
BFS		2 or more ³⁹
PPL		1
TOTAL		32-36

In addition to organizing OUs by type and number of evaluations completed, the study team attempted to structure its selection of country-level Missions to be representative on a regional basis, and with respect to Mission size (i.e., small, medium, and large based on the level of resources managed). The study team created a matrix to help structure this selection process and identified all of the Missions that fed into each matrix cell, as illustrated in Table 10, which is a version of the matrix shared with USAID in the weeks before the final set of Missions with which group interviews were conducted. It should be noted that while MSI used this type of systematic approach to Mission selection, the process was not mechanical. Some Missions that were identified through the matrix selection process were unable to participate, and alternates were chosen to replace them.

³⁷ Both of these terms were found in the sponsoring organization codes assigned by the DEC, but are presumed for purpose of this study to be roughly equivalent.

³⁸ Final number to be based on discussions with USAID, as this bureau has 12 separate offices

³⁹ Final number to be based on a more detailed analysis of this bureau's structure

Table 10. Distribution of Pre-Final Single Country Missions for Operating Unit Interviews

	AF/PAK	AFRICA	LAC	E&E	ME	ASIA
High ⁴⁰ Number of Evaluations (8 or more)	Afghanistan (15/41)	Uganda (7/13) ⁴¹ Liberia (0/13)	Nicaragua (1/10)	Armenia (8/9)	Jordan (0/10)	Indonesia (0/16) Bangladesh (4/15)
Low Number of Evaluations (7 or less)			Colombia (2/4)	Kosovo (0/7)	Morocco (0/5)	

The final set of OUs with which interviews were conducted included 12 bilateral. Three Missions were selected from Africa, one Mission was selected from Afghanistan/Pakistan, and two Missions were selected from each of the remaining regions. In cases where selected Missions declined to participate, another Mission from the same geographic region that most closely resembled it on the above criteria was selected to replace it. In one instance, a single interview was held with a Mission, instead of one at each level. Two Regional Missions were also selected as they produced a large number of evaluations and had been in existence long enough to have potentially established processes for evaluation utilization. The final list of Missions with which group interviews were conducted is shown in Annex 3. The bilateral set closely matched Table 10 including all Missions listed above, minus Liberia, plus Rwanda and Nigeria.

Four Technical Bureaus were selected based upon the number of evaluations completed during the study period. Specific offices within these Bureaus were selected based upon the frequency with which evaluations focused on their technical areas. All six Regional Bureaus were invited to participate in the group interviews. Interviews were held with all 10 Washington-based Bureaus invited to participate.

In addition to USAID group interviews, the study team conducted a small number of interviews with USAID partner organizations including three international NGOs and six firms, as listed in Annex 3. These organizations were asked about their experience with USAID evaluations and about their utilization. To identify USAID partners for group interviews the study team looked at the 44 evaluations commissioned by implementing partners and identified the five most common NGOs, of which three agreed to be interviewed. The team also looked at which firms authored the most evaluations and found the top seven firms, of which six agreed to be interviewed.

Data Analysis Methods

Data analysis methods used in this report are pursuant to the data collection methods used. The study employed a mixed method approach, using both qualitative and quantitative methods, and triangulated the findings to ensure accuracy of conclusions and recommendations.

Given the variety of data sources and data collection methods, the study team needed to identify the appropriate qualitative and quantitative data analyses methods to fully extract findings from the sources. While most of the data were processed in-house by MSI, some quantitative data were also reviewed and independently interpreted by an external statistical analysis specialist from the University of Pittsburgh. See Annex 4.3 for a detailed write-up of the logistic regression analysis.

⁴⁰ Op cit.

⁴¹ In the parentheses we present the number of evaluations on which PPRs reported some type of utilization over the number of evaluations in the study universe for that country.

Qualitative Data Analysis

For all data collected that contained any form of text or narrative, qualitative analyses were used. These data came from interviews, open-ended survey responses, and information extracted from USAID documents. These data were analyzed for content and patterns. To the extent possible, frequencies of themes were counted within each analysis. In some cases, narratives were used in text boxes or were further investigated to create case studies. Quotes or case studies were verified with USAID prior to use.

Quantitative Data Analysis

Quantitative data come from the survey responses and document reviews. Numerical data were entered into Excel-based databases and then transferred into SPSS. From the study universe, only the evaluations that had been sampled for the survey or scored using the quality checklist were included in these databases. In addition to utilization variables from individual survey questions, data were combined from the survey, CDCs, PPRs, and crowdsourcing to create variables which represented instrumental use (action) and conceptual use (learning) according to any source.

Frequencies and percentages were computed for all variables. Crosstabs and chi-squares were run between utilization variables and factors that could influence utilization. In cases where crosstabs produced very small numbers in some cells, Pearson correlations and Spearman correlations were also run. Significance was tested at the .05 level. From these correlations, the study team was able to identify factors associated with utilization. To determine the combined effects of groups of factors on utilization, logistic regressions were conducted by an external statistics specialist.

Study Limitations

The key study limitations for this study include:

General Limitations

- **Response bias** – The study’s survey and interviews produced self-reported data which were not independently verified. As a result, such data may not have been fully reliable.⁴² To the extent possible, the study team corroborated evidence from multiple sources to increase reliability.
- **Projects versus activities** – Though USAID guidance clearly distinguishes projects from activities, this distinction is not consistently upheld by staff and partners. Data from a variety of sources, including evaluation reports, interviews, and survey responses did not always reflect the appropriate terminology, making it impossible for the team to report at each level distinctly.
- **Information on project and activity design** – Due to procurement sensitivity issues, MSI was unable to analyze Project Appraisal Documents (PADs) from the study period. Insight into evaluations’ contributions to design documents was limited to survey and interview responses and a summary report on a review of PADs that USAID had conducted with in-house staff. The latter provided little useful information as evaluation utilization was not a focus of that study.

⁴² While MSI considers data that come exclusively from self-reporting to potentially be biased, it also notes that interviews with Agency staff are used as the basis for U.S. government reviews of evaluation utilization in federal agencies, conducted by the U.S. Government Accountability Organization (GAO). Given the paucity of alternative sources, this approach appears to be the norm. In this study, as described elsewhere, the study team made a deliberate effort at every stage of the evaluation to gather data that could be used to verify self-reported responses.

Survey

- **Generalizability of survey findings** – The stratified random sampling protocol for selecting evaluations for the survey was designed to be representative of the study universe. Due to the lower than anticipated survey response rate (57%), the responses received did not provide a sufficiently high confidence level to claim statistical representativeness. Survey results still provide a fair representation of the study universe and are reported as such, particularly with weighted percentages of survey results used throughout the report.
- **Institutional memory** – USAID Foreign Service Officers spend an average of two years at each post, taking their experience and memory of circumstances surrounding evaluations with them. This affected the team’s ability to find appropriate respondents for the survey, causing some evaluations to be removed from the sample and others to have respondents less than fully familiar with the entire evaluation process. This may affect the reliability of some of the survey responses.

Group Interviews

- **Inconsistency of Secondary Questions Asked** – Group interviews used a semi-structured interview guide. Major topics were always covered, but secondary questions under each topic were not always addressed, leading to an inability to compare consistently across OUs.

Development Experience Clearinghouse

- USAID provided the study team all evaluations submitted to the DEC for 2011-2014, but the team cannot confirm that all USAID evaluations were actually uploaded to the DEC, or therefore included in the study universe. For example, the study team found 40 evaluations listed in the PPR that could not be matched to evaluations in the DEC.
- The DEC uploading process results in inconsistencies including: duplication of evaluations; non-evaluations being labeled as evaluations; evaluations being labeled as something other than evaluations; and omitted or inaccurate descriptive data for evaluations. MSI reviewed evaluations to correct inaccuracies in the universe, though it is possible some may not have been identified.

Team Composition

The following provides a brief description of each of the members of the utilization study team. All team members have signed forms indicating they have no conflicts of interest by contributing to this study.

Molly Hageboeck (MSI) is the Team Leader and has significant experience working both for and with USAID on evaluations and evaluation quality. She has conducted numerous evaluations, including several meta-evaluations and quality review exercises. Ms. Hageboeck has served as Team Leader for previous meta-evaluations for USAID, including “A Review of the Quality and Coverage of A.I.D. Evaluations, FY 1989 and FY 1990;” “Trends in International Development Evaluation Theory, Policy and Practices;” “From Aid to Trade – Delivering Results;” and “Meta-Evaluation of Quality and Coverage of USAID Evaluations, 2009-2012.”

Micah Frumkin (MSI) is the team’s Evaluation Expert, and has experience in designing and maintaining M&E systems, training and certifying evaluators, and has contributed to numerous evaluations. His work has focused on evaluations at USAID, including as a team member on studies such as “Quality Review of Recent USAID Evaluation Statements of Work,” and “Trends in International Development Evaluation Theory, Policy and Practices” report, in addition to acting as the team manager and senior coder for the “Meta-Evaluation of Quality and Coverage of USAID Evaluations, 2009-2012.”

Jenna Lindeke Heavenrich (MSI) is the Research Assistant for the study. She has experience performing qualitative and quantitative data collection and analysis for scholarly articles on international development and intercultural relations and has designed an M&E system for a grassroots NGO in India.

Lala Kasimova (MSI) is the Utilization Study's Project Manager and has been managing and supporting field programs and evaluations, including M&E support platform projects, for MSI for three years. Most recently, she served on a team of technical quality reviewers for the sectoral synthesis of E3 Bureau evaluations, a gender analysis of E3 evaluations, as well as capacity development case study assessment.

Melvin M. Mark (Pennsylvania State University) was the academic advisor to the study, bringing his expertise on social psychology and the theory and practice of evaluations. Dr. Mark has written more than 130 articles and book chapters and edited 12 books. He has served as the President of the American Evaluation Association and an Editor of the *American Journal of Evaluation*. Dr. Mark also received the American Evaluation Association's Lazarsfeld Award for Contributions to Evaluation Theory.

Aníbal Pérez-Liñán (University of Pittsburgh) conducted the study's logistical regressions. He has used complex quantitative analysis such as logistic regression and qualitative comparative analysis to support his research on democratization, institutional performance, and the rule of law in Latin America. Dr. Pérez-Liñán has published one book and three articles on the subject.

ANNEX 3 – LIST OF INTERVIEWEES

Organization/ Operating Unit	Point of Contact	Participant Type	Number of Participants	Interview Date and Time (EST)
USAID Field Missions				
Nicaragua	Marcella Villagra	Mid-Level	5	June 5, 1-2pm
		Senior-Level	4	June 12 3:30-4:30pm
Afghanistan	Daryl Martyris	Mid-Level	6	June 9, 7-8am
		Senior-Level	4	June 10, 7:30-8:30am
Colombia	Liz Mendenhall	Mid-Level	7	June 11, 5-6pm
		Senior-Level	2	June 17, 3-4pm
Morocco	Alice Rowley	Mid-Level	6	June 16, 10am-11am
		Senior-Level	3	June 22, 7-8am
Indonesia	Adam Jung	Mid-Level	8	June 16, 9-10pm
		Senior-Level	5	June 22, 9-10pm
Kosovo	Melita Cacaj	Mid-Level	4	June 30, 7:30-8:30am
		Senior-Level	4	July 2, 8-9am
Bangladesh	Farheen Khurram	Mid-Level	13	June 30, 10-11pm
		Senior-Level	9	July 13, 11-12pm
Armenia	Mervyn Ellis	Mid-Level	9	July 1, 7:30am-8:30am
		Senior-Level	4	July 9, 7:30-8:30am
Jordan	Kenana Amin	Mid-Level	6	June 24, 7-8am
		Senior-Level	5	June 24, 8:30-9:30am
Rwanda	Daniel Handel	Mid-Level	3	July 14, 8-9am
		Senior-Level	4	July 13, 8-9am
Uganda	Lane Pollack	Program Office Learning Advisor	1	July 24, 11am
Nigeria	Joyce Elele	Mid-Level	11	June 30, 10-11am
		Senior-Level	12	July 7, 10:30-11:30am
USAID Regional Missions				
Southern Africa Regional Mission	Dinah Zeltser & Lloyd Jackson	Mid-Level	8	July 6, 9:30-10:30am
		Senior-Level	6	July 13, 9:30-10:30am
Regional Development Mission for Asia (RDMA)	Suzanne Polak	Mid-Level	4	June 22, 9:30pm
		Senior-Level	5	June 23, 9:30pm
USAID Regional Bureaus				
OAPA	Matthew Hermerding	Mid-Level	4	June 9, 2-3pm
		Senior-Level	4	June 12, 11:00am- 12:00pm
E&E	Kraemer Lovelace	Mid-Level	3	June 16, 2-3pm
		Senior-Level	4	June 18 3:30-4:30pm
Middle East	Christine MacAulay	Mid-Level	4	June 17, 9:30-10:30am
		Senior-Level	2	June 24, 2-3pm
LAC	Tanushree	Mid-Level	10	June 29, 11am-12pm

	Isaacman	Senior-Level	7	July 2, 2pm-3pm
Africa Bureau	Alphonse Bigiramana	Mid-Level	5	July 1, 12-1pm
		Senior-Level	6	July 6, 3-4 pm
Asia Bureau	Mike Wagg & Jennifer Kim	Mid-Level	3	July 9, 4:30-5:30pm
		Senior-Level	4	July 9, 1-2pm
USAID Technical Bureaus				
Global Health	Kristin Saarlax	Mid-Level	3	June 17, 11am-12pm
		Senior-Level	1	July 10, 3:30-4:30pm
Bureau for Food Security	Lesley Perlman	Mid-Level	8	June 29, 1-2pm
		Senior-Level	6	June 30, 1-2pm
E3	Bhavani Pathak & Rebecca Maestri	Mid-Level	6	June 17, 9-10:30am
		Senior-Level	2	July 6, 2-3pm
Democracy, Conflict, and Humanitarian Assistance (DCHA)	Amber Ussery	Mid-Level	4	June 29, 10-11am
		Senior-Level	8	July 1, 1-2pm
Other USAID Interviews				
PPL	Noam Unger	USAID Policy Key Informant	1	July 15, 4:30pm
PPL	Tjip Walker	Evidence Summits Key Informant	1	July 8, 3pm
Total USAID Interviewees:			254	
NGOs that conduct evaluations of USAID activities they implement				
Save the Children	Tanya Guenther	Evaluation Specialists	8	June 22, 10:30-11:30am
Mercy Corps	Jon Kurtz	Evaluation Specialists	1	July 13, 1-2pm
Catholic Relief Services	Jose Thekkiniath (TD)	Evaluation Specialists	1	July 10, 8:30-9:30am
Firms that conduct USAID evaluations				
Checchi & Co	Aimee Rose & James Agee	Evaluation Specialists	2	July 1, 11am-12pm
Social Impact	Jim Fremming	Evaluation Specialists	7	June 15, 3-4pm
IBTCI	Ed Allen	Evaluation Specialists	1	June 25, 11:30am-12:30pm
dTS	Janet Kerley	Evaluation Specialists	1	July 7, 4-5pm
Mendez and Associates	Mirela McDonald	Evaluation Specialists	1	July 2, 10-11am
MSI	Lala Kasimova	Evaluation Specialists	5	June 24, 12:30pm-1:30pm
Total Non-USAID Interviewees			27	

ANNEX 4 – DATA ANNEXES

4.1 – SURVEY RESULTS

(118 Responses)

Evaluation and Respondent Identification

**1) This survey is for the following evaluation: *Evaluation Title, Mission or Office, and Evaluation URL*
Does this information match what you received in the introductory email?**

One survey participant accidentally swapped the surveys for the two evaluations (s)he was responding to. The data was straightened out before analysis and no other cases of surveys not matching evaluations were found.

2) Years with USAID (N=118)

	Number of Responses	Percent	Weighted Percent
1 or Less	6	5.1%	3.99%
2 – 5	29	24.6%	23.20%
More than 5	83	70.3%	72.81%

3) When did you first become aware of this evaluation? (N=118)

	Number of Responses	Percent	Weighted Percent
When it was being planned	84	71.2%	69.37%
During the evaluation implementation period	7	5.9%	6.54%
The initial briefing on its findings	4	3.4%	4.43%
Draft report review	4	3.4%	3.26%
Final report acceptance or dissemination	4	3.4%	3.85%
Post-evaluation decisions/action planning	10	8.5%	9.94%
Just recently, to take this survey	5	4.2%	2.61%

4) How involved were you during various stages of the evaluation process? (N=118)

	Highly Involved	Weighted Percent	Moderately Involved	Weighted Percent	Not Involved	Weighted Percent
Planning for the evaluation	65 (55.1%)	54.75%	20 (16.9%)	14.80%	33 (28.0%)	29.04%
Evaluation implementation or oversight	48 (40.7%)	38.46%	27 (22.9%)	24.28%	43 (36.4%)	37.26%
Initial briefing on findings	64 (54.2%)	53.83%	23 (19.5%)	21.54%	29 (24.6%)	24.63%
Draft report review	72 (61.0%)	57.08%	23 (19.5%)	19.88%	25 (21.2%)	23.05%
Final report acceptance or dissemination	68 (57.6%)	58.45%	23 (19.5%)	19.11%	27 (22.9%)	22.43%
Post-evaluation decisions or action planning	57 (48.3%)	47.89%	39 (33.1%)	34.60%	22 (18.6%)	17.52%

5) What type of USAID evaluation is this particular evaluation? (N=118)

	Number of Responses	Percent	Weighted Percent	Percent Excluding Don't Know (N=116)	Weighted Percent Excluding Don't Know (N=117)
Performance Evaluation	96	81.4%	83.3%	82.8%	83.3%
Impact Evaluation	11	9.3%	7.8%	9.5%	7.8%
Other	9	7.6%	7.9%	7.8%	7.9%
Don't know	2	1.7%	1.0%	—	

Other Responses:

- Combination but primarily a performance evaluation
- End of project evaluation
- Final project Evaluation
- Hybrid - lit review, key informants, focus groups, observations, qualitative data
- Mid-term performance evaluation
- Performance per the definition but we wanted it to go beyond the performance of current activities, but tell us the impact on economic growth of Mission investments in [Location]
- Post activity evaluation
- Study of a new paradigm “Decent Care”

Genesis of the Evaluation

6) When did USAID decide that an evaluation of the program, project(s), or activities discussed in this evaluation report would be undertaken? (N=118)

	Number of Responses	Percent	Weighted Percent	Percent Excluding Don't Know (N=106)	Weighted Percent Excluding Don't Know (N=106)
When planning the strategy for the Operating Unit or sector (e.g., in the PMP or CDCS)	13	11.0%	11.1%	12.3%	12.3%
When designing the project or activity (e.g., in an M&E Plan)	34	28.8%	29.9%	32.1%	33.1%
During the project or activity implementation	38	32.2%	30.9%	35.8%	34.3%
After the project or activity ended	13	11.0%	10.9%	12.3%	12.0%
Other	8	6.8%	7.5%	7.5%	8.3%
Don't know/can't recall	12	10.2%	9.8%	—	

Other Responses:

- [Implementing Partner], the [Activity] implementer decided to conduct the evaluation to complement the SHOPS mid-term evaluation. These were not required, but they were built into the [Activity] work plan.
- Before I joined the office
- By office policy, all [Office] programs undergo an external final evaluation.
- During PPR
- During PPR process
- Not sure, because it was a series of 1 year extensions that all preceded me.
- Upon winning an evaluation competition from PPL
- When USAID determined in @2011 that it could be helpful and appropriate to evaluate whether and how the estimated \$700 million in restoration assistance to northern [Country] after the [Conflict] conflict was an effective delivery of resources

7) What “triggered” or led USAID to initiate this evaluation? (Select all that apply)
(N=53, only those that responded “during” or “after project” in question 6)

	Number of Responses	Percent	Weighted Percent	Percent Excluding Don't Know (N=50)	Weighted Percent Excluding Don't Know (N=49-52) ⁴³
A key management decision was required, but there was not enough information	12	22.6%	10.2%	24.0%	24.3%
There were problems with cost-effectiveness or sustainability	2	3.8%	1.0%	4.0%	2.4%
The validity of Results Framework hypotheses or critical assumptions was questioned	3	5.7%	2.1%	6.0%	4.7%
Project review identified questions that need to be answered	10	18.9%	7.9%	20.0%	18.9%
Information was needed for other on-going or future programs	29	54.7%	22.7%	58.0%	51.6%
Feedback suggested that there were implementation problems or unmet needs	3	5.7%	2.6%	6.0%	5.9%
The project or activity produced unexpected results that required explanation	2	3.8%	1.0%	4.0%	2.4%
Other	13	24.5%	10.1%	16.0%	24.1%
Don't know/can't recall	3	5.7%	0.0%	—	—

Other Responses:

- Compliance with Evaluation Policy
- Detail evaluation described on page I of [Program Title] Program Evaluation report
- EG portfolio review needed
- Mission had chosen this project in light of new eval policy.
- Realization that it was required
- The Mission became more involved in evaluations in general due to changed USAID approach to evaluations overall
- The project activities were very successful and the Mission was thinking to validate the results and for granting extension to the project.
- The project was of the nature that results were not expected during the life of the project and therefore doing an evaluation a couple years post project made sense so we could actually understand the impact/follow on benefit of the project.
- This is a 20 years bilateral project. Government of [Country] expressed the interest in having an impact evaluation done on this project.

⁴³ Not every respondent answered every question, so the weighted N is 49 for some responses, and 52 for others.

- To assess the project effectiveness after realignment with the Feed The Future objectives during fourth year of implementation
- USAID Forward Guidance; and, Agency's Evaluation Policy
- Validate the management and implementation structure of [Project Title] - confirm fit for purpose to achieve the ambitious objectives
- Wanted to evaluate and document the successes/challenged of the end of the project

8) What process or person called for this evaluation? (Select all that apply)

(N=53, only those that responded “during” or “after project” in question 6)

	Number of Responses	Percent	Weighted Percent	Percent Excluding Don't Know (N=52)	Weighted Percent Excluding Don't Know (N=49)
Portfolio Review process	11	22.6%	9.1%	23.1%	24.1%
Performance Plan and Report (PPR) process (e.g., annual evaluation plan)	7	13.2%	5.4%	13.5%	13.0%
Mission or Office Director/Deputy	14	26.4%	10.5%	26.9%	25.0%
Program Office	19	35.8%	16.9%	36.5%	40.4%
DO Team Leader	8	15.1%	6.2%	15.4%	14.8%
AOR/COR for the Project(s) or activities	19	35.8%	17.8%	36.5%	42.6%
Other	4	7.5%	3.1%	7.7%	7.4%
No one in particular	1	1.9%	0.9%	1.9%	2.0%
Don't know/can't recall	1	1.9%	0.0%	—	—

Other Responses:

- [Program Title] Program also involved FAO & WHO as implementing partners, their activities on [Program Title] were evaluated
- Assistant Administrator directed that the study be done.
- IP recommended and concurred by DO Team leader
- It has been planned in Cooperative Agreement

9) Was this evaluation required, or was it something the Operating Unit undertook voluntarily?
(N=118)

	Number of Responses	Percent	Weighted Percent	Percent Excluding Don't Know (N=105)	Weight Percent Excluding Don't Know (N=103)
Required	47	39.8%	44.0%	44.8%	50.4%
Elective or non-required	56	47.5%	43.4%	53.3%	49.6%
Don't know/can't recall	15	12.7%	12.6%	—	—

10) If this evaluation was required, to what requirement did it respond?
(N = 47, only those that responded “required” in question 9)

	Number of Responses	Percent	Weighted Percent	Percent Excluding Don't Know (N=44)	Weighted Percent Excluding Don't Know (N=52)
Evaluate larger than average projects in a unit	27	57.4	54.6%	61.4%	60.9%
Evaluate pilot projects and innovative interventions	7	14.9	16.7%	15.9%	19.6%
Other Requirement	8	17.0	17.3%	18.2%	19.6%
Don't know/can't recall	5	10.6	11.4%	—	—

Other Responses:

- As per the Cooperative Agreement
- Assessment whether or not to expand this program into the scope of the [Program Title] program
- Gather information for future direction.
- Government of [Country]'s request
- Simply to conduct some evaluations
- This was a new mechanism and project, with partners we had not worked with directly ever, not this way, or not in a long time
- To determine the extent to which the [Location] images sensitization sessions led to changes in ordinary citizens??? knowledge and attitudes about decentralization and democratic politics more generally, and whether the sessions prompted greater cognitive
- To help decide on upcoming [Activity Title] activity

11) At which stages in the evaluation process were other stakeholders actively involved? (Select all that apply for each stakeholder) (N=118)

	Implementing Partner (Contractor / Grantee)	Weighted Percent	Country Partner	Weighted Percent	USAID Staff from Other Missions or Offices	Weighted Percent
Planning for the evaluation	64 (54.2%)	54.3%	25 (21.2%)	24.0%	70 (59.3%)	58.2%
Evaluation implementation or oversight	83 (70.3%)	71.0%	53 (44.9%)	43.3%	47 (39.8%)	38.7%
Initial briefing on findings	77 (65.3%)	68.6%	44 (37.3%)	38.6%	62 (52.5%)	51.8%
Draft report review	72 (61.0%)	65.2%	26 (22.0%)	23.7%	68 (57.6%)	56.7%
Final report acceptance or dissemination	69 (58.5%)	57.8%	35 (29.7%)	29.7%	72 (61.0%)	60.5%
Post-evaluation decisions or action planning	52 (44.1%)	44.0%	32 (27.1%)	26.8%	58 (49.2%)	48.2%

12) What do you feel was the quality of the participation in evaluation process of various types of stakeholders? (N=118, not all participants answered all rows)

	High Quality	Weighted Percent	Medium Quality	Weighted Percent	Low Quality	Weighted Percent	Don't Know	Weighted Percent
Implementing Partner	79 (66.9%)	69.4%	27 (22.9%)	21.7%	3 (2.5%)	2.0%	9 (7.6%)	6.9%
Country Partner	32 (27.1%)	28.2%	43 (36.4%)	37.3%	6 (5.1%)	4.4%	26 (22.0%)	21.0%
USAID Staff from Other	53 (44.9%)	47.4%	22 (18.6%)	16.7%	6 (5.1%)	3.7%	25 (21.2%)	21.2%

USAID Perceptions of the Evaluation Report

13) Are the following statements about this evaluation true? (Select all that apply) (N=118)

	Yes	Weighted Percent	No	Weighted Percent
The evaluation was designed to be relevant for upcoming decisions of the Operating Unit	111 (94.1%)	93.9%	7 (5.9%)	6.1%
It was delivered on schedule	99 (83.9%)	84.7%	19 (16.1%)	15.3%
The length of the report was appropriate	110 (93.2%)	93.0%	8 (6.8%)	7.0%
It was received in time to make decisions or for other intended uses	106 (89.8%)	90.7%	12 (10.2%)	9.3%
When the evaluation report was received, it was still considered relevant for upcoming decisions	106 (89.8%)	87.8%	12 (10.2%)	12.2%

14) You indicated above that the relevance of this evaluation changed between the time the evaluation was planned and when it was received. Please describe below what occurred that explains this difference. (N=5, optional question that only appeared when question 13 indicated a change in relevance)

- At the time, branch office in [Country] was about to be closed down. The evaluation purpose was to give clear picture of program's performance. However, the findings influenced the decisions on close-out and helped better plan of forthcoming activities.
- By the time the evaluation was underway, a decision had already been made to not fully fund the contract or continue in this field.
- The evaluator was too concentrated in matters that were not that relevant.
- This evaluation was a final evaluation of one of the Rule of Law Projects in the Mission. By the time we got evaluation results the Mission already had a new rule of law project ([Project Name]) up and running for around a year. The Mission made a decision not to work in judicial administration area and focus more on judicial independence (under the new [Project Type] project) prior the final evaluation would take place.
- While it did not affect the use of the recommendations completely, [Activity Title] was planned to be the first activity of a group of four [Activity Title]s to be evaluated. At some point during the planning stage it became clear that logistically that would be challenging given that it was too close to the winter time when the weather in the North would be limiting. The timeline was shifted to the right which meant that the report was produced way into the implementation of the option period.

15) To what degree did this evaluation meet the expectations of staff in your USAID Office (the office that sponsored the evaluation) on each of the dimensions listed below? (N=118)

	Excellent	Weighted Percent	Good	Weighted Percent	Inadequate	Weighted Percent	Can't Judge	Weighted Percent
Level of evaluation expertise on the evaluation team	40 (33.9%)	36.3%	53 (44.9%)	43.1%	14 (11.9%)	11.5%	11 (9.3%)	9.2%
Appropriateness of the study design and methods	36 (30.5%)	19.7%	68 (57.6%)	62.2%	4 (3.4%)	5.4%	10 (8.5%)	12.7%
Data quality and completeness	28 (23.7%)	13.7%	70 (59.3%)	62.9%	9 (7.6%)	7.3%	11 (9.3%)	16.2%
Clarity and coherence of the reporting	35 (29.7%)	26.4%	66 (55.9%)	46.1%	9 (7.6%)	16.7%	8 (6.8%)	10.8%
Adequacy and value of the data analysis provided	27 (22.9%)	19.2%	64 (54.2%)	58.0%	13 (11.0%)	10.7%	14 (11.9%)	12.1%
Clear links between findings, conclusions and recommendations	37 (31.4%)	25.9%	66 (55.9%)	63.9%	9 (7.6%)	1.4%	6 (5.1%)	8.9%
Overall credibility as a basis for learning or decision-making	35 (29.7%)	27.9%	69 (58.5%)	59.4%	9 (7.6%)	6.4%	5 (4.2%)	6.3%

16) Which of the following statements about this evaluation are true? (Select all that apply) (N=118)

	Number of Responses	Percent	Weighted Percent
Confirmed what USAID already thought it knew	85	72.0%	73.7%
Included important new information, not known to USAID	62	52.5%	51.6
Included gender specific findings when relevant	55	46.6%	49.0%
None of the above	7	5.9%	0.0%

17) Compared to pre-existing views on the performance of the project or activity at the time, how positive were the evaluation results? (N=118)

	Number of Responses	Percent	Weighted Percent
More Positive	31	26.3%	25.9%
As Expected	77	65.3%	65.9%
More Negative	10	8.5%	8.2%

18) Did this evaluation report include recommendations? (N=118)

	Number of Responses	Percent	Weighted Percent
Yes	114	96.6%	95.3%
No	4	3.4%	4.7%

19) Which of the following statements are true about the recommendations listed in the evaluation report? (Select all that apply) (N=114, question only appeared if question 18 was answered “yes”)

	Number of Responses	Percent	Weighted Percent	Percent Excluding None of the above (N=111)	Weighted Percent Excluding None of the above (N=112)
Easy to find in the report	103	90.4%	85.2%	92.8%	90.2%
Well organized in relation to questions and findings	86	75.4%	74.4%	77.5%	76.8%
Action-oriented, structured for use	71	62.3%	61.0%	64.0%	61.6%
Specific, not vague	71	62.3%	64.5%	64.0%	65.2%
Practical and/or feasible	66	57.9%	57.7%	59.5%	58.0%

	Number of Responses	Percent	Weighted Percent	Percent Excluding None of the above (N=111)	Weighted Percent Excluding None of the above (N=112)
Cost-conscious	21	18.4%	16.4%	18.9%	16.1%
Clear about who should take action	63	55.3%	54.7%	56.8%	54.5%
Well supported by evidence, findings and conclusions	58	49.2%	49.5%	52.3%	50.0%
None of the above	3	2.6%	0.0%	—	

20) On which of the following topics or issues did these recommendations focus? (Select all that apply) (N=114, question only appeared if question 18 was answered “yes”)

	Number of Responses	Percent	Weighted Percent	Percent Excluding None of the above (N=113)	Weighted Percent Excluding None of the above (N=112)
Design of current or future projects or activities	89	78.1%	76.3%	78.8%	76.8%
How the implementing partner manages the project	63	55.3%	59.1%	55.8%	58.9%
How USAID manages projects or activities (current or future)	43	37.7%	38.6%	38.1%	38.4%
How the project or activity could improve its performance	80	70.2%	68.3%	61.9%	61.0%
How gender-specific issues can be better addressed	35	30.7%	31.2%	31.0%	31.3%
How to better monitor projects or activities (current or future)	30	26.3%	26.3%	26.5%	26.8%
How to evaluate projects or activities (current or future)	16	14.0%	13.7%	14.2%	13.4%
Other	10	8.8%	11.2%	8.8%	11.6%
None of the above	1	0.9%	0.0%	—	

Other Responses:

- Focused on how our government counterpart could continue the project work or build on the lessons from the project, e.g. mainstreaming tools and strategies

- How should project target beneficiaries
- Possible future direction
- Project design
- Recommendations about how USAID managed the project were excised from drafts of the report
- Recommendations to the implementer on where to strengthen its own systems, use of data, tools
- Relations with host country partner
- Specific to possible program design
- The effectiveness of support for regional institutions and how to focus that support
- The validity of data and extension of the project.

21) How useful were the recommendations provided in this evaluation report?

(N=114, question only appeared if question 18 was answered “yes”)

	Number of Responses	Percent	Weighted Percent (N=114)
Not Useful	4	3.5%	3.8%
Moderately Useful	66	57.9%	58.1%
Highly Useful	44	38.6%	36.3%

22) For this particular evaluation, were the number of recommendations the evaluation team provided appropriate? (N=114, question only appeared if question 18 was answered “yes”)

	Number of Responses	Percent	Weighted Percent (N=112)
Too Few	5	4.4%	5.1%
About Right	95	83.3%	83.1%
Too Many	14	12.3%	11.8%

Dissemination of the Evaluation Results

23) How was each audience (listed in the columns below) exposed to the evaluation findings and recommendations? (Select all that apply) (N=118)

	Briefing prior to report completion	Weighted Percent	Final report or its executive summary	Weighted Percent	Dissemination on event or other briefing materials	Weighted Percent
USAID Staff	106 (89.8%)	88.0%	114 (96.6%)	96.9%	70 (59.3%)	62.6%
Implementing Partner Staff	77 (65.3%)	66.3%	90 (76.3%)	76.2%	48 (40.7%)	41.6%
Country Partner	37 (31.4%)	33.6%	50 (42.4%)	42.5%	38 (32.2%)	32.3%
Direct Beneficiaries	13 (11.0%)	11.5%	16 (13.6%)	12.7%	20 (16.9%)	16.8%
Other Donors/Stakeholders	14 (11.9%)	13.1%	27 (22.9%)	24.9%	29 (24.6%)	26.5%

24) Was any part of the evaluation report posted online? (Select all that apply) (N=118)

	Number of Responses	Percent	Weighted Percent	Percent Excluding Not to my knowledge (N=97)	Weighted Percent Excluding Not to my knowledge (N=118)
Full report was sent to USAID's Development Experience Clearinghouse (DEC)	90	76.3%	73.6%	92.8%	73.6%
Full report was posted on the Mission or Office website	11	9.3%	8.8%	11.3%	8.8%
Executive summary or one page briefer was posted on the Mission or Office Website	3	2.5%	2.1%	3.1%	2.1%
Other online location	8	6.8%	7.7%	8.2%	7.7%
Not to my knowledge	21	17.8%	20.4%	—	—

Other Responses:

- [Program Name] program's website
- A cable was prepared and circulated
- Circulated via email many times
- Implementing partner website

- Implementing partner website
- Program website
- Redacted version put on DEC
- The final report also is saved at public drive (P drive) at USAID [Mission]

25) If USAID received feedback from its partners or other stakeholders about this evaluation process or report, please describe what they told USAID about its value or usefulness to them.

(N=58)

- Both the Implementing partner and Government partners thought some components of the document could have used measured language to describe the findings.
- Can't recall the details but partner did submit a response to the Evaluation.
- Comment was request only for concurrence on the final report. No specific comment was requested on its value or usefulness to them.
- Do not know about the feedback from stakeholders.
- Donors and development partners interested in similar areas of intervention constantly asked for the report and often provided positive feedback on the usefulness of the report.
- Don't know.
- Don't know, also don't know answers to #19
- During briefing and presenting draft evaluation report, the implementing partners, government-country partners, and related stakeholders (CSOs) were provided an opportunity for clarification or correction to data/ findings from evaluation team. Most of them accepted the report and see the report as a good reference for future direction. They also addressed during the briefing that they found lesson learned and best practices from implementation of the project.
- Evaluation team considered implementing partner's suggestion as appropriate.
- Feedback reactions were numerous, all about concurrence and acceptance.
- Feedback was received from the implementing partner ([Implementing Partner]). They provided 5 comments to help clarify the evaluation report. Comments were general and related to the time-frame of the project; specificity of the context; evaluation methodology used and general comment on the [Country]'s transition process. In addition to that, implementing partner provided specific comments on election process in MNE, political party and parliamentary development and project staffing at the end, in order to give some more details on the project's performance.
- Found useful.
- Government of [Country] commented on the quality of the report. They felt the information on the flag ship government programs had factual errors, and they were not happy about the kind of data analysis that were presented.
- I do not know if USAID received feedback of any form.
- I don't know.
- Implementing partners did not feel all of the evaluation results were valid and that they were not adequately listened to by the evaluators when information was being collected.
- Including validation workshops with several groups of stakeholders was key for having a high quality report and increased the usefulness of the recommendations, increasing the level of stakeholders' appropriation of results to follow recommendations.
- [Implementing Partner] felt that some of the recommendations were applicable and valid, but that some were not applicable. [Implementing Partner] felt that they were not given the opportunity to discuss the findings with the evaluation team post the report. Therefore

recommendations were published in the report without room for discussion with the implementing partner ([Implementing Partner Name]).

- It is useful.
- It was useful because some definitions were reviewed regarding Complementary Package Plus and Minimal Package Plus of activity to be implemented in the health facilities.
- No.
- No feedback from stakeholders.
- No feedback received.
- No feedback.
- None.
- Not applicable.
- Not applicable. Due to the fact the evaluation was the project final evaluation and it was already known that the Mission was not going to continue working in the particular area the [Project Name] project would focus, the evaluation report did not get much attention of the stakeholders.
- Not aware if feedback was provided but the report informed the development of the new NRM Strategy.
- Not sure.
- Not to my knowledge.
- Numerous implementing partners submitted comments on the usefulness of the report and it has been referenced several times in other projects.
- Other donors that are engaged in the clinical legal education sector found the report useful and relevant.
- Our government counterparts appreciated the recommendations specific for them.
- Overall there was a dissatisfaction with the results. However, the partner failed to bring valid arguments in support of their disagreement with evaluation findings.
- Partner commented on evaluation and we considered most of the comments. Not much comment about value.
- Stakeholders commended USAID and the evaluation team for involving a wide array of stakeholders in gathering information for the evaluation and during the dissemination of preliminary findings.
- Stakeholders found the evaluation informative, but they felt that the IRI component did not tell enough if radio instruction is or not recommendable in [Country] context. Another concern was about the sustainability analysis of the self-directed teacher training methods: stakeholders wanted more insights for that.
- The COR and other staff who handled this are no longer at post. Hence unable to comment.
- The evaluation was an opportunity for partners to tell us how they believe future program can improve.
- The following feedback was received from implementing partner, including from volunteers working under this project: 1. The key recommendations were fair and reflect the real situation. 2. More efforts will be made to improve the project implementation and meet the recommendations. 3. The participation in this evaluation process was an excellent learning opportunity.
- The [Partner Country Government] stakeholders appreciated the findings of the report
- The host nation Ministry of Health had hoped we would include and expand this program into [Project Name] but the cost, personnel issues and sustainability of the program clarified in the study made USAID stay away from including this in the larger program.
- The implementer was very unhappy about the evaluation, as it viewed the report as instrumental in ceasing further USAID funding for it.

- The implementing partner took to heart the recommendations and we have seen a very positive change in a number of areas of work as a result (coordination with management team, greater emphasis and effort on gender mainstreaming, stronger results reporting, strengthening HMIS and use of data for decision making, etc).
- The Ministry of Finance (MoF) highly appreciated USAID third-party performance evaluation of the [Activity Name]. MoF managed the [Activity] which was direct G2G/on-budget assistance program funded by several donors.
- The report did not circulate outside of USAID as it contains procurement-sensitive information. It was used to design out new PAD for HIV.
- The [sic] will inform that the report was very useful for their project implementation.⁴⁴
- There was a validation activity with all partners and stakeholders (by groups and specific health sector) where the final report was presented. The participants were requested to validate data, conclusions and recommendations. This process was highly appreciated by them, and recommendations were improved.
- They accepted and appreciated the report.
- They did not have any objection about the report, they assumed all [was accurate], strength and weaknesses.
- This particular evaluation was not at all useful for two reasons: 1) strategic decisions had already been made in the time between when the evaluation was decided upon and when the contractor arrived. 2) it was of poor quality and even with improvements by the contractor was of no added value although minimally met standards.
- This process was useful for the partner, as it provided an opportunity to get feedback from the ministry of health in a structured and constructive way on how they could link better and report to the ministry regularly through various platforms.
- This was end of project evaluation. By the time the project... [sic]⁴⁵
- Three USAID/[Mission] implementing partners that were subject of the evaluation submitted formal Statement of Difference(s); USAID/[Mission] enclosed them (as Annexes 1, 2, and 3) with the Final Report that was posted to USAID's DEC (note: the latter is the same with information I received in the introductory email – [URL])
- USAID did not receive any comment from its partners.
- USAID did not request stakeholders feedback
- We have received many requests for copies of the evaluation, which many have found enlightening.
- We shared finding widely with local northern [Country] development partners as well as Office of the Prime Minister here in [City] who had detailed a statistician to join the evaluation team, but never did we receive feedback.

26) Were you involved in the evaluation process in some way? (e.g., SOW development, evaluation team member, data collection, draft report peer review, etc.) (N=118)

	Number of Responses	Percent	Weighted Percent
Yes	92	78.0%	76.7%
No	26	22.0%	23.3%

⁴⁴ Not all respondents provided complete responses to open-ended questions with free-form response options.

⁴⁵ Ibid.

27) How did exposure to this evaluation affect what you know or your views about the program, project or activity that was evaluated? (N=118, not all participants responded to all parts of this question)

	Evaluation Process	Weighted Percent	Evaluation Report	Weighted Percent	Both	Weighted Percent	Neither/Not Affected	Weighted Percent
My opinion of the merits of the project or activities	5 (4.2%)	4.2%	40 (33.9%)	35.4%	63 (53.4%)	51.8%	10 (8.5%)	8.6%
My views about the soundness or adequacy of the design	9 (7.6%)	8.1%	28 (23.7%)	26.3%	66 (55.9%)	54.7%	12 (10.2%)	10.9%
Understanding of schedule and budget problems beyond the implementing partner's control	8 (6.8%)	8.5%	18 (15.3%)	15.8%	33 (28.0%)	28.3%	52 (44.1%)	47.5%
Why some results were or were not achieved	4 (3.4%)	3.3%	48 (40.7%)	41.9%	39 (33.1%)	30.8%	23 (19.2%)	23.9%
The way I think about other activities in the same sector or topic	3 (2.5%)	2.8%	39 (33.1%)	34.2%	52 (44.1%)	44.8%	21 (17.8%)	18.2%
The way I think about partner collaboration in the Program Cycle	7 (5.9%)	5.7%	23 (19.5%)	20.9%	43 (36.4%)	37.5%	40 (33.9%)	35.9%
Understanding of the unplanned consequences of projects like this	2 (1.7%)	2.0%	31 (26.3%)	25.3%	40 (33.9%)	30.5%	43 (36.4%)	42.2%
Understanding of how to make this type of project more effective in the future	2 (1.7%)	2.0%	38 (32.2%)	36.0%	62 (52.5%)	50.6%	13 (11.0%)	11.4%
Other	1 (0.8%)	2.5%	0 (0.0%)	0.0%	5 (4.2%)	4.2%	13 (11.0%)	13.4%

Other Responses:

- The challenges of not having a consistent mechanism to execute project evaluations. The greatest challenges we experienced in this evaluation was the absence of a mechanism, the use of contract bridges (we worked over 2 contract bridges), which greatly delayed this process and then the report product, and we weren't as able to use the findings to inform new designs.
- Understanding why project design, PMP, and M&E are important and associated to achieve goals for the project.

28) How did exposure to this evaluation affect your knowledge or views about evaluation?

(N=118, not all participants responded to all parts of this question)

	Evaluation Process	Weighted Percent	Evaluation Report	Weighted Percent	Both	Weighted Percent	Neither/Not Affected	Weighted Percent
My understanding of the purpose and role of evaluation	17 (14.4%)	15.4%	18 (15.3%)	16.3%	54 (45.8%)	45.6%	27 (22.9%)	22.7%
My understanding of evaluation types and methods	20 (16.9%)	18.4%	14 (11.9%)	11.6%	51 (43.2%)	40.2%	32 (27.1%)	29.8%
Understanding of how to oversee an evaluation team	22 (18.6%)	19.3%	6 (5.1%)	5.2%	50 (42.4%)	40.6%	39 (33.1%)	34.9%
Ability to engage stakeholders in an evaluation processes	20 (16.9%)	17.6%	16 (13.6%)	14.2%	49 (41.5%)	40.8%	31 (26.3%)	27.4%
Ability to review or critique evaluation products	7 (5.9%)	8.3%	29 (24.6%)	24.2%	55 (46.6%)	43.0%	26 (22.0%)	24.5%
Ability to lead or support post-evaluation action planning and follow-up processes	8 (6.8%)	7.9%	23 (19.5%)	19.5%	52 (44.1%)	43.2%	34 (28.9%)	29.4%
My interest in monitoring and evaluation	8 (6.8%)	7.9%	12 (10.2%)	9.6%	59 (50.0%)	50.9%	38 (32.2%)	31.5%
Other	1 (0.8%)	2.5%	2 (1.7%)	1.4%	7 (5.9%)	4.6%	14 (11.9%)	13.8%

Other Responses:

- My understanding how M&E is crucial for program
- My understanding of the purpose and role of evaluation within USAID
- The value of our office's standing team that reviews evaluation SOWs and helps with the design. It was a stronger evaluation because of this body

USAID Post-Evaluation Review

29) Did USAID organize a post-evaluation review meeting to decide how it would respond to the evaluation's findings and/or recommendations? (N=118)

	Number of Responses	Percent	Weighted Percent	Percent Excluding Can't Recall (N=95)	Weighted Percent Excluding Can't Recall (N=93)
Yes	62	52.5%	53.4%	65.3%	67.7%
No	33	28.0%	25.4%	34.7%	32.2%
Can't Recall	23	19.5%	21.2%	—	—

30) How long after this evaluation was completed did USAID hold the evaluation review or action planning meeting for this evaluation? (N=62, question only appeared if "yes" was selected in question 29)

	Number of Responses	Percent	Weighted Percent
Less than 1 month	30	48.4%	50.1%
1-3 months	25	40.3%	36.5%
3-6 months	3	4.8%	5.5%
6-12 months	3	4.8%	5.1%
More than 1 year	1	1.6%	2.8%

31) Who chaired the post-evaluation review meeting? (N=62, question only appeared if "yes" was selected in question 29)

	Number of Responses	Percent	Weighted Percent	Percent Excluding Don't Know (N=57)	Weighted Percent Excluding Don't Know (N=57)
Mission or Office Director/Deputy	6	9.7%	8.4%	10.5%	9.4%
Program Office Director/Deputy	5	8.1%	7.7%	8.8%	8.6%
Mission M&E Officer	8	12.9%	13.2%	14.0%	14.7%
DO Team Leader	10	16.1%	15.4%	17.5%	17.1%
AOR/COR for the Project(s) or activities	19	30.6%	29.8%	33.3%	33.2%
DO Team M&E Coordinator	2	3.2%	2.3%	3.5%	2.5%
Other	7	11.3%	13.0%	12.3%	14.5%
Don't know/can't recall	5	8.1%	10.2%	—	—

Other Responses:

- A team consisting of all M&E officers in all the DOs and Program Office
- AOR with Division Chief
- Assistant Administrator
- Cannot remember
- More informal meeting/ discussion
- Was discussed during portfolio review
- Was discussed during portfolio review

32) Which of the following decisions were made in the post-evaluation review meeting? (Select all that apply) (N=62, question only appeared if “yes” was selected in question 29)

	Number of Responses	Percent	Weighted Percent	Percent Excluding None of the above (N=61)	Weighted Percent Excluding None of the above (N=61)
Which evaluation findings were most important for USAID to focus on	46	74.2%	70.3%	75.4%	71.4%
Which recommendations to accept and act on	40	64.5%	63.4%	65.6%	61.5%
Who would be responsible for taking action based on decisions from this post-evaluation review	28	45.2%	46.1%	45.9%	43.8%
Timeframes or deadlines for completing those actions	22	35.5%	32.8%	36.1%	33.3%
Whether and how USAID would follow up on the implementation of those actions	36	58.1%	56.2%	59.0%	57.1%
Other specific steps to learn from this evaluation	18	29.0%	26.8%	29.5%	27.2%
Who would prepare and/or monitor a written version of the post-evaluation Action Plan	15	24.2%	22.5%	25.6%	22.9%
Other	7	11.3%	12.3%	11.5%	12.5%
None of the above	1	1.6%	0.0%	—	—

Other Responses:

- Can't recall since it could happen in 2011 or 2012
- How would implementation of recommendations be funded

- Instead of a post-e action plan, specific recommendations were incorporated in each IM's action plans, as adjustments
- No decision is made as the Ambassador decided to cut funding for private sector activities
- Terminate the Project
- The evaluation data, findings and recommendations supported planning of [Activity Title] II a follow-on activity to [Activity Title]. Again, I wanted to note that USAID did not request or participate in planning this evaluation. It was planned by the implementing partner who...[sic]⁴⁶
- There were two parallel evaluations on the basis of which the project was granted extension.

33) Was a post-evaluation review report or Action Plan written? (N=62, question only appeared if “yes” was selected in question 29)

	Number of Responses	Percent	Weighted Percent	Percent Excluding Can't Recall (N=54)	Weighted Percent Excluding Can't Recall (N=63)
Yes	23	37.1%	32.7%	42.6%	38.1%
No	31	50.0%	54.1%	57.4%	61.8%
Can't Recall	8	12.9%	13.2%	—	—

34) To whom were copies of the post-evaluation review meeting report or Action Plan distributed? (Select all that apply) (N=23, question only appeared if “yes” was selected in question 33)

	Number of Responses	Percent	Weighted Percent
Mission or Office Director/ Deputy	8	34.8%	36.2%
Program Office Director/ Deputy	14	60.9%	63.2%
Mission or Office M&E Officer	9	39.1%	46.4%
DO Team Leader	12	52.2%	53.8%
AOR/COR for the Project(s) or activities	19	82.6%	90.9%
DO Team M&E Coordinator	5	21.7%	24.7%
Other	5	21.7%	19.3%
None of the above	0	0.0%	0.0%

Other Responses:

- ENV Office Director/Deputy, implementing partners
- Everyone in the Office
- I don't know
- Implementing partners, [Local Task Force]
- Posted publicly on pmi.gov, disseminated through PMI list serve

⁴⁶ Not all respondents fully answered the open-ended survey questions in the free-form response.

Status of Recommendations

35) Approximately how many recommendations were included in this evaluation report?
(N=118)

	Number of Responses	Percent	Weighted Percent
None	2	1.7%	2.1%
1-5	23	19.5%	21.4%
6-10	28	23.7%	23.0%
11-15	27	22.9%	22.7%
16-25	19	16.1%	15.8%
More than 25	9	7.6%	6.2%
Recommendations exist but not in countable format	10	8.5%	8.9%

36) About what portion of these recommendations did USAID adopt or otherwise decide to take action on? (N=116, question did not appear if “none” is selected in question 35)

Percent of Recommendations Accepted	Number of Responses	Percent	Weighted Percent (N=116)
100%	11	9.5%	10.1%
75% to 99%	24	20.7%	20.7%
50% to 74%	32	27.6%	28.7%
25% to 49%	30	25.9%	23.5%
Less than 25%	9	7.8%	7.4%
None	10	8.6%	9.7%

37) If some of the evaluation recommendations were not accepted for action by USAID, why were those recommendations not accepted? (Select all that apply) (N=105, question does not appear if “100%” is selected in question 36)

	Number of Responses	Percent	Weighted Percent (N=104)
Not directed to USAID	31	29.5%	32.8%
Not action-oriented	18	17.1%	16.0%
Not timely – not relevant to the time the report was received	21	20.0%	20.3%
Not relevant to a changed context around the project or activity	29	27.6%	26.3%
Too vague – could not understand what action was supposed to be taken	14	13.3%	11.1%
Too complicated – capacity to implement them was lacking	10	9.5%	9.5%
Too costly to implement	22	21.0%	18.0%
Too many recommendations	9	8.6%	7.7%
Insufficient evidence to be compelling	17	16.2%	14.9%
Insufficient interest or political will to pursue action	19	18.1%	18.1%
Not possible within USAID's constraints or practices	37	35.2%	34.4%
Other	21	20.0%	18.7%

Other Responses:

- As requested by SOW, the evaluation report had recommendations for USAID and other stakeholders. For USAID there were 5, MOH 8, [Other Donor] 5, donors 5 and other sectors 3. Approximately 80% of them have been accomplished to date.
- Certain senior Mission official(s) had rather fixed views, despite the evidence
- Change in USAID's programmatic focus
- Don't know how many adopted
- Embedding consultants hired by the [Project Name] project in [Partner Country Government] agencies was not implemented. USAID thought that they will not be used solely for project purpose but for additional work as assigned by the government agency and out of their defined scope of work.
- Funding priorities
- I'm not entirely sure, but these are my best guesses.
- It was Ambassadors decision to cut future funding for private sector project
- Mission cut the follow on program due to staffing constraints
- Mostly an issue of our own internal capacity - our team still needs to find the time to act on some of the recommendations
- No formal process to accept or reject recommendations
- Not seen as feasible or as the priority at the time in reviewing partner country's feasibility and priorities

- Procurement action to actualize the recommendation is not yet completed.
- Some recommendations were very obvious
- Some were beyond USAID manageable interests
- Specific to that activity, which was ending; could possibly inform design in future but not certain
- The Health Program terminated
- They were not relevant considering the Mission's decision to terminate the Project.
- This programming is not completely controlled by USAID - this programming is political and dictated in part by State (which funds the activity) and [Regional Intergovernmental Organization]

38) Of the recommendations accepted for action, what portion have been fully implemented or acted upon? (N=105, question does not appear if “none” is selected in question 36)

Percent of Accepted Recommendations Implemented	Number of Responses	Percent	Weighted Percent (N=104)
100%	21	20.0%	22.3%
75% to 99%	33	31.4%	31.6%
50% to 74%	23	21.9%	24.6%
25% to 49%	10	9.5%	8.4%
Less than 25%	17	16.2%	12.1%
None	1	9.5%	1.0%

39) If some of the recommendations that were adopted for action have not been fully implemented, why has action on them not been completed? (Select all that apply) (N=84, question does not appear if “100%” is selected in question 38)

	Number of Responses	Percent	Weighted Percent (N=104)
Time – it is too soon after decisions were made to expect actions to be completed, or too much time has already passed and implementation is no longer likely	28	33.3%	39.0%
Availability of funds to take the steps the recommendation requires	29	34.5%	32.4%
Capacity – skills, knowledge, experience needed to implement the recommendations is not available	7	8.3%	8.6%
Relevance – the context or environment changed	40	47.6%	45.1%
Lacking an advocate in the Mission/Office for seeing it through to full implementation	8	9.5%	3.8%
Resistance	3	3.6%	4.7%
Other	12	14.3%	14.7%

Other Responses:

- [Country] government capacity development is a long long-term process that will take at least 2 generations
- Again, I'm not sure, but these are my best guesses
- All recommendations had been taken into account
- Changes introduced into a new solicitation. Awaiting award in order to operationalize the recommendations.
- Don't know how many implemented
- Findings are being incorporated into a new project design, which has taken a long time
- Funding
- New procurement to actualize most of the recommendations is still in process
- Not sure
- Recommendations are related to the partner itself and up to them to improve
- Recommendations were used in another Local Governance Program
- Relevant for follow on activity, which has not been designed yet

40) Please explain the resistance that prevented recommendations from being implemented. (N=3, question only appeared if "resistance" was selected in question 39)

- Recommendation to stop further issue of Small Grant RFAs as well as the commitment of further Small Grant funds and use the unspent funds to implement activities more clearly focused on MRP Objectives. It was hard at the beginning to convince implementing partners to stop issuing the small grants as they had commitments already, and they kept proposing poor quality of small grants proposal but USAID did not approve it.
- Technical team feels it doesn't have sufficient leverage with the partner to make changes to existing grants. Lack of capacity of technical team to monitor program implementation sufficiently.
- Resistance from some key stakeholders to continue receiving certain types of assistance (For example, preference to receive funding instead of TA).

41) If USAID were to implement a policy for how many recommendations should be included in an evaluation report, what do you think is an appropriate number? (This was an optional question.)

Number of Recommendations	Minimum	Weighted Percent	Maximum	Weighted Percent
0	2 (1.9%)	1.2%	—	—
1	13 (9.3%)	11.2%	—	—
2	12 (11.2%)	11.6%	—	—
3	28 (26.2%)	21.2%	9 (8.1%)	7.4%
4	4 (3.7%)	4.0%	1 (0.9%)	0.9%
5	37 (34.6%)	32.7%	19 (17.1%)	16.2%
6	1 (0.9%)	0.9%	4 (3.6%)	3.3%
7	4 (3.7%)	2.9%	5 (4.5%)	4.1%
8	1 (0.9%)	0.3%	10 (9.0%)	8.6%
9	0 (0.0%)	0.0%	1 (0.9%)	1.0%
10	5 (4.7%)	4.7%	32 (28.9%)	28.9%
12	—	—	5 (4.5%)	3.2%
15	—	—	12 (10.8%)	10.7%
17	—	—	2 (1.8%)	1.7%
20	—	—	8 (7.2%)	7.0%
25	—	—	2 (1.8%)	1.2%
100	—	—	1 (0.9%)	0.2%

Descriptive Statistics		
N	111	115
Valid N*	107	111
Range	0 – 10	3 – 100
Mean	3.9	10.6
Std. Deviation	2.2	9.9
Mode	5	10

*Not all participants provided responses in a useable, numeric format

Actions and Changes Made in Response to this Evaluation

42) Were any specific actions taken/changes made in response to this evaluation's findings? (N=2, question only appears if "none" was selected in question 35)

	Number of Responses	Percent	Weighted Percent	Percent Excluding Don't know (N=1)	Weighted Percent Excluding Don't know (N=1)
Yes	0	0.0%	0.0%	0.0%	0.0%
No	1	50.0%	50.0%	100.0%	50.0%
Don't know	1	50.0%	50.0%	—	50.0%

43) Which of the following actions were taken based on the findings or recommendations? (Select all that apply) (N=102, question does not appear if "none" is selected in question 36 OR if "no" or "don't know" are selected in question 42; due to a survey error, 4 participants that should have responded to this question did not)

	Number of Responses	Percent	Weighted Percent	Percent Excluding None of the above (N=92)	Weighted Percent Excluding Don't know (N=102)
Modified the project or activity	29	28.4%	26.3%	31.5%	30.4%
Used to design a direct follow-on project or activity	49	48.0%	42.7%	53.3%	49.3%
Used to design a new activity in the same sector or area (not a direct follow on)	36	35.3%	27.7%	39.1%	32.0%
Changed the way USAID manages projects or activities (e.g., type of partner)	14	13.7%	12.2%	15.2%	14.1%
Used to revise or develop a strategy for this country, region or office	22	21.6%	17.8%	23.9%	20.6%
Used to revise or develop a USAID policy for this sector, particular problem or topic	14	13.7%	12.0%	15.2%	13.8%
Other*	1	1.0%	0.9%	1.1%	1.0%
None of the above	10	9.8%	13.3%	—	—

*No descriptions of other activities were given by the participant

44) What most influenced decisions to take action or make changes? (Select all that apply) (N=106, question does not appear if "none" is selected in question 36 OR if "no" or "don't know" are selected in question 42)

	Number of Responses	Percent	Weighted Percent	Percent Excluding None of the above (N=103)	Weighted Percent Excluding Don't know (N=102)
USAID staff participation in the evaluation processes	34	32.0%	28.4%	33.0%	32.2%
Strength of the evidence presented in the report	59	55.7%	48.9%	57.3%	55.4%
Strong USAID backing for action on the evaluation findings/recommendations	41	38.7%	33.4%	39.8%	37.8%
Appropriateness of the evaluation methods employed	20	18.9%	14.6%	19.4%	16.5%
Informal discussions with colleagues and partners about the report	25	23.6%	19.8%	24.3%	22.3%
Usefulness of the Executive Summary	15	14.2%	10.8%	14.6%	12.2%
Formal post-evaluation meeting(s) on evaluation's implications and actions to be taken	30	28.3%	25.9%	29.1%	29.3%
Recommendations that were evidence-based, specific and actionable	60	56.6%	50.1%	58.3%	56.7%
Other	4	3.8%	3.0%	3.9%	3.4%
None of the above	3	2.8%	11.6%	—	—

Other Responses:

- An accompanying economic growth assessment
- Documentation of government option
- New project designed afterwards
- Strong sector background of the evaluation team

45) Over how many months after this evaluation was completed did your Office/Mission continue to actively learn from or take actions based on this evaluation? (N=115, optional question)

	Number of Responses	Percent	Weighted Percent	Percent Excluding Don't Know (N=92)	Weighted Percent Excluding Don't Know (N=90)
3 months	24	20.9%	20.5%	26.1%	26.6%
6 months	14	12.2%	9.5%	15.2%	15.5%
12 months	31	27.0%	28.6%	33.7%	34.4%

24 months	8	7.0%	6.2%	8.7%	8.8%
More than 24 months	15	13%	13.4%	16.3%	16.6%
Don't know	23	20.0%	21.8%	—	—

Impact of Learning and Actions Taken Based on this Evaluation

46) Did actions taken on the findings or recommendations of this evaluation improve USAID's overall development effectiveness (e.g., change the status of outcomes at the country or regional level)?

(N=118)

	Number of Responses	Percent	Weighted Percent	Percent Excluding Don't Know (N=64)	Weighted Percent Excluding Don't Know (N=66)
Yes	47	39.8%	41.6%	73.4%	74.2%
No	17	14.4%	14.5%	26.6%	25.8%
Don't know	54	45.8%	43.9%	—	—

47) What positive results occurred as a result of taking action that might have improved USAID's overall development effectiveness? (Select all that apply) (N=47, question only appears if “yes” is selected in question 46)

	Number of Responses	Percent	Weighted Percent (N=47)
Effectiveness of existing USAID programs, projects, activities has improved	29	61.7%	63.8%
Cost-effectiveness of existing USAID programs, project, or activities has improved	9	19.1%	16.8%
Sustainability of existing USAID programs/projects/activities has improved	23	48.9%	47.5%
Results from new program/project strategies built on the evaluation are positive	23	48.9%	47.6%
Implementing partners have taken complementary actions with positive results	22	46.8%	48.0%
Partner government strategies and/or policies reflect learning from the evaluation	9	19.1%	19.2%
Other donor programs/projects have applied lessons from this evaluation	8	17.0%	19.0%
Other	1	2.1%	2.1%
None of the above	0	0.0%	0.0%

Other Responses:

- Informed new project design

Stories about the Benefits of Evaluations

(N=12 participants, 18 benefits stories; optional question)

- After the Evaluation, we understood that some expected results were not achievable.
- After this Evaluation, the implementing partner which had staff in the remote part of the country was obliged to have a representative staff in the head country of the region to have more visibility and dialog with provincial authorities.
- Assessment of successful pilot studies such as this identify that the costs sometimes cannot be sustained to reap the obvious benefits
- Findings and recommendations from this evaluation were fully used for the design of a new PAD for HIV which covers the period of five years, 2012-2017. This new PAD was endorsed by the Mission and used as a strategic guidance for the program.
- [Project Name]: The project's work plan was developed to strengthen support groups to serve as volunteers supporting various components of the project in the facility and community. The project built their the capacity to enable them take part in some parts of service delivery such as HIV Counseling, and testing, medication adherence support and defaulter tracking. This has resulted in improved retention and adherence to treatment.
- In the second phase the implementer took improvement from what's the evaluation recommendation, such as in the area selection, working closely with country partners

- Included hard targets for banks to encourage loans to female beneficiaries
- Lesson learned and best practices conducted by an implementing partner on local community and CSO engagements have been adopted by another partner on this program.
- One evaluation finding was 'A specific budget line for institutional capacity building should be introduced to encourage partners to make better use of this grant to develop internal capacity (a specific objective of the USAID Development Grants Program.) In the past, with awards to local partners (whether funded by the DGP program or not) we helped them build their capacity to implement development programs and increase their impact by 1) offering USAID staff as resources and 2) using a grant with [Implementing Partner] to provide targeted capacity building support. Over time, it became clear that local partners preferred to manage their own capacity building resources and find their own assistance. So, as a result of the findings of this evaluation and feedback from our partners, now we put extra funding in their award for them to directly manage to obtain their own capacity building support (for example, help to improve their M&E approach, their organizational sustainability, their financial management practices, etc.)
- One important finding: local government funding support on participatory disease surveillance and response (PDSR) activity decreased. Responding this finding, USAID discussed to the implementing partner to intensify their advocacy to the local government (districts) for more funding allocation for PDSR
- Since the [Program Name] in [Country] involving cross sector line ministries including Ministry of Agriculture, Ministry of Health, Ministry of Trade, and Coordinating Ministry of People Welfare, many stakeholders, CSOs and communities, and implementing by FAO, WHO, international and national partners, with activities covered human health, animal health, public awareness, health promotion, poultry value chain, the evaluation recognized how these cross sectoral approaches and comprehensive activities are important aspects to prevent and control AI in [Country].
- Targeted capacity building was an action resulting from the evaluation findings as general capacity building was not effective
- The assessment team recommended the uninterrupted continuation of the project's support for the [National Council]. The project continued to support the [National Council], and as a result of USAID's technical assistance the [National Council] is the government main partner for public-private dialogue (PPD). It represents the voice of the private sector serving as a member on a few committees, and advocates for policy changes that will contribute to greater economic development.
- The experience with this project and the evaluation that describes the implementation process has been critical in informing the Mission about how to design future activities.
- The follow on activity included more direct relationship with partner banks as well as counterpart donor [Other Donor].
- The other key impacts of this evaluation includes: a) Most technical recommendations were used to shape the focus of the new project as well as make changes to existing activities, b) [Country] PEPFAR program has completely transitioned out of the direct service delivery (transition to Global Fund) and become a TA focus program, c) Two new projects were designed and awarded (The HIV Flagship and the HIV innovate and evaluate projects).
- This specific award was for a Development Grants Program (DGP), a first-timer partner USAID awardee. The implementer really learned about integrating the topics of climate change adaptation and disaster risk reduction, at the same time about fulfilling USAID requirements. The implementer worked hard in the program implementation and improving the organization management. The evaluation recommended some follow up action to make the program more sustainable. As the program works in a prone disaster areas, [Office] was able to fund the

second phase of this DGP award, it's also the areas (eastern Indonesia) where [Office] and the Mission will work in the next five years

- We used this evaluation as does the U.S. Ambassador to show the need to fund elections through the electoral cycle. We still use this evaluation 3 years later as the recommendations help us make our case for more funding.

May we contact you to follow-up on any of these impact stories? (N=12)

	Number of Responses	Percent	Weighted Percent (N=14)
Yes	7	58.3%	44.4%

Other Tools and Approaches for Improving Evaluation Utilization

(N=21, optional question)

- After evaluation, tracking tables are attached to project regular reports.
- Already mentioned in the previous survey.
- Discussion with staff from the [City] Mission's Office of Democracy and Governance indicated that the evaluation report process took a lot of work to get this product into a form that was minimally acceptable. Moreover, the [City] ODG has undertaken three evaluations, one prior to this report and one after. The consensus was that all three lacked of quality and timeliness, and that there is a flaw in the way the products are commissioned and developed. A review should be undertaken with the Office of Program Development (OPPD) in [City] to determine how the process can be improved, so that better quality analysis and reports can be produced.
- Evaluation recommendation implementation action plan will be submitted to [MSI Staff]
- For now, I do not know of such instrument. If I run across one, I would send it to [MSI Staff].
- For USAID recommendations: After each evaluation, we organize specific meetings with implementing mechanisms involved, and analyze each recommendation, identifying the expected contribution (alone or in coordination) of each one of them, to accomplish each recommendation. We define responsible resources and time line, and adjust IM's action plans in accordance. For non-USAID recommendations: in quarterly meetings we revised other sector's advances related to their recommendations. Since USAID graduated its cooperation on FP and MCH, we can only provide direct follow-up to HIV recommendations.
- Hello [MSI Staff], I am sending (as a Word document attachment) to your email account the [Mission] portion of the 'USAID Forward and Evaluations (thru 1/31/13) Quality Review Reports', which PPL shared with our Mission in the past; the latter covers the evaluation subject of this questionnaire (among other USAID/[Mission] evaluations). Hope you find this useful. Cheers!
- I am not aware of such tool at our Mission.
- In preparation for the evaluation assignment, as a COR, I asked the project to do a non-biased self-evaluation of their performance. I was glad to see that their findings were almost identical with those from the evaluation team. Note: The internal self-assessment report has not been shared with the evaluation team. I would suggest that to the extent possible, the implementing partners to think of the evaluation as a "corrective" action rather than a "punishment" by taking a proactive approach in identifying their strengths and weaknesses.
- Post evaluation report review tables.
- Technical Working Group (TWG) composed of all health implementing partners with country partner meet to discuss learning from the evaluation.

- The ability to share and later discuss the report with host country partners proved to be a valuable tool for the activity management team. The exercise was successful because it created an interaction point that reviewed the report's recommendations to ensure were understood by the partners and action would be taken in not in progress already
- The Mission uses one of the two annual portfolio reviews (the detailed portfolio review) to review progress in implementation of evaluation findings. This has proved effective in ensuring that DO team draw up evaluation recommendation implementation plans (example emailed to [MSI Staff]) and AORs keep following up on progress.
- The timing of the evaluation should be that it should inform new project or activity design hence the need to spell out the type and time for conducting evaluations in our Project and IP M&E Plans
- USAID/[Mission] is using Dashboard for their M&E capture tool that is created couple of months ago.
- We are about to engage in an evaluation that will use the utilization-focused approach.
- We developed chart for utilization and will send to [MSI Staff], also evaluation reports and findings were part of portfolio review
- We have those tools mentioned above, but there's been a lack of commitment and interest from senior management to follow up the recommendations and make sure that relevant technical office monitor and report the progress.
- While utilization is super important, we are still challenged in getting good structure around evaluations. In the belief the Agency's own Eval policy has still not been directive enough in these matters, USAID/[Mission] created supplemental guidance we use in post-award orientation meetings and when discussing the quality of reports expected: [Pasted evaluation guidance into the survey]
- Working groups after evaluations are done to strengthen change and adaptation of programs to recommendations.
- Yes, for the tracking of high quality evaluations I've created High Quality Evaluation Matrix, very simple excel spreadsheet but useful when you need info about general trends. It is not perfect cause it is based on the subjective estimation of post-evaluation review team members, however it gives some useful information. Could be improved. I would be more than glad to share with you if needed.

Additional Comments

(N=31, optional question)

- Application of Theory of Change in biodiversity funded projects or those with blended funding has to be made more popularized
- As previously stated, the past three evaluation processes related to Rule of Law have been less than satisfactory. A discussion needs to take place with OPPD to determine what steps can be taken in order to improve the timeliness and quality of these evaluations. There are several issues that at first glance need to be addressed (and others may need to be after more careful consideration): 1) There were challenges in getting the contractor to deliver a quality report -- some of these appear to be related to the contractor's capacity, but others were related to the communication process among the contractor, OPPD, and ODG. 2) Related to this, one year assignments and frequent R&Rs in Afghanistan affect the Mission's institutional capacity and knowledge. Some efforts to come up with work-arounds needs to be considered,
- At the time [Activity Name] was planned the new USAID Evaluation Policy was not in place and some of the questions did not apply. Some questions were asked in a way that did to generate the right response. For example question 29b does not leave room for situation where some

recommendations were adopted by USAID but the proportion is unknown, there should be option of don't know.

- Evaluation of USAID-funded projects is very important and should be planned from the inception of the project.
- For some reason implementing partner was reluctant to cooperate in full capacity.
- For this type of strategy evaluation, it was very important to have a mixed evaluation team (external contractors and USAID staff Mission and WDC). We were able to replicate this evaluation design to other strategies evaluation (MCH, FP and HIV, and health program). We are still learning from the lessons identified, and sharing with other national, regional and international stakeholders, and applying them to current HIV programming and sustainability plans.
- I might not have answered some of the questions accurately especially with regards to the percentages. I decided to ignore them but the questionnaire will not allow. There are a few I do not recall the answers. USAID had input into the pre solicitation SOW.
- In this specific evaluation we had a lot of problems with the first evaluation team (international/national). They were not able to comply with the SoW and to provide a good draft report. The mechanism was not able to provide an appropriate follow up of the evaluation deliverables, including reading drafts in Spanish or assuring compliance with the SoW. Even though we did not accept the draft report, all the funds were used. The Mission had to find additional funds to pay for a second evaluation team (national) to complete the evaluation. All the data had to be revalidated and the report had to be rewritten. At the end, the result of this second process was good, but the overall process for this evaluation was not good. We invested a lot of time and financial resources in an evaluation mechanism that did not guarantee the quality of the process and the product.
- It is quite a challenge to find a good evaluation team, if there are database of consultants or companies that are familiar with USAID evaluation that would be helpful.
- [Activity Name] (in [Country]) Evaluation was conducted in the period when [Country] Office (branch office of the USAID [Mission] at the time) was about to be closed. That process went faster than expected and that is why the Mission didn't act in accordance with recommendations in most of the cases. Speaking about lack of actionable recommendations in the evaluation report, most probably the prominence of expected closure was the reason why some of the findings were not transferred into the actionable recommendations.
- Less questions. Many of the questions seemed to be almost repetitive or too similar in nature.
- Mid-term evaluations seem best suited for USAID context. They can be beneficial both for ongoing activities/improvement of existing project performance and for design of any follow-on projects. Planning for the mid-term evaluation should start during Year 2 of the five-year project in order to have the full report and recommendations available in time to inform the re-design process.
- Most of the recommendations made came in late while changes were already in progress. The report failed to recognize these efforts in many ways. Additionally, some findings (though important), the language used for their presentation diverted attention away from their utility.
- My knowledge of actions take on the recommendations are limited by the fact that I left the Mission a month after the report was finalized; I was no longer there to see what actions were taken on the recommendations.
- One fact that the evaluators did not comment much is the fact that [Project Name] started to work with 3,000 schools at once, in three wide provinces. That created too much stretch to all resources (staff, funds, time, equipment, ..) while the scaling-up could be done progressively, for example starting by 500 the first year, then 800, then 1,200 schools (even without piloting!).

- Our Unit has used other evaluations to a larger extent; this one was a particularly poor evaluation and had changed context.
- Performance management and monitoring tools a project should be a living document. They are required to navigate the project and at the same time they need to be responsive with dynamic situation that may not be anticipated when the project was designed
- Please, involve technical/field teams the most you can so that their thoughts are taken in account. If they participate from the beginning, findings may bring change to the system and process. Evaluation team members should avoid behaving like policemen or auditors. This refrains [discourages] people from telling what they really think of the project being evaluated. Very important: changes should follow quickly after evaluation.
- Studies such as this can help focus program design and execution; especially when resources can or could be constrained.
- The above mentioned evaluation was conducted prior to Parliamentary elections which changed the whole dynamic of political developments in the country. It was quite close to elections and was hard to implement some recommendations.
- The draft evaluation report was not of a good quality and it took several months back and forth with the contractor to improve the report quality. By the end we still were not happy with the final product but it was satisfied.
- The [Activity Name] evaluation report was very useful. I specifically liked the way the report was presented. Different grading system was provided (such as A+, A, etc.) for different components and why the grading was given. The report identified some of the drawbacks with the mechanisms/process adopted for the different activities under the project. We ensured that the recommendations related to these drawbacks were adopted in our future program (such as formal system for assessing training effectiveness, common drive for storing information).
- The Evaluation targeted a program that I supported but I was not there when the evaluation took place. I am however involved in the implementation of some of the Evaluation recommendations.
- The PEPFAR priorities essentially changed on a pivot that rendered the findings of this report non priority making it non-strategic to prioritize further.
- The qualitative component of this evaluation had limited usefulness and was determined that it would be a more appropriate methodology for implementing partners as a feedback loop. The quantitative component was extremely useful and determined to serve as a baseline. The Mission is currently repeating the quantitative component in order to measure changes/progress/improvement following in the quality of community-based service delivery.
- The response to this survey reflects the opinions of the Program Office at USAID/[Mission] and the relevant Technical Team members.
- The timing/delay in receiving the evaluation really undermined its potential usefulness, both for the project it was evaluating and in terms of providing information for the follow-on project, [Activity Name]. In spite of this delay, the COR (me) made considerable effort to incorporate findings into the new design (had [Activity Name] not experienced its own delays in procurement, this probably would not have happened).
- There was a RIG-initiated performance audit that was earlier conducted prior to the [Project Name] final evaluation. Most of the findings and recommendations were mainly adopted from the RIG report, hence, there was little value added to the knowledge base about the project. The [Project Name] final evaluation was conducted, despite the earlier RIG audit, in order to comply with Mission evaluation requirements.
- These responses reflect the collective knowledge about the evaluation of staff in the Program Office and the Education and Youth Office.

- This is a good survey. I hope you will be use the information provided to improve the way we conduct and use evaluations at USAID. Good Luck!
- This was an extremely important, complex evaluation to undertake. The technical shortcomings were detailed in the CPAR report to [Contractor]. Northern [Country] remains the “great problem region” for [Country], and USAID has considerable assets up there, but we do not have any one [Region] Unit to drive such affairs, and our equities in these matters, and fungibility of financing, dilute what we can hope to get out of the evaluation. [Country] commissions more and better evals than many, but we are still consumed with Quality Assurance, much less about looking more carefully at utilization and making Mission management deliberate and decide over the recommendations in any formal review process. However, this evaluation, as well as others, have co-opted and benefited from [Partner Country Government] participation and we, unlike most USAID Missions, have conducted a number of joint evaluations, drawing in the [Partner Country Government] and using [Country] eval actors, including [Country]'s [Organization], the [Country] Evaluation Association.
- We took (together with the evaluation team at USAID) a lot of time back and forth with lead consultant to polish the report. It was very clear that the evaluation team was not coherent.

4.2 – EVALUATION UTILIZATION STUDY SUB-STUDIES

1. PPR Evaluation Registry Use Statements
2. Country Development Cooperation Strategies (CDCS)
3. USAID Agency-Wide Policies, Strategies, Frameworks, and Visions
4. USAID Evidence Summits
5. Evaluation Purposes & Questions
6. Evaluation Recommendations
7. Standard Mission Orders on Evaluation
8. Evaluation Utilization Monitoring Systems



EVALUATION UTILIZATION SUB-STUDY I: PPR Evaluation Registry Use Statements

Introduction

Under the Evaluation Utilization Study for USAID/PPL/LER, the study team conducted a series of sub-studies looking at distinct evaluation-related processes or products to identify whether and how they might either utilize or encourage utilization of evaluations across the Agency. This particular study is focused on use statements from the Evaluation Registry of USAID's Annual Performance Plan and Report (PPR).

Main Findings

The Evaluation Utilization Study team conducted a review of USAID's Evaluation Registry in the PPR and identified 136 evaluations claimed in the PPR to have been used in some manner, and in some cases that it was used in multiple ways. The following are the major findings identified from this review. Of the 136 evaluations claiming some form of utilization:

- One hundred twenty-two (90%) were found to claim instrumental use, or use in modifying existing projects or activities, designing future projects or activities, or developing strategies.
 - Sixty-three (52% of instrumental claims) evaluations claimed 80 instances of modifications to the projects or activities being evaluated
 - Fifty-eight (48% of instrumental claims) evaluations claimed 61 instances of evaluations influencing the designs of either follow-on or new projects or activities
 - Thirteen (11% of instrumental claims) evaluations claimed to have been used to modify other existing projects or activities aside from the one being evaluated
 - Seven (6% of instrumental claims) evaluations claimed to have been used in the development of a strategy such as a CDCS
- Thirty-eight (28%) were found to claim conceptual use, or use in changing the way USAID thinks about projects or activities.
- Thirteen (10%) claimed to have disseminated evaluation results to audiences such as partner-country governments, implementing partners, and other donors.

Purpose, Scope, and Methodology

On an annual basis, using the PPR Evaluation Registry, USAID Operating Units are required to provide an update on all planned, conducted, or completed evaluations in that year. For completed evaluations, guidance states that the PPR should, “[d]escribe specific ways the evaluation findings were used to inform decisions or fill knowledge gaps [including] a description of any policy, program management, budgetary or other decisions and changes made as a result of the evaluation findings.” To identify examples of evaluation utilization from PPRs, the study team reviewed the 1,077 Evaluation Registry entries from the 2011-2014 study period and removed all entries that were not directly related to evaluation utilization, leaving 136 entries. These entries were then run through a qualitative analysis to identify patterns in how the evaluations were used.

Sub-Study Findings

While the major evaluation utilization related findings are presented above, additional findings are included below and are organized into instrumental use, conceptual use, and dissemination even though there are instances where one PPR statement included all three.

Instrumental Use

Instrumental use refers to instances where a direct action was taken as a result of an evaluation, and may include modifying existing activities, designing new activities, or the development of a new strategy or policy. Of the 136 PPR statements indicating utilization, 122 (90%) were instrumental use. Since PPR statements could include multiple types of use, or even multiple types of instrumental use, there were a total of 167 instances of instrumental use. Specifically, PPRs claimed that evaluations were instrumentally used to:

- Modify an existing project or activity;
- Modify, influence, or inform a different existing project or activity;
- Design a direct follow-on project or activity or a new activity (not a direct follow on);
- Change the way USAID manages projects or activities (e.g., type of partner); or
- Revise or develop a strategy or policy for the country, region, or office that sponsored or managed the evaluation.

Modify an existing project or activity

Of the 122 evaluation statements claiming instrumental use, 63 (52%) modified a project or activity based on the evaluation's results, for a total of 80 instances of modification. Modifications to existing projects or activities took a variety of forms. A detailed breakdown and frequency of modifications includes:

- A refocusing or prioritizing of interventions within the project or activity (45 instances)
- A revision of the project or activity's work plan (15 instances)
- A revision of the activity M&E plan or PMP (11 instances)
- The project or activity was extended (4 instances)
- The project or activity was expanded to cover additional geographic or technical areas (2 instances)
- The project or activity was terminated (1 instance)
- An exit plan and sustainability strategy was developed (1 instance)
- The project or activity was renamed (1 instance)

The most common type of modification, claimed in 83 percent of all evaluations resulting in modified projects or activities, was a refocusing or prioritization of interventions within a project or activity.

Modify, influence, or inform a different existing project or activity

Analyses of PPRs identified 13 evaluations (12%) that affected an existing project or activity different than the one evaluated. In four of those instances, the affected project or activity was a direct follow-on to the evaluated project or activity, but it was already awarded so the design could not be affected. An example of this type of use is, "the findings and recommendations were used to inform implementation of the newly awarded ... program, which is also implemented by [the same implementing partner]." In the remaining instances, evaluation results were used to reorient other newly awarded projects or activities that were not follow-ons. For example, "findings were used to revise work plan of related project..." In all of these instances the modified projects or activities were at early stages in their project life cycle, indicating that they may have influenced the designs of these projects or activities had evaluation results been available earlier.

Design a direct follow-on or new future project or activity

The study team found that 58 evaluations (48% of instrumental use evaluations) influenced the design of new or follow-on projects or activities. There were no examples of an evaluation influencing the design of both new and follow-on projects or activities.

Direct follow-on projects or activities refer to instances where USAID designed a project or an activity as a “Phase II,” continuation of an existing project or activity, or scaling-up of a pilot. According to the PPR, 20 evaluations were used to affect these kinds of projects or activities, and in one instance a PPR claimed that a single evaluation influenced the design of two different follow-on projects or activities.

New project or activity designs may refer to interventions in a new sector or region, or perhaps the same sector and region but having a different focus. Thirty-eight evaluations influenced the design of this type of project or activity, with one instance where the PPR claimed that an evaluation was used in the design of three new projects or activities.

While PPR statements were generally vague when detailing how evaluations affected project or activity designs, in some instances language was provided that indicated new indicators were created, scopes were narrowed, and in one case the evaluation itself was included in a project solicitation document and thereby influenced the proposed designs submitted by interested firms bidding on the project.

Change the way USAID manages projects or activities

In four instances, an evaluation changed the way a USAID Mission or Office manages its projects or activities. This type of use differs from other uses in that it focuses on USAID behavior as opposed to dictating how an activity or project should be implemented. For example, one PPR statement indicated that USAID “sharpened quality assurance processes.”

Revise or develop a strategy or policy

Of the 122 evaluations resulting in instrumental use, eight (7%) were used to revise or develop a strategy or policy for the country, region, or office that commissioned or managed the evaluation. Analyses indicated that in two instances, evaluations were used directly to influence countries’ CDCSs or equivalent country strategies. In four instances, evaluations informed specific sector strategies without directly referencing a change to a CDCS. For example, “findings were used to design future assistance in the area of small scale infrastructure development and [Mission’s] involvement in the rural development sector” and “findings informed the design of work plans for the follow on projects...as well as the overall [Mission] Health Strategy.”

In one instance, the PPR stated that an evaluation was used to revise or develop a USAID policy for a sector, particular problem or topic. Specifically, “findings [from this evaluation] were used to design future USAID assistance in the sector of political process strengthening.”

Conceptual Use

Conceptual use differs from instrumental use in that it does not immediately and directly affect the way in which USAID strategies, policies, projects, or activities are managed or implemented. Instead, conceptual use changes the way USAID staff perceive or think about the work that they are doing. Of the 136 evaluations reviewed, 38 (28%) included some form of conceptual use, with a total of 51 claimed instances. Content analyses of conceptual use claims identified that evaluations were used to:

- Change or validate views about the soundness or adequacy of the design;
- Change the way USAID thinks about other activities in the same sector or topic;
- Change the way USAID thinks about partner collaboration;
- Assess the way the project, activity, or implementing partner performed;

- Understand how to make this type of project or activity more effective in the future; or
- Change the way USAID partners think about activities in the sector.

Change or validate views about the soundness or adequacy of the design

Seven evaluations indicated eight instances where findings or recommendations were used to validate or change USAID staff views about the design of the project or activity being evaluated. In three of the instances, the evaluation validated the design or programmatic approach already in place. For example, “the evaluation report was extremely useful to the Mission in refining programmatic approaches based on lessons learned... [Evaluation] findings and recommendations validated the programmatic directions that ... project has taken during the next two years of implementation.” In the remaining seven instances, the evaluation was used to inform USAID about major successes or lessons learned, or otherwise assess the design of the project or activity. Unfortunately, insufficient language was provided in these cases to determine whether changes in views were positive or negative.

Change the way USAID thinks about other activities in the same sector or topic

This type of conceptual use was more prevalent than others, with 20 of the 38 (53%) evaluations changing the way USAID thinks about other activities in the same sector or topic, for a total of 25 instances. One quarter of the evaluations (5 of 20) were used to document best practices and lessons learned. In 11 cases, evaluations influenced the way that USAID will approach future projects or activities in this sector, for example, “the lessons learned from this evaluation were used to enhance USAID’s understanding of program performance and inform decisions about future USAID programming in the area of decentralization and local governance.”

In six instances, findings from evaluations were used to think about or inform how other current projects or activities were being implemented at the time. There was no consistency in these claims, and conceptual use here ranged from changing how they think about:

- Projects or activities in other sectors but with comparable implementation approaches;
- The way USAID approaches the monitoring and evaluation of projects in a sector; to
- The way they think about how the host country government could approach projects or activities in a given sector.

Change the way USAID thinks about partner collaboration

Of the 32 conceptual statements, there were three instances (10%) where the evaluation findings or recommendations affected the way the Mission thought about partner collaboration. One instance touched on how to maintain better relationships with stakeholders while another discussed the manner in which partner organizations are engaged to better implement a project or activity. The third instance related to the host country government and how to better engage a particular ministry in government-to-government projects or activities.

Understanding the way the project, activity, or implementing partner performed

In seven instances, USAID claimed to have used the evaluation to assess how a project, activity, or implementing partner was performing, and in some cases how it could be improved. In three of these cases, the conceptual use related to the project or activity’s performance, such as effectiveness or expected impact. In the other four cases, conceptual use related to the implementing partner for the project or activity, or perhaps a component of the project or activity. Two instances stated that “the findings were used by [Mission] to assess activities of the [Implementing Partner] and its local partner ...” and “...the evaluation report is also being used as a source for the analysis on [the project’s] performance at the mid-term of the project.” Unfortunately, there was no additional information provided on the results of these project or partner assessments.

Understanding of how to make this type of project or activity more effective in the future

There were four instances where USAID claimed to have learned how to make similar projects or activities more effective in the future. In one instance, the claim complemented an instrumental use claim where the follow-on project or activity design was affected. In the other three instances, claims were more specific to the types of project or activity, such as increasing health communication strategies or using a localized producer to market strategies.

Change the way USAID partners think about activities in the sector

In four instances, USAID indicated in the PPR that evaluation results were used not by USAID, but by USAID partners to rethink activities in particular sectors. In three instances, the partners came from host country governments. Two instances state that “other implementing agencies and [the host government] are also using the evaluation’s finding to inform their decisions” and “The evaluation’s findings were used by the project development team, [USAID’s technical office], and the [host country government]. The purpose of the evaluation was to identify lessons that will be used to improve current and future HIV/AIDS and TB programs in [the country].” In the fourth instance, it was other donors that utilized the evaluation findings to rethink their activities. There was no further information on how exactly perceptions or thoughts changed.

Dissemination

Though not technically a distinct type of use, there were 13 evaluations resulting in 19 instances where evaluation results were disseminated to partners and stakeholders, though there is no information on how these partners or stakeholders might have used the information. Analyses of PPR statements indicated that evaluation reports, summaries, results, findings, and/or recommendations were disseminated to the following groups:

- Host country governments (6 instances)
- Other donors (4 instances)
- Implementing partners (3 instances)
- Other stakeholders (3 instances)
- Conferences attendees (2 instances)
- Other USAID Offices (1 instances)



EVALUATION UTILIZATION SUB-STUDY 2: Country Development Cooperation Strategies

Introduction

Under the Evaluation Utilization Study for USAID/PPL/LER, the study team conducted a series of sub-studies looking at distinct evaluation-related processes or products to identify whether and how they might either utilize or encourage utilization of evaluations across the Agency. This particular study is focused on how evidence from evaluations is used in the development of Country Development Cooperation Strategy (CDCS) and Regional Development Cooperation Strategy (RDCS) documents.

Main Findings

The Evaluation Utilization Study team conducted a review of 51 publically available CDCS (45) and RDCS (6) documents, collectively referred to from here forward as CDCSs. That review identified that 30 of the 51 (59%) CDCSs referenced or cited USAID evaluations as evidence for their strategies. It should be noted that CDCSs had a higher rate of USAID evaluation use than RDCSs, with the former citing or referencing evaluations in 28 of the 45 (62%) CDCSs compared to two of the six (33%) of the RDCSs. Of the 30 CDCS and RDCS documents citing USAID evaluations:

- Fourteen (47%) used evaluations to inform sector strategies at the DO level.
- Eighteen (60%) used evaluations to inform implementation strategies at the IR level.
- Eight (27%) used evaluations to inform CDCS development in general.
- Four (13%) used evaluations to indicate the extent to which stated objectives were or were not achieved through past and present interventions.
- One (3%) used an evaluation to facilitate learning about evaluation practices.
- Five (17%) provided contextual knowledge about a country or region.
- One (3%) cited an evaluation sponsored by a different Operating Unit other than the authoring unit of the strategy document. All other identifiable citations to USAID evaluations were to evaluations commissioned by the same Operating Unit that produced the CDCS.

In the 30 CDCS and RDCS documents that referenced USAID evaluations, there were 73 separate references to USAID evaluations. Some but not all of these references identified the specific USAID evaluation on which USAID staff were drawing by name. It is thus impossible to say exactly how many USAID evaluations these 30 CDCSs used. Fifteen of the evaluations referenced in the body of the reports and an additional 28 evaluations listed in bibliographies provided sufficient information to be matched to evaluation reports in the study universe, totaling 43 evaluations.

Purpose, Scope, and Methodology

As required by USAID guidance, including ADS 201.3, Bilateral and Regional Missions are responsible for creating 3-5 year strategy documents detailing the goal and development objectives they intend to achieve over the stated time period, as well as their underlying evidence-based development hypotheses. To understand the extent to which USAID evaluations comprised this evidence base, the study team reviewed 51 CDCSs currently uploaded to USAID's public web site, including 6 RDCSs. The team conducted qualitative content analyses of the CDCSs, identifying all instances where USAID evaluations

were referenced as evidence. A content analysis of strategies for learning from and disseminating evaluations found in the Monitoring, Evaluation, and Learning section of the reports was also conducted.

Sub-Study Findings

While the major evaluation utilization-related findings are presented above, additional findings are included below. Findings are broken down into distinct sections on the use of evidence, references to evaluations in CDCS documents, and policies for learning from evaluations.

Use of Evidence in CDCSs

To identify the different sources of evidence used in CDCSs, the study team performed a keyword search for the terms “evidence” and “evaluation” in the body of each document then skimmed footnotes and in-text parenthetical citations, if present, for further references. Lists of sources referenced, if included at the end of a CDCS, were considered separately and will be discussed below. The following counts and sources of evidence were identified from the 51 CDCSs reviewed:

- Twenty-seven (53%) claimed evidence without referencing a source
- Thirty (59%) referenced or cited USAID evaluations
- Forty-five (89%) referenced or cited USAID documents other than evaluations
- Eight (16%) referenced or cited non-USAID evaluations such as those from other donors
- Twenty-one (41%) referenced or cited published research
- Forty-nine (96%) referenced or cited other non-USAID documents

Table 1: Number of Evaluations Referenced per Section

CDCS Section	USAID Evaluations	Non-USAID Evaluations
Introduction	1	0
Development Context	10	2
Development Hypothesis	4	1
Results Framework	56	3
Annex	7	2

Evaluation References

USAID guidance, per ADS 201.3.4, refers to specific sections of the CDCS as needing to be evidence-based, though that does not mean additional evidence should not be included elsewhere. These particular sections include the Development Context, Development Hypothesis, and Results Framework sections. As a part of the analysis, the study team identified those sections in which evidence from evaluations was found, as reflected in Table 1.

Types of utilization of USAID evaluations

Of the 51 CDCSs reviewed, 30 (59%) referenced USAID evaluations. Numerous CDCSs referenced evaluations multiple times, giving a total of 73 references to evaluations, for an average of nearly 2.5 references per CDCS that included such evaluations. The maximum number of references per CDCS was eight, and 21 CDCSs did not contain any such references.

In addition to the number of evaluation references, the team also identified the total number of distinct evaluations referenced. While some CDCSs were ambiguous as to the evaluations they drew from, the team was able to determine that, where identified, between one and eleven distinct USAID evaluations were referenced in CDCSs, for an average of 0.9 USAID evaluations referenced per CDCS.

The 73 references to USAID evaluations in CDCSs fell into one of the six following types of use:

- Thirty-six instances (49%) informed implementation strategies
- Twenty instances (27%) informed sector strategies

- Ten instances (14%) informed CDCS development in general
- Seven instances (10%) provided contextual knowledge about the country or region
- Five instances (7%) demonstrated progress or levels of success towards goals
- One instance (1%) facilitated learning about monitoring and evaluation practices

Instances of each type of evaluation use were scattered through all sections of the CDCSs, without clear patterns of a particular category of use appearing more frequently in any one section over another.

There were 20 instances in 14 CDCSs where evaluations were used to inform sector strategies through development objectives (DO), supporting objectives, or cross-cutting priorities. Five of these instances informed sector strategies in a general manner. Eleven instances informed how USAID focuses efforts within a sector, either by identifying new focus areas (4 instances) or by shifting between existing programs within the sector. Two of these instances used evaluations to inform which beneficiary groups focuses within the sector (4 instances). In one instance, an evaluation informed the geographic focus of interventions. Programmatically, evaluations identified successful programs as models for future interventions (2 instances), justified continued use of an existing model (1 instance), and led to revising the strategy for a current model in a sector (2 instances).

In 36 instances, evaluations informed implementation strategies at intermediate results (IRs) or Sub-IR levels. These evaluations either identified best practices (12 instances), justified continuing a particular program, project, activity, or approach (10 instances), or informed how to modify and improve existing projects, activities, or programs in the country or region (10 instances). In other cases, evaluations were used to identify internal and external conditions for success (5 instances) or to identify needs to be addressed by modifying current programs or developing new programs (5 instances).

Considerations for data quality in evaluation references

We hypothesized that higher quality evaluation data might result in increased utilization, but no CDCSs mentioned data quality or reliability in relation to evaluations.

Evaluation citation quality

References to USAID evaluations in the body of CDCSs were inconsistent. Of the 73 USAID evaluation references, only 32 included a proper in-text citation or footnote, and none provided links or USAID IDs. Thirty-five references provided the full name of the evaluation and 36 provided the year the evaluation was published. Only 16 specified the report’s authors.

There were 14 instances where evaluations were claimed as having been used as evidence, but no further information was provided. Generally, the precision of the information about evaluations in CDCSs was low compared to other types of sources cited.

USAID evaluation references by region, sector, and year

The study team examined references to USAID evaluations by CDCS region, sector, and starting year. While 59 percent of the 51 CDCSs reference USAID evaluations overall, the numbers varied widely between regions, as demonstrated in Table 2. Asia, Europe & Eurasia, and the Middle East all had a lower percent of CDCSs referencing USAID evaluations than the global average.

Table 2: USAID Evaluation References by Region

Region	Number of Public CDCSs	CDCSs Referencing USAID Evaluations
Africa	18	13 (72%)
Asia	12	5 (42%)
E&E	9	4 (44%)
LAC	9	6 (67%)
ME	3	1 (33%)

The study team also looked at which sectors were most commonly supported by evaluations. This information is presented in Table 3 below. In nine cases, there was insufficient information to determine which sector an evaluation was supporting. The most common sector referencing evaluations was democracy and governance, followed by economic growth, with few education evaluations referenced.

The same table also presents information on sectors included in CDCS DOs and IRs, to better understand whether evaluation references for a given sector are dependent on a CDCS focusing on that sector. All CDCSs that referenced an evaluation in a particular sector also contained a DO and/or IR in that sector.

Table 3: CDCS and Evaluation References by Sector

Sector	CDCSs with a Sector Focus	Percent of CDCSs Referencing Evaluations
Agriculture	24	6 (25%)
Democracy & Governance	48	11 (23%)
Education	25	7 (28%)
Economic Growth	47	11 (23%)
Health	35	10 (29%)

The percent of CDCSs with a DO and/or IR in a sector that referenced an evaluation of that sector ranged from 23 percent to 29 percent, roughly the same across all sectors.

Looking at starting dates for the current 51 public CDCS documents, the team found they ranged from 2011 to 2015. Table 4 shows the distribution of CDCSs by year as well as the number of CDCSs published each year that reference USAID evaluations. CDCSs that reference evaluations range from 44 percent to 63 percent and do not demonstrate any patterns over time.

References to non-USAID evaluations

Eight CDCSs (16%) each referenced one non-USAID evaluation. Sponsors of these evaluations included the World Bank, the World Health Organization, and partner governments, among others. In six of the eight cases, a full citation with evaluation title, year, and author is given. One case only provided the name of the program evaluated and sponsoring evaluation. In other words, seven of the eight (88%) non-USAID evaluations were easily identifiable from the CDCS text. This is a far higher proportion of identifiable evaluations than evaluations of USAID programs.

Table 4: USAID Evaluation References by Year

CDCS Starting Year	Number of CDCSs	CDCSs Referencing USAID Evaluations
2011	9	4 (44%)
2012	8	5 (63%)
2013	15	7 (47%)
2014	15	9 (60%)
2015	4	2 (50%)

Evaluations in reference lists

Of the 23 CDCSs that included a references section or a list of influential assessments and evaluations as an annex, 14 listed USAID evaluations. Because some of these CDCSs listed evaluations in their reference lists that were not cited anywhere in the body of the report, and it could not be determined how or if these evaluations were used, evaluations the reference lists were excluded from the earlier analyses. Sixty-one USAID evaluations were found in reference lists, and each CDCS had between zero and thirteen USAID evaluations listed. Five CDCSs referenced non-USAID evaluations, with one to three non-USAID evaluations listed each.

Monitoring and Evaluation Dissemination Policies

To determine how Missions disseminate evaluations, dissemination policies for monitoring and evaluation (M&E) were examined. ADS 201.3.4.5 encourages Missions to include a Learning section under the M&E section of the CDCS that details approaches to encourage learning through all stages of the Program Cycle. The study team focused its search for language on dissemination or other forms

knowledge sharing with stakeholders in this section of the CDCS. It is possible that there were other instances of language about dissemination in other sections of the CDCS, but those were not considered for this analysis.

Twenty-six CDCSs (51%) included language about dissemination or knowledge sharing with stakeholders in their Learning sections. Each of these instances was analyzed to understand the type of dissemination activities and intended audience. Identified dissemination activities represented a wide range of formal and informal means for knowledge sharing with stakeholders. The target groups for these dissemination and learning activities were as follows:

- Nine (34%) called for dissemination to implementation partners
- Seven (27%) called for dissemination to country partners
- Seven (27%) called for dissemination to other donors
- Four (15%) called for dissemination to direct beneficiaries
- Eleven (42%) did not specify target groups for dissemination

There were 17 CDCSs that called for dissemination in general, without giving specifics on the type of dissemination activities. Of these, nine called for disseminating results, findings, lessons, and recommendations from evaluations. Three called for dissemination of monitoring or other progress information and two called for dissemination of process knowledge.

Twenty-five of these described traditional dissemination activities and fifteen described other forms of shared learning from evaluations with stakeholders. Many CDCSs described multiple types of dissemination and learning activities. The six types of traditional dissemination activities described in CDCSs were:

- Sixteen described general dissemination to stakeholders
- Ten described stakeholder meetings, discussions, or briefings
- Five described internal dissemination activities
- Five described dissemination through communities of practice or other networks
- Five described online dissemination
 - Two of these described uploading evaluation reports to the DEC
- Two described dissemination of materials translated into local languages

Of the ten CDCSs that called for dissemination at stakeholder meetings, seven suggested meetings for disseminating results, findings, lessons, and recommendations from evaluations. Four called for meetings to disseminate monitoring and progress information, and four others called for meetings to disseminate process knowledge. Three CDCSs included language that called for meetings to disseminate hypotheses.

The 15 CDCSs that described other forms of knowledge sharing with stakeholders described three types of learning activities:

- Seven described stakeholder participation in evaluations
- Five described stakeholder participation in Portfolio Reviews
- Four described stakeholder participation in Performance Management Plan development

The following table provides the full list of CDCS documents reviewed, including the year they were finalized.

Table 5: List of CDCS Documents in the Sub-Study

Mission	Evaluation Citation	Year	Mission	Evaluation Citation	Year
Azerbaijan		2011	Liberia	Y	2013
Bangladesh	Y	2011	Malawi	Y	2013
Albania		2011	Moldova		2013
Sri Lanka		2011	Serbia	Y	2013
Ethiopia	Y	2011	South Africa	Y	2013
South Sudan		2011	Zimbabwe		2013
Zambia	Y	2011	Kosovo	Y	2014
Uganda		2011	Cambodia	Y	2014
Southern Africa	Y	2011	Nepal		2014
Bosnia		2012	Tanzania	Y	2014
Central Africa (CARPE)	Y	2012	Vietnam		2014
Philippines	Y	2012	Yemen		2014
Guatemala	Y	2012	Indonesia	Y	2014
India		2012	Dominican Republic	Y	2014
Peru	Y	2012	Jamaica	Y	2014
Senegal	Y	2012	Mozambique	Y	2014
Ukraine	Y	2012	Paraguay		2014
Bureau for Africa		2013	Angola	Y	2014
Armenia	Y	2013	Colombia	Y	2014
Morocco	Y	2013	Kenya		2014
Jordan	Y	2013	RDMA		2014
Timor-Leste	Y	2013	Kyrgyz Republic		2015
Nicaragua		2013	Honduras	Y	2015
El Salvador		2013	Central Asian Republics		2015
Georgia		2013	Democratic Republic of Congo	Y	2015
Ghana	Y	2013			



EVALUATION UTILIZATION SUB-STUDY 3: USAID Agency-Wide Policies, Strategies, Frameworks and Visions

Introduction

Under the Evaluation Utilization Study for USAID/PPL/LER, the study team conducted a series of sub-studies looking at distinct evaluation-related processes or products to identify whether and how they might either utilize or encourage utilization of evaluations across the Agency. This particular study is focused on how evaluation results are used to influence USAID policy.

Main Findings

The Evaluation Utilization Study team conducted a review of the 19 USAID Policies and the USAID Policy Framework, 2011-2015 to determine how often and in what manner evaluations influenced USAID's policy decision making. The following are the major findings identified from this review. Of the 20 documents reviewed:

- Only one explicitly referenced an evaluation of a USAID project or activity, and that evaluation was commissioned and managed by an implementing partner, not USAID directly.
- One referenced an Evidence Summit in the policy document, and that Evidence Summit mentioned the use of a USAID evaluation.

Purpose, Scope, and Methodology

According to USAID's website, USAID policies, strategies, frameworks, and visions are designed in a collaborative, evidence-based process. As such, the study team looked at such documents to identify any instances where USAID evaluations might have provided such evidence. The team conducted a review of the 20 such documents posted to the USAID website. Nineteen of them were at a sub-agency level, but the twentieth was the USAID Policy Framework 2011-2015, which is at the Agency level. A list of these documents is provided below. The review looked for specific references to evidence used, namely:

- USAID documents or experience (including Evidence Summits);
- USAID evaluations;
- Non-USAID evaluations;
- Published research from journals or other scholarly literature; and
- Other documents from non-USAID sources.

Sub-Study Findings

Nineteen of the twenty (95%) policy documents either directly referenced USAID documents or indicated having learned from USAID's development experiences, though did not indicate specific documents. The USAID documents referenced included other policy documents, assessments, and project or activity reports. Policy documents commonly discussed outcomes or lessons from successful or unsuccessful USAID programs as evidence for policy decisions without citing a specific source. It is possible that USAID evaluations provided evidence for these case studies during the policy development process, but the team was unable to determine the influence of uncited evaluations on the policies.

Only one policy—focused on ending child marriage—referenced an evaluation of a specific USAID project or activity. The evaluation, commissioned by the implementing partner, was used to demonstrate what an effective program for supporting married children looked like. This case study accompanied an assertion that such programs were necessary. There were no references to USAID-commissioned evaluations of USAID programs in any of the policies.

Of the non-USAID sources cited, 15 policy documents (75%) referenced published research while 18 policy documents (90%) referenced documents from other non-USAID sources. One policy document—on local systems—referenced an evaluation of a non-USAID program. In this instance, evidence from an independent impact evaluation was used to illustrate the value of local accountability, justifying that aspect of the policy. In total, only 2 (10%) of the 20 policy documents referenced evaluations of any kind.

We also found that two policy documents referenced Evidence Summits. In both of these cases, an Evidence Summit on the same topic provided the genesis for the policy. Learnings from each summit were synthesized into the final policy document. For more information on these Evidence Summits see Sub-Study 8 on Evidence Summits.

Policy Documents Reviewed	
USAID Policy Framework	LGBT Vision for Action
Counter Trafficking In Persons Policy	USAID Education Strategy
USAID Biodiversity Policy	USAID Scientific Research Policy
Multi-Sectoral Nutrition Strategy	USAID Water and Development Strategy
USAID Policy: Youth	USAID Evaluation Policy
USAID Climate Change & Development Strategy	Gender Equality and Female Empowerment Policy
USAID Guidance on Programming in Closed Spaces	USAID Strategy on Democracy Human Rights and Governance
Local Systems: A Framework for Supporting Sustained Development	Ending Preventable Maternal Mortality: USAID Maternal Health Vision for Action
Sustainable Service Delivery in an Increasingly Urbanized World: USAID Policy	Building Resilience to Recurrent Crisis: USAID Policy and Program Guidance
The Development Response to Violent Extremism and Insurgency	Ending Child Marriage & Meeting the Needs of Married Children: The USAID Vision For Action



EVALUATION UTILIZATION SUB-STUDY 4: USAID Evidence Summits

Introduction

Under the Evaluation Utilization Study for USAID/PPL/LER, the study team conducted a series of sub-studies looking at distinct evaluation-related processes or products to identify whether and how they might either utilize or encourage utilization of evaluations across the Agency. This particular study is focused on how evaluations are used in USAID Evidence Summits.

Main Findings

The Evaluation Utilization study team conducted a review of nine USAID Evidence or Experience Summits (referred to as Evidence Summits moving forward) held since December 2011. The following are the major findings identified from this review. Of the nine Evidence Summits reviewed:

- One provided an evidence base for respective USAID policy documents.
 - The *Development to Counter Insurgency* Summit provided evidence used in the *Development Response to Violent Extremism and Insurgency* policy paper
- Two included USAID evaluations as resources, and each of those two only included one USAID evaluation apiece.
 - The *Strengthening Country Systems* Summit referenced the 2011 *Development Grants Program Global Evaluation* (http://pdf.usaid.gov/pdf_docs/PDACT867.pdf)
 - The *From Microenterprise to Inclusive Market Development* Evidence Summit included as recommended readings the 2010 *Profit Zambia Impact Assessment Final Report* (http://pdf.usaid.gov/pdf_docs/PDACR843.pdf)
- Seven provided lists of speakers online; where provided, USAID staff never represented more than 50% (11/22) of speakers at any summit and reached as low as 8% (2/24).

Purpose, Scope, and Methodology

Evidence Summits are coordinated by USAID Missions, Bureaus, and Offices to develop an evidence-based perspective on a particular subject of focus, gathering information from USAID projects as well as other donors' perspectives and academic literature. The study team identified 11 summits to review based on a combination of information posted to USAID's website, USAID's Learning Lab, a general internet search, and conversations with USAID staff. Upon review, the team identified that not all summits were Evidence Summits; two were Experience Summits and a third was an Evaluation Summit, though no differentiations were provided. One of the two Experience Summits was delayed for security reasons, and the Evaluation Summit had a different focus; therefore neither was included in analyses. The process of reviewing summits involved identifying resources available via summit websites (e.g., transcripts, agendas, reference lists, etc.) and identifying the people and evidence sources from the summits. All resources were reviewed except for video or audio recordings of sessions within summits.

Sub-Study Findings

In addition to the major evaluation utilization related findings presented above, the study team also identified the following findings.

Of the nine Evidence Summits that were completed, they varied in terms of who organized them, the geographic and sector focus of the summits, and even the type of summits. The following focuses were represented:

- USAID summit hosts
 - PPL/LER – five summits
 - Global Health Bureau – four summits
- Types of summits
 - Eight Evidence Summits
 - One Experience Summit (RDMA attempted to host one additional Experience Summit that was cancelled due to political unrest)
- Focus of summits
 - Agricultural Technology and Food Security in Africa
 - Broad-Based Economic Growth
 - Community and Formal Health System Support for Enhanced Community Health Worker Performance
 - Development to Counter Insurgency
 - Enhancing Provision and Use of Maternal Health Services through Financial Incentives
 - Microfinance and Inclusive Market Development
 - Protecting Children Outside of Family Care
 - Population-Level Behavior Change for Child Survival Strengthening Country Systems
 - Strengthening Country Systems

The availability of results from past Evidence Summits was inconsistent, with some summits having nothing more than a brief description available and others having entire websites developed and maintained in response to the summit. As a result, the study team was not always able to find general information such as agendas, lists of participants, or resources referenced during summits.

In those cases where rich information was available, it often went beyond the provision of agendas, participant lists, and resource lists to include a summary brief document from the summit as well as recordings or typed transcripts from sessions. In some instances additional documents were created which clarified aspects of Evidence Summits. In two instances, for two different summits, documents were created which identified the specific processes by which evidence was gathered and assessed prior to being included in Evidence Summits. In another instance, the summit created a document referred to as an “evidence to action strategy” documenting how the evidence would be used to affect programming. These documents are available respectively via the following links:

- <http://www.tandfonline.com/doi/full/10.1080/10810730.2014.918215#.VZr-GfViko>
- http://www.usaid.gov/sites/default/files/documents/1864/mh_summit_evidence_synthesis.pdf
- <http://www.childreninadversity.gov/docs/default-source/default-document-library/apca.pdf?sfvrsn=4>

In terms of the content of Evidence Summits, very little of the evidence presented came from USAID. While not all summits provided agendas or lists of speakers, of the eight summits that did, USAID on average comprised 21% of speakers or presenters, or 6 out of the average 29 speakers at each summit. Of the resources referenced at summits, USAID evaluations were rarely included or referenced in materials, and were only referenced at two different summits each only referencing one evaluation. In two instances Evidence Summits provided full lists of all resources referenced, identifying 129 and 194 respectively, but neither of those lists included a single USAID evaluation reference.

A full list of the Evidence Summits and links to where additional information is available is provided below:

Evidence and Experience Summits with Links to Webpages	
Agricultural Technology and Food Security in Africa	http://agrilinks.org/agexchange/agricultural-technology-adoption-food-security-africa-evidence-summit OR http://cega.berkeley.edu/events/usaids-atai-evidence-summit-agricultural-technology-adoption-food-security-in-africa/
Broad-Based Economic Growth	https://usaidlearninglab.org/events/usaids-evidence-summit-promoting-broad-based-growth OR http://kdid.org/events/usaids-evidence-summit
Community and Formal Health System Support for Enhanced Community Health Worker Performance	http://www.usaid.gov/what-we-do/global-health/chw-summit#overlay-context=node/21216
Development to Counter Insurgency	http://blog.usaid.gov/2010/09/our-first-evidence-summit-tackling-tough-challenges-of-counterinsurgency-and-counterterrorism/
Enhancing Provision and Use of Maternal Health Services through Financial Incentives	http://www.usaid.gov/what-we-do/global-health/maternal-and-child-health/us-government-evidence-summit
Microfinance and Inclusive Market Development	https://www.microlinks.org/events/evidence-summit-microfinance-inclusive-market-development
Protecting Children Outside of Family Care	http://www.childreninadversity.gov/news-information/in-the-press-events/news--full-view/u.s.-government-evidence-summit-on-protecting-children-outside-of-family-care
Population-Level Behavior Change for Child Survival Strengthening Country Systems	http://plbcevidencesummit.hsaccess.org/home
Strengthening Country Systems	http://kdid.org/events/experience-summit OR http://usaidlearninglab.org/events/strengthening-country-systems-experience-summit



EVALUATION UTILIZATION SUB-STUDY 5:

Evaluation Purposes & Questions

Introduction

Under the Evaluation Utilization Study for USAID/PPL/LER, the study team conducted a series of sub-studies looking at distinct evaluation-related processes or products to identify whether and how they might either utilize or encourage utilization of evaluations across the Agency. This particular study is focused on evaluation purpose statements and evaluation questions.

Main Findings

The Evaluation Utilization Study team conducted a review of evaluation purpose statements and questions from 45 USAID evaluations randomly selected from the overall Evaluation Utilization Study universe. The following are the major findings identified from this review. Of the 45 evaluation reports analyzed:

- 33 (73%) included a management purpose or clear description of how the evaluation would be used to answer management decisions. Of these, 10 included some form of purpose statement such as identifying lessons learned, but 2 included no reference to purpose statements at all.
- 40 (89%) identified specific evaluation questions to be answered by the evaluation team.

Purpose, Scope, and Methodology

Starting in the evaluation statement of work (SOW) development phase, and even earlier, USAID has an opportunity to direct an evaluation on the path of utilization. Specifically, by indicating a clear management purpose—or identifying key management decisions that will be answered using evidence from an evaluation—USAID can ensure that the information provided by an evaluation team is in line with how the evaluation will be used. When a purpose statement is aligned with a small number of clear, focused, and relevant evaluation questions, the likelihood of evaluation utilization increases. To understand these critical elements of a SOW, the study team analyzed the purpose statements and questions from 45 evaluation reports. These 45 reports are a randomly selected subset of the 237 evaluations for which evaluation utilization surveys were sent to USAID staff. Questions and purpose statements were extracted from the body of evaluation reports except in cases where they were not provided in the body of the report and then the study team looked for them in the SOW, if included as annex. The study team conducted both qualitative and quantitative analyses for this sub-study. Quantitative analyses included the counts of distinct evaluation questions as well as the count of question marks included in the list of evaluation questions. Qualitative analyses were conducted of evaluation purpose statements, using content and pattern analyses to determine the types of purpose statements used in evaluations and frequency with which they appear.

Sub-Study Findings

While the major evaluation utilization-related findings are presented above, additional findings are included below and are organized by evaluation purpose statements and evaluation questions.

Evaluation Purpose Statements

Purpose statements tend to be short, leaving little opportunity for in-depth analyses. However, the study team’s review did identify two different types of purposes within purpose statements: management and other. Management purposes include specific language on how an evaluation will be used in making management decisions such as modifying activities, designing future or follow-on projects or activities, or developing strategies and policies, to name a few. Other purposes indicated that evaluations would be used to identify lessons learned, assess whether the activity achieved its objectives, or to determine the potential sustainability of a project or activity, among others. The sub-study found that:

- Twenty-nine evaluations (64%) included both a management and other purposes
- Four evaluations (9%) included only a management purpose
- Ten evaluations (22%) included only an “other” purpose and no management purpose
- Two evaluations (4%) did not include any purpose statement at all

Management Purposes

The study team found that 33 (73%) of the 45 evaluations identified at least one management purpose, with 13 (29%) of those 33 including multiple management purposes in a single purpose statement. The management purpose statements claimed that evaluations would be used to:

- Modify the existing project or activity (17 instances)
- Design a direct follow-on project or activity (4 instances)
- Design a new project or activity that is not a direct follow-on (12 instances)
- Modify, influence, or inform a different existing project or activity, other than the one that was evaluated (3 instances)
- Change the way USAID manages projects or activities (3 instances)
- Revise or develop a strategy for the country, region, or office that sponsored or managed the evaluation (12 instances)
- No management purpose identified (12 instances)

Other Purposes

In addition to, or sometimes instead of, identifying management purposes, evaluation reports identified a number of other purposes for conducting the evaluation. For example, nearly all of the evaluations (82%) indicated that the purpose of the evaluation was to assess some aspect of the project or activity. The following are other purposes for the evaluations in the sub-study with additional explanations in some cases:

- To assess the current project or activity (37 instances)
- To identify gender-related issues (2 instances)
- To improve the way implementing partners manage the project or activity (3 instances)
- To document lessons learned and best practices (16 instances)
- Other (6 instances)

Of these sub-categories, two are worth further analysis: assessing current projects or activities and other. The assessment of current projects or activities is a fairly broad catch-all category that can be further broken down into assessments of project performance, impact, achievement of objectives, and so forth. “Other,” however, captured some of the more interesting and distinct statements mentioned.

To Assess the Current Project or Activity

- Effectiveness (12 instances)
- Performance (10 instances)

- Results, outcomes, or impacts (10 instances)
- Achievement of objectives (9 instances)
- Sustainability (7 instances)
- Implementation or management (6 instances)
- Design (5 instances)
- General assessment of the project or activity (2 instances)
- Monitoring and evaluation
- Cost-effectiveness

Other

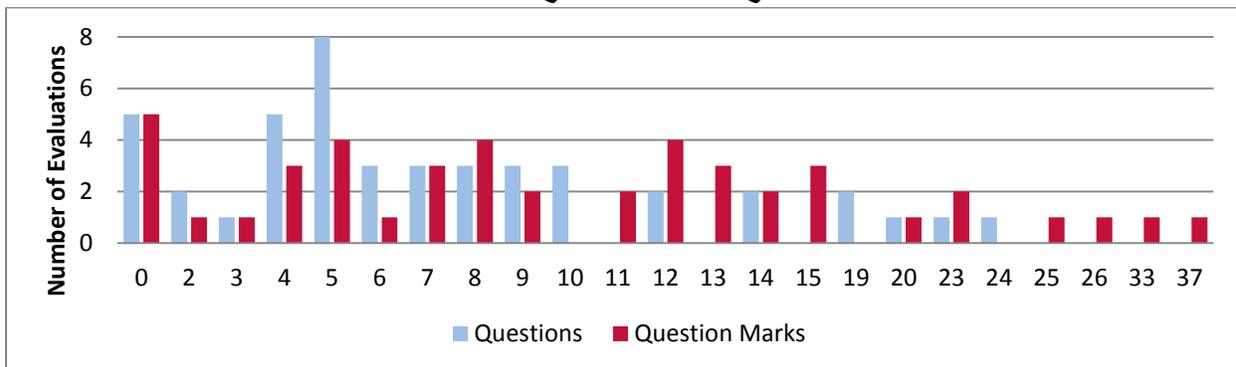
- Identify success stories
- Add to the knowledge base of all stakeholders so that they can move forward with appropriate strategies to meet sector objectives
- Assess the extent to which recommendations from a mid-term evaluation were implemented and the effects of implementing them
- Provide recommendations on how to scale up an activity within the country (not just USAID)
- Recommend specific opportunities to enhance regional-level impact
- Recommend specific opportunities to strengthen the regional Wildlife Enforcement Network (WEN) and sustainability approach

Evaluation Questions

When identifying evaluation questions, the team initially looked at the body of the evaluation report and moved on to the SOW when unavailable in the report. There were 5 evaluations of the 45 total (11%) where no evaluation questions were provided.

Two counts of evaluation questions were conducted. The first was of the stated number of questions, excluding sub-questions, included in the body of the report or in the SOW. The second count was of the actual number of question marks included in the questions section or the report or SOW. While a breakdown of the number of questions and question marks are provided in the chart below, the team found that the average number of questions was 7.6, which was lower than the average number of question marks at 10.7. Of the 40 evaluations that included evaluation questions, 15 (38%) had the same number of questions as question marks while 25 (63%) had more question marks than questions, indicating that evaluation teams were often asked to respond to more questions than initially indicated.

Chart I: Number of Evaluation Questions and Question Marks in Evaluations





EVALUATION UTILIZATION SUB-STUDY 6: Evaluation Recommendations

Introduction

Under the Evaluation Utilization Study for USAID/PPL/LER, the study team conducted a series of sub-studies looking at distinct evaluation-related processes or products to identify whether and how they might either utilize or encourage utilization of evaluations across the Agency. This particular study is focused on evaluation recommendations.

Main Findings

The Evaluation Utilization Study team conducted a review of evaluation recommendations from 45 USAID evaluations randomly selected from the overall Evaluation Utilization Study universe. The following are the major findings identified from this review. Of the 45 evaluation reports analyzed:

- There was an average of 18 formal recommendations in each evaluation, ranging from zero to seventy.
- Four evaluations (9%) did not include any recommendations at all, and one (2%) had recommendations buried in paragraph-form, making them uncountable.
- Ten evaluations (22%) included informal recommendations, or recommendations spread throughout the evaluation report, and not in a distinct section.

The team also qualitatively reviewed the first five recommendations from those 40 reports that had clear recommendations. From that review, the team found:

- There are inconsistencies in the content, structure, and quality of recommendations.
- Recommendations are largely not meeting USAID policy expectations in regards to being directed at specific parties, either not indicating actors at all or being ambiguous in naming them.
- Many recommendations are clear, concise, and action-oriented, but at least as many, if not more, are unclear, unnecessary, impractical, and not even truly recommendations.

Purpose, Scope, and Methodology

Evaluation recommendations are the foundation for evaluation utilization in that they are meant to provide USAID and its partners with evidence-based guidance on how to improve programming in direct response to the evaluation questions and purpose of the evaluation as stated in the evaluation statement of work (SOW). USAID guidance states that recommendations should be supported by a specific set of findings as well as being action-oriented, practical, and specific, with defined responsibility for the actions to be taken.

To understand the extent to which evaluation recommendations meet USAID guidance, the study team analyzed recommendations from 45 evaluation reports. These 45 reports are a randomly selected subset of the 237 evaluations for which evaluation utilization surveys were sent to USAID staff. The study team conducted both qualitative and quantitative analyses for this sub-study. Quantitative analyses included the counts of recommendations in the report. Qualitative analyses were conducted on a randomly selected subset of recommendations. Due to the fact that some evaluations included large numbers of recommendations (a total of 70 in one report), qualitative analyses only included the first

five recommendations from each report that included recommendations, resulting in 200 recommendations. To put recommendations in better context, 6 evaluations were randomly selected from the 45 for additional comparisons of recommendations in the body of the report to recommendations and information provided in executive summaries from reports.

Sub-Study Findings

While the major evaluation utilization related findings are presented above, additional findings are included below.

Quantitative Counts of Recommendations in Evaluations

The study team began by distinguishing between formal and informal recommendations, with formal recommendations being those included in distinct recommendations sections of evaluation reports. Informal recommendations refer to statements scattered throughout the evaluation report, using language such as “USAID should...”

Using this approach, the team conducted a count of evaluations. The number of formal recommendations in each of the 45 evaluations ranged from 0 to 70, with an average of 18. Four evaluations did not contain any formal recommendations and one provided them in paragraph form, preventing the team from being able to identify distinct recommendations. Only ten evaluations included informal recommendations, ranging from 2 to 18.

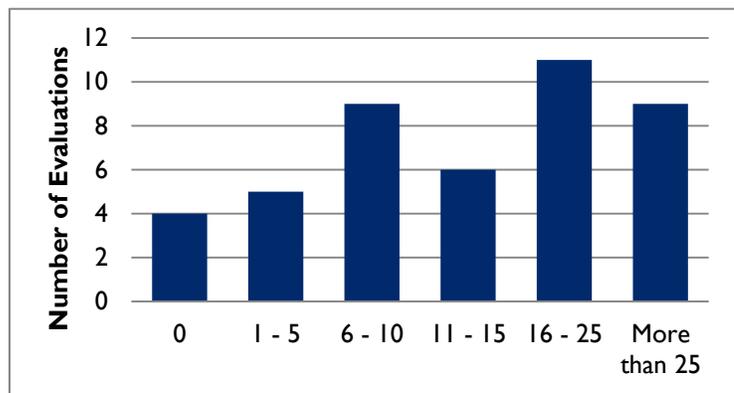


Figure 1: Distribution of Formal Recommendations (N= 44)

Qualitative Analyses of Evaluation Recommendations

For qualitative analyses, the study team identified the first five recommendations from each of the 40 evaluations that included easily identifiable recommendations, which resulted in a total of 200 recommendations. These recommendations were analyzed to assess the extent to which they met USAID policy guidance of being evidence-based, specific, practical, action-oriented, and directed to a specific audience. Due to the nature of recommendations, however, a formal content analysis could not be conducted as recommendations inherently require additional context from the evaluation report to determine factors such as practicality and use of evidence. Instead, the 200 recommendations were reviewed to identify overall perceptions from the study team. A subset of six randomly selected evaluations were reviewed in further depth by comparing recommendations in the body of the report to the recommendations in the executive summary, identifying from the executive summary whether sufficient evidence was provided and determining aspects such as specificity.

Overall Perceptions of Recommendations

The single strongest reaction that the study team had to recommendations was the extreme variability in their structure, content, and quality, indicating that insufficient guidance is being provided on what quality recommendations look like. Recommendations varied in length from being a single sentence to being nearly 1/3 of a page or more. While many recommendations were fairly succinct and direct, others were more rambling with findings and conclusions mixed in, or even other seemingly non-relevant recommendations being included.

Overall, recommendations appeared to be specific, practical, and action-oriented, as best could be assessed without additional context. When looking at being directed at certain audiences, recommendations fell short. It was fairly common for single recommendation statements to include multiple recommendations, often for different actors, if specific actors were even identified. In many cases, recommendations were simply statements, with no actor identifiable. In even more cases, ambiguous actors were identified, such as “USAID” or “the project.” In both of these cases the team could not tell who at USAID should be responsible, or what aspect of the project was being referred to; either the USAID or implementing partner side. The lack of distinct recommendations or clear actors is particularly problematic for Missions wishing to create post-evaluation Action Plans, should recommendations be buried in narrative or without clear audiences for each recommendation.

Another issue identified in recommendation statements was the inclusion of unclear, unnecessary, or impractical recommendations, or even statements that did not actually include a recommendation. This was manifested through vague language, the restatement of common knowledge, or recommending actions that are not feasible for cost, timing, or political reasons. For example, “The length of overseas assignments for [Mission] personnel is obviously key, and this too needs to be thought about more.”

The review did find a number of examples of high quality recommendations which could be used as a model for guidance moving forward. These statements largely did well at concisely providing specific action-oriented recommendations to discrete audiences:

- USAID/[Mission] must ensure that implementers have sufficient budget and capacity to design and carry out longer, more hands-on trainings to truly build capacity that can affect an entire sector or community. Programs should also incorporate more rigorous training evaluation frameworks/components to gauge the impact of capacity building activities. In addition, projects should seek to build local capacity to conduct future trainings once foreign experts leave/donor funding ends.
- [Project] should address transparency in party financing in order to promote a level playing field for all candidates and prevent the laundering of corrupt and other criminally sourced funds through political party and candidate financing. This area needs a comprehensive program to help prevent crime and illicit funding of political activities by transnational organized criminal networks, including terrorist cells.
- In the design of future programs, USAID and implementing partners should increase the focus on vulnerable groups such as youth and high-risk women. While, [Project] did implement activities directed at these vulnerable populations in selected regions, the success of these activities provides evidence that they should be replicated and further developed for additional regions, perhaps with even a national focus. Accordingly, future projects should continue to focus family planning and teen reproductive health, which are areas particularly salient for youth and high-risk women.

Findings from Executive Summary Reviews

For the six evaluations randomly selected for deeper analysis, the study team read the executive summaries with a focus on whether the recommendations included in them were sufficiently evidence-based, as this was a factor the team could not assess from looking solely at recommendation statements on their own. Part of the reason for this additional analysis came from interviews with Mission Directors and Senior Staff from Missions indicating that frequently these were the only portions of an evaluation they would see, aside from presentations.

The team found that only half of the executive summaries provided sufficient support for recommendations, with the remaining three evaluations providing support for only some of the recommendations, or not clearly distinguishing recommendations within the executive summary. Though this is too small of a sample to make any significant claims, and the body of the report may

provide substantially more evidence than the executive summary, the results indicate that recommendations overall are not meeting USAID expectations for being evidence-based.

When comparing the recommendations in the executive summaries to the recommendations provided in the body of the report, four of the six reports had the same recommendations in both places, though nearly all of them went into greater detail in the body of the report, providing greater specificity. In the remaining two cases, discrete recommendations were difficult to identify as they were mixed in the narrative.

In one of the executive summaries, the team found an excellent example of how to present potentially complex recommendations in a simple manner. There were 13 recommendations, addressing three different aspects of the project or activity, which were relevant for eight different audiences. The evaluation team presented their recommendations in a matrix format, with additional information provided in the body of the report. A copy of that recommendations table is provided below in Figure 2.

On the other side of the spectrum, one of the executive summaries reviewed did not do very well at presenting a large number of recommendations. This evaluation report had 20 recommendations. With this many recommendations it was quite difficult to provide sufficient support within the executive summary to justify each of them. These recommendations lacked specificity and an evidence base, and more importantly seemed to go into a territory of impractical recommendations.

Figure 2: Example Recommendations Matrix

No.	Recommendation	Worldreader	Future Evaluation Team	Donor Partners (ex: USAID)	Government Agencies (ex: MoE/GES)	Manufacturers (ex: Amazon)	Microfinance Organizations & Potential Device Distributors	Publishers	Teachers/ Administrators/ Students
Methodological/Study Design Recommendations									
1	Use a genuine random sample for the purposes of drawing conclusions on Ghanaian public schools	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>						
2	Limit influences and exposure within the control group	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>						
3	Modify data collection tools to capture data on student reading habits	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>			
Programmatic Recommendations									
4	Focus on primary schools and OCE activities to maximize benefits	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				
5	Pilot potential funding mechanisms to explore sustainability	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>		
6	Gain greater stakeholder buy-in	<input checked="" type="checkbox"/>							
7	Integrate e-readers more fully into the entire curriculum	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
8	Reduce logistical challenges for iREAD 2011-2012	<input checked="" type="checkbox"/>							
9	Continue to build the capacity of teachers so that the tool is used to its maximum potential	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>
10	Introduce e-readers to teacher training colleges and teachers' unions	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>				
11	Expand iREAD Activities to underserved areas	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				
Technological Recommendations									
12	Reduce the number of e-reader breakages	<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>			
13	Develop an improved e-reader management system	<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>			



EVALUATION UTILIZATION SUB-STUDY 7:

Standard Mission Orders on Evaluation

INTRODUCTION

Under the Evaluation Utilization Study for USAID/PPL/LER, the study team conducted a series of sub-studies looking at distinct evaluation-related processes or products to identify whether and how they might either utilize or encourage utilization of evaluations across the Agency. Each sub-study is keyed to one of the boxes included in the Evaluation Utilization Study Theory of Change. This particular study is keyed to Box 11 of the Utilization Study's theory of change in Figure 1 of the methodology section, and is focused on USAID's Standard Mission Order (MO) on Evaluation.

MAIN FINDINGS

The Evaluation Utilization Study team conducted a review of 41 customized MOs on Evaluation and identified the following major findings where specific language was added in an effort to improve evaluation utilization in Missions. Of the 41 MOs reviewed:

- Three encouraged staff to consider alternative dissemination methods, beyond DEC submission, to reach broader audiences including: the sharing of abstracts or recordings, uploading to non-USAID websites, offering public presentations of findings, or emailing summaries of findings directly to stakeholders.
- One encouraged incorporation of a partner Mission government representative on the team to encourage Mission buy-in to the evaluation and results.
- One included an appendix on the "Steps of the Evaluation Cycle" with specific utilization-focused steps highlighted.
- Six clarified responsibilities of the M&E Working Group, including at least one of the following: learning, dissemination, evidence-based decision making, or acting on recommendations.
- Two encouraged staff to engage with evaluation communities to exchange knowledge, share best practices, and ensure findings are integrated into strategies, programming, and project design.

PURPOSE, SCOPE, AND METHODOLOGY

To understand how Missions support evaluation utilization and what best practices emerged from Mission thinking, the study team reviewed 41 customized MOs on Evaluation, downloaded from ProgramNet in early 2015. MOs on Evaluation allow Missions to take Agency guidance and customize it within the Mission's context and culture. The study team conducted both qualitative and quantitative analyses of modifications to MOs on Evaluation, using content and pattern analyses for qualitative research and quantitatively using frequency counts of modifications on both Mission and MO sub-section bases.

Any MOs on Evaluation not uploaded to ProgramNet have not been included in this study. Additionally, quantitative counts of modifications to MOs do not appear to be indicative of whether and how utilization might be affected, so result from that analysis are not included below.

SUB-STUDY FINDINGS

While the major findings of the study relating to improving utilization at Missions were presented above, additional findings that may affect evaluation utilization are detailed below and are organized around nine main areas of focus. These include:

- Responding to evaluation findings & post-evaluation action tables
- Mission management
- Evaluation data
- Evaluation team
- Evaluation oversight
- Dissemination
- M&E Working Group
- Evaluation Statements of Work
- Evaluation planning

Responding to Evaluation Findings & Post-Evaluation Action Tables

In USAID's Standard Mission Order on Evaluation, there is a paragraph that addresses a section in USAID's ADS that describes actions to be taken after evaluations are completed. The basic wording for this standard paragraph is shown below:

Responding to Evaluation Findings

Following the completion of an evaluation, the PO will lead relevant Mission staff through the process of responding to evaluation findings as described in ADS 203.3.1.9. This process will lead to the development of an action plan for addressing the evaluation findings, conclusions, and recommendations the Mission has accepted or any other actions it deems necessary as a result of evaluation findings, conclusions, and recommendations. If necessary, the Mission may submit a statement of differences as an annex to the evaluation report. During Portfolio Reviews, the status of action plans for evaluation findings and their use in respective decisions will be discussed and documented (see Portfolio Review MO).

Inclusion of this paragraph in MOs that Missions and Regional Missions have shared indicates an awareness of these post-evaluation expectations. Across the 41 MOs the study team reviewed, it found 36 bilateral Missions and four Regional Missions that included this paragraph or most parts of it in their MOs; it should be noted that on paper the West African and Ghana Missions share one MO on Evaluation, making the total number of MOs that included a paragraph of this sort 37. One other Mission, USAID/India, included much of this information, but presented it separately under a section called Utilization. The India Mission was one of only five that used the word "utilization" in their Mission Orders on Evaluation.

A post-evaluation action table identifies specific recommendations accepted by a Mission and assigns staff the responsibility of taking action on each recommendation. In addition to an actor, a timeline is often also included. This analysis identified three Missions that provided templates for such tables as appendices to their MO. One of these three countries went a step further and created a figure titled "Steps of the Evaluation Cycle," which included the use of this action table as one of the critical steps.

Evaluation Data

The submission to USAID of raw data collected during an evaluation is required in USAID's Evaluation Policy and in ADS 579. An underlying reason for this requirement is that USAID may find future reasons to reanalyze or otherwise use the raw data provided. Having such data available at the Mission may

increase the likelihood of future utilization. Two Missions added language to their MO which required the submission of evaluation data to the Mission.

Dissemination

In order for an evaluation to be used beyond the Office that commissioned or managed it, the evaluation report, or at least findings and recommendations, must be disseminated to a wider audience. This audience may include USAID staff from other Offices or Missions, implementing partners, or other stakeholders such as partner Mission governments. Four Missions were identified as having modified their MOs to better address evaluation dissemination in a variety of ways.

Three of the four Missions that mentioned dissemination encouraged staff to consider alternative dissemination methods beyond traditional submission to USAID's Development Experience Clearinghouse (DEC) or posting to the website. Such alternative methods included the sharing of abstracts or recordings, uploading the reports to the internet beyond USAID websites, offering public presentations of findings, and emailing summaries of findings directly to stakeholders.

All four Missions included language regarding the public dissemination of evaluation findings, with three of them requiring it and the fourth indicating that public dissemination, via the DEC, would be at the discretion of the Program Office and DO teams. One of the three Missions requiring public disclosure further required that exemption to this Mission policy would only be granted following a written request directly to the Mission Director and approved by the Program Office.

Two of the four Missions included language requiring internal Mission-wide dissemination via the Intranet, with one of them additionally requiring an internal discussion of findings either by organizing an evaluation after-action meeting or introducing the findings during periodic project or portfolio reviews. One other Mission stated that intended dissemination approaches should be considered during the evaluation statement of work (SOW) development stage.

Evaluation Planning

Encouraging utilization of evaluation results can begin very early in the evaluation cycle, and as early as project design or CDCS stages, by ensuring a clear use for every evaluation conducted. Ten Missions incorporated language that might affect utilization, starting at evaluation planning stages.

Three of the ten Missions referenced potential triggers for evaluations which staff should be aware of. Such triggers include the upcoming designs of projects in new sectors or strategic areas, projects with perceived performance issues, or key upcoming management decisions with inadequate information, among others.

Two of the ten Missions stated that additional non-required evaluations could only be added to the Mission Evaluation Plan once the intended use for that evaluation was clear. One of these further stated that, during the individual evaluation planning process, the intended utilization of that evaluation must justify the cost and effort involved with conducting the evaluation, or it could not proceed.

Two of the ten Missions incorporated language requiring annual updates to the Mission Evaluation Plan to more accurately report upcoming evaluations and intended uses in the PPR's Evaluation Registry. Another Mission mentioned holding an annual evaluation planning meeting to identify priority evaluations and Mission-level evaluation questions. One last Mission mentioned that, even in the absence of a CDCS/RDCS, a PMP was still required to cover the projects and activities at that Mission, which would include the planning of any evaluations.

Evaluation Team

The composition of an evaluation team may have a direct effect on whether and how an evaluation is utilized, either by lending credibility of experts to the evaluation results, providing an internal champion to see recommendations through to completion, or by garnering support of external stakeholders through their engagement. Four Missions elected to include additional language in their MOs which touch on the composition of evaluation teams.

Three of the four Missions stated that Mission staff should be included as evaluation team members in some capacity. Specifically, in two different Missions, roles relating to evaluation were established, and those individuals were the ones to be included on evaluation teams, at the Program Office's discretion. One Mission indicated that non-required evaluations should be conducted internally, to build evaluation capacity of staff, with external consultants brought in as needed. The fourth Mission encouraged incorporating a partner Mission government representative on the team to encourage Mission buy-in to the evaluation and results.

M&E Working Group

The standard MO, from which each Mission adapted unique versions, included language on the development of an M&E Working Group. While many Missions deleted or did not modify this section of the MO, eight Missions chose to improve upon the standard language provided.

Six of the eight Missions opted to further clarify the roles and responsibilities of the M&E Working Group to facilitate utilization. These Missions included language on "identifying opportunities for learning," "disseminating and utilizing findings," "promoting decision-making based on high-quality M&E evidence," or "acting on findings and recommendations."

Two Missions felt that additional support beyond the M&E Working Group was needed, and created subgroups to provide more specific support for creating high quality evaluations and increasing evaluation capacity at their Missions.

Mission Management

Senior Mission management fostering a strong culture of evaluation at a Mission is thought to be one of the most effective ways to encourage evaluation utilization practices. Seven Missions used their MO to encourage this kind of support. Five Missions did so by establishing evaluation or M&E POCs (EPOCs or MEPOCs) in technical or DO Offices to work with EPOCs or equivalents at the Mission level.

Two additional Missions, of the seven total, specifically reference the role of Mission management in supporting a culture of evaluation either by encouraging staff to participate in relevant evaluation communities to exchange knowledge and share best practices or to ensure evaluation findings are integrated into decision making about strategies, program priorities, and project design.

Evaluation Oversight

Another means of increasing the likelihood of an evaluation being utilized is by improving how USAID oversees the implementation of an evaluation. Adding quality control checkpoints into the evaluation management process can increase the quality of evaluation products and therefore the likelihood of use. Seven Missions clarify the roles and responsibilities of staff in the oversight of evaluations at various stages of the evaluation cycle, ranging from SOW development to dissemination of reports. One of the seven went so far as to establish a small internal "evaluation team" to support the AOR/COR in reviewing products and deliverables and providing general oversight.

In two of the seven Missions, MOs added quality control checkpoints that did not directly relate to roles and responsibilities of staff, but rather steps in the evaluation cycle that should be included. One Mission includes in its evaluation process the use of a validation workshop to present findings to stakeholders prior to finalization of the report. The other Mission includes an evaluation report quality checklist as an appendix to support the production of high quality evaluation reports.

Evaluation Statements of Work

Nine Missions adjusted the language in MOs to affect how evaluation SOWs are written and reviewed. One Mission was so direct as to encourage the consideration of dissemination, and more specifically name dissemination methods, at the SOW development stage.

Five of the nine Missions chose to focus not only on how SOWs are written, but also provided guidance on who should be involved in writing SOWs and how they should be written. Two Missions require an evaluation specialist's involvement in the SOW writing process, with one going so far as to state that only graduates of USAID's course on Evaluation for Evaluation Specialists could write SOWs. Two other Missions provided a template for the drafting of SOWs as an annex, one of which provided written guidance on the actual writing process. Another Mission chose to direct staff to a USAID document on SOW "good practice examples" for reference while developing a SOW.

Five of the nine Missions chose to focus on the SOW review process, with three of them including a SOW quality review checklist as an annex and one of them requiring its use. A different annex was provided in another instance which provided technical guidance to Mission staff on how to peer review a SOW. In other cases, guidance was provided on who should review a SOW. One Bilateral Mission requires SOWs to be reviewed by its Regional Mission while another requires the review be done by an M&E specialist. A third Mission indicated that it may require that SOWs for any evaluations of G2G projects be reviewed by partner Mission government representatives to ensure that the host government is supportive of the planned evaluation.



EVALUATION UTILIZATION SUB-STUDY 8:

Evaluation Utilization Monitoring Systems

INTRODUCTION

Under the Evaluation Utilization Study for USAID/PPL/LER, the study team conducted a series of sub-studies looking at distinct evaluation-related processes or products to identify whether and how they might either utilize or encourage utilization of evaluations across the Agency. Each sub-study is keyed to the boxes included in the Evaluation Utilization Study Theory of Change. This particular study is keyed to Box 6 of the Utilization Study's theory of change, found in Figure 1 of the methodology section, and is focused on how evaluation utilization is monitored.

MAIN FINDINGS

The Evaluation Utilization Study team reviewed a number of studies on evaluation use and evaluation development trends; interviews and surveys with USAID staff; and evaluation monitoring trackers provided by USAID to understand how USAID compares to the development community in terms of monitoring utilization of evaluation results. The following are major findings identified from this review:

- USAID inconsistently tracks acceptance or implementation of recommendations, with some Operating Units (OUs) not tracking them at all, while others track them at either the evaluation level or at the Mission level.
- USAID's Annual Performance Plan and Report (PPR) tracks use of evaluations, but current structural issues within the Agency appear to prevent such monitoring.
- Utilization of evaluation evidence is easier to monitor for project/activity modification and design than for development of strategies and policies.
- "About two thirds of OECD DAC members have a mechanism in place to ensure management responds to and follows up on evaluation findings."⁴⁷

PURPOSE, SCOPE, AND METHODOLOGY

Evaluation utilization monitoring systems provide development organizations with critical information on whether and how evaluations are being used, and therefore that evaluation budgets are not just wasting valuable financial and staff resources. When structured effectively, such systems can also provide stories on how evaluations are used to improve program performance and even development outcomes.

The study team identified several studies commissioned by USAID, the Organization for Economic Cooperation and Development/Development Assistance Committee (OECD/DAC), and independent authors which touched on how various donor organizations approach the issue of evaluation and evaluation utilization. Information from these studies was compared to primary data collected for this current Evaluation Utilization Study, including actual evaluation recommendation trackers provided to the team by OUs.

⁴⁷ <http://www.oecd.org/dac/peer-reviews/12%20Less%20eval%20web%20pdf.pdf>

SUB-STUDY FINDINGS

In addition to the major evaluation utilization-related findings presented above, the study team also identified the following findings regarding how evaluation utilization is being monitored both at USAID and among the development community writ large.

Evaluation Utilization Monitoring at USAID

USAID currently has in place guidance and systems that could, if strengthened, provide an effective monitoring mechanism to track the utilization of evaluations within USAID.

USAID Guidance on Post-Evaluation Action Planning

USAID guidance, provided in its Automated Directives System (ADS), clearly encourages USAID to hold meetings following the completion of an evaluation. Such meetings are intended to provide the platform for USAID to systematically accept or reject evaluation recommendations. This guidance, however, encourages but does not require such meetings or actions take place, as they use “should” instead of “must” when providing guidance. Specifically, ADS 203.3.1.9 suggests that Missions should:

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 the 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, 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.

What is not included in the ADS is any discussion of how to formally track recommendations once accepted, or how to follow up on those recommendations in the future to determine whether they were implemented. Guidance is provided elsewhere, such as in USAID’s recently released Evaluation Toolkit and on USAID’s Learning Lab and Project Starter websites. These sources provide guidance in terms of Post-Evaluation Action Plan templates and trackers as well as a USAID Evaluation Resource document on utilizing and learning from evaluations. The tools provided in the Evaluation Toolkit for accepting recommendations and tracking recommendations are provided below.

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		

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						

In addition to these templates, now shared Agency-wide, USAID staff interviewed or surveyed provided the team with similar trackers created and used in their OUs. Such tools were received from eight Missions; these tools were largely applicable to specific evaluations and were not necessarily used consistently within the Missions.⁴⁸ USAID/Afghanistan was the only Mission to provide a tracker that appeared to track recommendations and utilization across the entire Mission’s evaluation portfolio, though other Missions provided summary tables indicating how recommendations had been acted upon, suggesting that some formal mechanism was in place. Interestingly, several Missions appear to be tracking recommendations in a disaggregated manner, categorizing recommendations based on their focus, such as recommendations focused on a particular sector or on aspects of effectiveness, efficiency, sustainability, etc. Common items tracked through these tools were the recommendation, USAID response/action, person responsible, deadline, and status.

USAID’s Annual Performance Plan and Report

Another mechanism USAID has to aid in tracking how evaluations are used is the PPR. The PPR, while not able to track on a recommendation-by-recommendation basis, can track evaluation utilization on an evaluation basis. On an annual basis, using the PPR Evaluation Registry, USAID OUs are required to provide an update on all planned, conducted, or completed evaluations in that year. For completed evaluations, guidance states that the PPR should, “[d]escribe specific ways the evaluation findings were used to inform decisions or fill knowledge gaps [including] a description of any policy, program management, budgetary or other decisions and changes made as a result of the evaluation findings.”

According to interviews and validation workshops with USAID staff in Washington, several issues exist with using the PPR’s Evaluation Registry in this manner. The first issue is that there is low confidence in the claims of use in the PPR, as staff are often too overburdened with other responsibilities to actually

⁴⁸ Missions were requested to provide these tools if they exist, but were not required to do so. Any tools provided were done so voluntarily, suggesting that similar tools may exist beyond those provided to the team.

update information, and instead simply restate what they had entered the previous year, which often comes directly from the purpose statement in the evaluation statement of work.

Another identified issue has to do with timing of evaluations in regards to the PPR. PPRs are updated annually after the close of the fiscal year, reflecting all evaluations from that year; PPRs do not include evaluations from previous years. For evaluations completed close to the end of the fiscal year, there may not be any reportable use yet. According to survey responses, there may be a lag of up to two years between approval of an evaluation and actually using that evaluation.

Evaluation Utilization Monitoring in the Development Community

To understand how other development agencies approach the issue of monitoring evaluation utilization, the study team drew on four studies that discussed the topic. The first study, Trends in International Development Evaluation Theory, Policy and Practice, was written for USAID in 2009 and shows that even then major donors were formally tracking utilization. According to that study, “a number of donors have also strengthened their guidelines for post-evaluation follow-up and several organizations (SIDA, World Bank, DFID and UNDP) are monitoring the frequency with which evaluation recommendations are accepted and, once accepted, whether or not they are implemented and the degree of thoroughness of the implementation.”⁴⁹ That study went on to provide examples demonstrating that the World Bank systematically collects and analyzes data on evaluation utilization, citing its 2008 Annual Review of Development Effectiveness with quotes such as, “95 percent adoption/implementation (of those accepted) up from 85 percent the prior year.”

Other studies, such as two released by OECD/DAC, took a more in-depth look at evaluation practices at a wider variety of organizations. These studies referred to “management responses,” which typically consist of a written, formal response identifying the agreed-upon follow-up action on recommendations made by evaluators. According to one study, 20 of 24 agencies had a mechanism in place to ensure such management responses are drafted in response to evaluation findings, and that follow-up actions take place, even if the mechanisms did not always work well; 17 percent claimed that they did not work well. Some DAC members, such as the Asian Development Bank, published these management responses alongside the evaluations on their websites.⁵⁰

The other DAC study, published more recently, claimed that about two-thirds of DAC members have such mechanisms in place and goes on to highlight how each approach is used.

⁴⁹ Hageboeck, Molly. “Trends in International Development Evaluation Theory, Policy and Practices.” Management Systems International (August 2009). http://pdf.usaid.gov/pdf_docs/PNADO464.pdf.

⁵⁰ OECD (2010), Evaluation in Development Agencies, Better Aid, OECD Publishing. <http://dx.doi.org/10.1787/9789264094857-en>

4.3 – LOGISTICAL REGRESSION ANALYSIS

Analysis for Evaluation Utilization Study, USAID FTF-KDAD-PPL

Aníbal Pérez-Liñán, University of Pittsburgh

Version 5 – November 29, 2015

Contents

1. Main Findings	144
2. Analytic Strategy	145
3. The Pooled Evaluation Sample (N = 310)	148
3.1. Instrumental Use	148
3.2. Conceptual Use	151
4. The Quality Score Sample (N = 241)	156
4.1. Instrumental Use	156
4.2. Conceptual Use and Dissemination	159
5. Analysis of the Survey Sample (N = 118)	162
5.1. Instrumental Use	163
5.2. Conceptual Use	168
5.3. Discernible Development Impacts	170
5.4. Summary Lessons and Conclusions	173

1. Main Findings

1. The analysis of a large sample of evaluations (N=310) shows that:
 - 1.1. **Conceptual Use** and Dissemination are key explanatory factors to understand the probability of **Instrumental Use**.
 - 1.2. **Dissemination** of results promotes the **Instrumental Use** of evaluations in two ways: directly—presumably by creating interpersonal and collective pathways of influence—and indirectly, by promoting greater levels of **Conceptual Use**.
 - 1.3. The evidence shows strong effects of **Dissemination** on **Conceptual Use** even after we account for the possibility of reverse causality (i.e., the possibility that learning will encourage further dissemination).
2. The **Average Quality** of the evaluations conducted by an Operating Unit is directly related with greater rates of utilization, particularly **Instrumental Use**. The analysis of an intermediate sample for which quality scores are available (N=241), however, provides no evidence in support of the hypothesis that the quality of individual evaluations influences their Instrumental Use, Conceptual Use, or even their Dissemination. This suggests that institutional process of evaluation planning and design, more than the quality of any individual report, facilitates utilization in some units more than others.
3. The study of a smaller number of evaluations for which information was collected through the **survey** instrument (N=118) helps clarify the previous findings:
 - 3.1. Statistical estimates suggest that involving a **country partner** in the evaluation expands the expected rate of **Conceptual Use** from 88% to 99%; that a **timely report** expands the probability of conceptual use from 70% to 98%; and that changing from a low-quality to a **high-quality environment** expands the rate of conceptual use for an Operating Unit from 57% to almost 100%.
 - 3.2. “**Proactive**” **post-evaluation** meetings promote greater levels of **Instrumental Use**. Proactive meetings involve additional steps to follow up on evaluation results (they accept recommendations, assign responsibilities, etc.). **Dissemination** of results to implementing partners, the capacity of an evaluation to generate **novel information**, and the **overall quality** of the average evaluation conducted by the Operating Unit also foster instrumental utilization:
 - Statistical estimates suggest that the adoption of proactive post-evaluation protocols increases the expected rate of utilization from 87% to 99%. Proactive meetings also expand the probability of observing high levels of instrumental use, measured as the implementation of more than 75% of all recommendations.
 - Dissemination to implementing partners expands expected rates of instrumental use from 83% to 94%.
 - Predicted utilization is about 87% when the report does not contain novel findings and 97% when it does.
 - The predicted probability of instrumental use is 47% for units working in a low-quality environment, while it is almost a 100% for Operating Units with access to top-quality evaluations.
 - 3.3. At the end of the evaluation life cycle, the main factors promoting discernible **Developing Outcomes** are an active pattern of instrumental use and **dissemination among local country partners**. Because development outcomes are determined by many external factors unrelated to the evaluation process, none of those favorable conditions by themselves is likely to push the expected probability of discernible

outcomes above 54%, but the combination of both factors increases the expectation of discernible outcomes to 63%.

4. Taken together, these results document a chain of effects from “lower” levels of utilization into “higher” levels of utilization:
 - 4.1. Officers display an ability to make conceptual use of recommendations when those recommendations are presented in a timely manner, if they incorporate the knowledge of local partners, and if evaluations are routinely conducted in a high-quality environment.
 - 4.2. In turn, officers are able to transform those recommendations into actionable plans by adopting a proactive post-evaluation protocol and by sharing results with implementing partners. Instrumental use is more likely when evaluations generate novel findings and in high-quality environments.
 - 4.3. Instrumental use, as well as dissemination to country partners, facilitates the translation of those plans into development outcomes. But this translation is mediated by multiple external conditions beyond any direct control, and therefore observed rates of discernible outcomes will be lower than conceptual or instrumental use.

2. Analytic Strategy

The data employed for the quantitative analysis in this study involve three overlapping samples:

1. **A Random Survey Sample** (n = 118): MSI conducted a survey based on a stratified sample of 237 evaluations. The survey yielded 134 responses, 16 of which referred to evaluations selected non-randomly. To avoid bias in favor of utilization, we dropped these cases from the analysis.⁵¹ The main advantage of this sample is the very large number of variables coded from survey responses, which include features of the evaluation process (e.g., partners involved, timeliness), post-evaluation activities (e.g., meetings, dissemination), and forms of utilization. The stratification procedure was designed to secure a broad representation of evaluations based on their scope and timing. Because this procedure oversampled smaller scope-timing configurations, we employed post-stratification weights in the statistical analysis to reflect the original distribution of evaluation types in the population. The use of weights had only marginal impact on the findings. For example, the rate of instrumental use is 92% in the unweighted sample and 90% in the weighted sample.
2. **Scored Evaluations** (n = 241): The second dataset combines a sample of 138 evaluations (from the PPL/LER commissioned 2013 Meta-Evaluation) and 103 evaluations for 2013-14 (from USAID E3’s 2015 Sectoral Synthesis). MSI coded all observations using a common evaluation report checklist and scoring system which includes, for example, whether the executive summary is accurate, whether an evaluation specialist was part of the team, or whether specific methods were used to address specific questions. To summarize these quality indicators, the MSI team developed an additive index ranging between 1 and 10. In this study, an average for this index was also computed for each Operating Unit to provide a measure of each unit’s overall access to quality reports. This sample includes a relatively small number of variables because survey responses are not available. By contrast to the survey, it is possible that this sample underestimates rates of utilization, because utilization

⁵¹ The 16 evaluations dropped from the survey sample were selected non-randomly for qualitative purposes because they guaranteed good examples of utilization.

has been coded from secondary sources (the rate of instrumental use in this sample is about 35%).

3. **The Pooled Sample** (n = 310): The two samples were pooled in some analyses to gain greater inferential leverage from a larger sample size. This pool includes 49 evaluations that overlap (i.e., they were randomly selected into both samples), 69 that only belong to the survey sample, and 192 that only belong to the scored sample. At this level, the number of cases is large but the number of variables is quite restricted because survey responses are only available for the first sample and quality scores are only available for the second sample. However, the dataset contains useful information for several descriptive variables, including the timing (mid-term, final, ex-post, continuous), the region, and the year of the evaluations. More important, information for a small number of relevant variables capturing instrumental use, conceptual use, and the dissemination of evaluations is available for all cases. These variables are described next.

The main variables employed in the quantitative analysis were coded using a combination of secondary sources and survey data. This information is available for all cases, even though, unfortunately, not every source could be used to score every evaluation:

Instrumental Utilization (eval_use). This dichotomous indicator was coded 1 when at least one of the following conditions was met (and 0 otherwise):

- Survey respondents indicated that at least some of the recommendations that were accepted were implemented (survey question 38);
- Survey respondents indicated that at least one concrete action was taken as a result of the evaluation (survey question 43);
- The evaluation was cited in a CDCS or RDCS strategy document to provide justification or action or a policy;
- The use statement for the evaluation in the PPR described a concrete action that had already been taken based upon the evaluation;
- Crowdsourcing narratives described an action that had been taken based upon the evaluation;
- The evaluation was used in an Evidence Summit, which was then cited in a policy document.

Conceptual Use (eval_use.f). This dichotomous measure was coded 1 when at least one of the following conditions was met (0 otherwise):

- The use statement in the PPR for the evaluation described learning or opinion change that occurred based upon the evaluation;
- Survey respondents indicated that they learned something or that their opinions changed because of the evaluation (at least one item was selected for survey questions 27 or 28);
- The evaluation was cited in a CDCS or RDCS strategy document to provide contextual information about the country or region.

Dissemination (eval_dism.both) was coded 1 when at least one of the following conditions was met (0 otherwise):

- Survey respondents indicated that evaluation findings and recommendations were disseminated to at least one stakeholder group in some way other than uploading the report online (survey question 23);
- The evaluation's PPR use statement described dissemination of the evaluation to other stakeholders.

To leverage the different sample sizes, the analysis was conducted in stages, moving from the most inclusive sample (n = 310, small number of variables) progressively into the more restrictive sample (n = 118, many variables). To represent the causal model implicit in the *Evaluation Utilization Conceptual Framework*, at each stage logistic regression was used to reconstruct a backward explanatory sequence, in which:

- Discernible Effects on Developing Outcomes are shaped by levels of Instrumental Use, Conceptual Use of evaluations, and additional facilitators, while in turn
- The Instrumental Use of evaluations depends on levels of Conceptual Use and on additional drivers; and finally
- Conceptual Use depends exclusively on prior drivers such as Dissemination. In section 3.2, an ancillary model is also estimated in which Dissemination is treated as an endogenous variable, to account for the possibility of reverse causation with conceptual utilization.

When possible, similar models were tested using different samples, including the pooled sample, in order to verify the consistency of the findings. Although each sample presents its own distinctive challenges (given the possibility that different sources may over- or under-estimate utilization), similar results across samples reinforce the credibility of the results.

QCA Analysis. In early stages of the study, several QCA (qualitative comparative analysis) tests were also developed, with quite limited results. The QCA protocol proceeds in four stages: (1) it creates a matrix for all possible combinations of the values of the independent variables (e.g., three dichotomous explanatory variables will create a matrix of 8 cells); (2) locates all cases in their respective cells; (3) identifies cells with a very high density for the outcome of interest (“true” configurations); and (4) combines these cells using Boolean logic to simplify the solution. QCA was in principle a promising technique to explore systematic interactions among survey responses, but two features of the survey undermined the applicability of the technique. First, the rate of instrumental use captured by the survey was above 90%, which created too many “true” cells. The large number of highly consistent cells undermined the capacity of the protocol to discriminate configurations and identify a meaningful solution. Second, the survey included a large number of questions, and the initial exploratory work considered a relatively large number of variables. This complicates the implementation of QCA, because the number of combinatory cells grows multiplicatively with the number of variables; as the classification grid becomes larger, many cells become devoid of cases. For this reason, QCA is often successfully employed in studies involving a moderate number of variables and a moderate number of cases (with the additional advantage for small-N studies that the protocol does not rely on significance tests).

3. The Pooled Evaluation Sample (N=310)

Although no measure of Impact is available for this larger sample, the number of observations offers an advantage to test general conditions for the Instrumental Utilization and Conceptual Use.

3.1. Instrumental Use

Table 1 below presents the parameter estimates for three logistic regression models that treat **Instrumental Utilization** (eval_use) as the dependent variable. In the first model (1.1), only Conceptual Use (eval_use.f), Dissemination (eval_dism.both), and the Mean Quality score observed for all evaluations commissioned by the Operating Unit (ou_scorequal) are included as predictors for this outcome.⁵² The Mean Quality score does not reflect the quality of the specific evaluation (a topic addressed in the next section), but the overall capacity of the unit to commission work consistent with USAID standards.

The three independent variables present large and significant coefficients. The effect for Conceptual Use indicates that learning and attitude change are crucial to promote the incorporation of evaluation lessons into further projects, strategies, and activities. In turn, dissemination of findings has a direct effect on the likelihood of Instrumental Use, making a positive contribution above and beyond any indirect influences exercised through greater levels of conceptual use (an effect documented below). This hints at the idea that dissemination reinforces interpersonal and collective processes, which transcend strict individual learning. Similarly, the coefficient for the Mean Quality score suggests that the greater the quality of *the usual* evaluations processed by an Operating Unit, the greater the probability that *any single* evaluation will yield instrumental use. This hints at the importance of better evaluation planning and design at the institutional level, beyond any specific case, as a driver of utilization.

⁵² Two evaluations included in the survey originated in operating units for which no quality scores are available; because the average quality for the operating unit is a predictor in all models discussed in this section, sample size reduces to N = 308.

Table 1. Models of Instrumental Use (Pooled Sample)

	(1.1) Baseline	(1.2) Evaluation Features	(1.3) Other controls
Conceptual use	2.07** (0.53)	2.15** (0.55)	2.86** (0.63)
Dissemination	2.37** (0.58)	2.36** (0.60)	2.27** (0.66)
Mean Quality OU	0.47* (0.19)	0.52** (0.19)	0.31 (0.23)
Timing (ref. = Mid-term)			
Final		-0.77* (0.36)	-0.87* (0.39)
Ex-post		-0.26 (0.93)	0.15 (1.01)
Continuous		-0.42 (1.69)	-0.65 (2.01)
Region (ref. = Af/Pak)			
Africa			-1.30* (0.59)
Asia			-1.41* (0.65)
Latin America			-1.35+ (0.70)
Middle East			-3.99** (1.03)
Europe and Eurasia			-1.51* (0.69)
Global			-3.95** (1.23)
Evaluation Year			0.01 (0.19)
Constant	-5.07** (1.44)	-5.09** (1.45)	-25.84 (376.17)
N	308	308	308
Pseudo-R ²	.46	.47	.52
Chi ²	195.95	200.62	221.06
Log-L	-115.20	-112.86	-102.64

Note: Entries are logistic regression coefficients (standard errors). + p<0.1, * p<0.05, ** p<0.01

To assess the robustness of these findings, Models 1.2 and 1.3 incorporate additional control variables. Model 1.2 includes a series of dichotomous variables to capture the timing of the evaluation process: Mid-term (the reference category), Final, Ex-Post, or Continuous. These features do not exercise great influence on levels of Instrumental Utilization, although final evaluations are less likely to encourage instrumental use—possibly because new phases of the project are already in progress by the time the evaluation is completed.

Model 1.3 incorporates dichotomous variables that reflect potential differences across USAID regional contexts, and controls for the year in which the evaluation was conducted to identify any time trends. The coefficients for Africa, Asia, Latin America, the Middle East, East Europe & Eurasia (E&E), and Global programs are all negative and significant (at the .05 or the .01 levels, with the exception of Latin America, just significant at the .10 level). This pattern indicates that evaluations for all regions have significantly lower rates of Instrumental Use than evaluations for

Afghanistan and Pakistan, the reference category. The observed rate of utilization is roughly 68% in Afghanistan and Pakistan, while it ranges between 57% and 23% for the remaining regions. These regional patterns are confirmed by additional multivariate analyses of Instrumental Utilization conducted for a smaller sample (presented in the next section, Table 3).⁵³

To provide a more intuitive interpretation of the coefficients presented in Table 1, Figure 1 plots the predicted rates of Instrumental Use based on Model 1.3. The panels in this figure display the probability of evaluation utilization for hypothetical samples of 308 cases in which the selected independent variables are fixed at the desired values (0 or 1 for Conceptual Use and Dissemination, a realistic range between 4 and 10 for the Mean Quality score) and all other variables are kept at their observed values in the existing sample. The expected probabilities based on the model are then averaged across the sample to produce a point estimate for this counterfactual population.

Figure 1.1 indicates that if no evaluation in the sample displayed any evidence of Conceptual Use, the rate of Instrumental Use would drop from the actual level of 40% observed in the sample to about 34%. By contrast, if all evaluations reflected some form of Conceptual Use, rates of Instrumental Use would increase to 79%.

⁵³ Alternative models incorporated a set of dichotomous variables to capture the evaluation's scope (Single Project/Single Country, Single Project/ Multi-Country, Sector-wide, Regional, Multi-Project/Single-Country, Multi-Project/ Multi-Country, or Global) and the project sector (Agriculture, Democracy & Governance, Education, Economic Growth, Health, Multiple Sectors, and Others). The results presented in Table 1 remained unaltered when those controls were included. These variables did not emerge as strong predictors, albeit in few models Sector-wide evaluations appeared to be somewhat more influential on average. To simplify the presentation of results, these factors were omitted in the final models presented in Table 1.

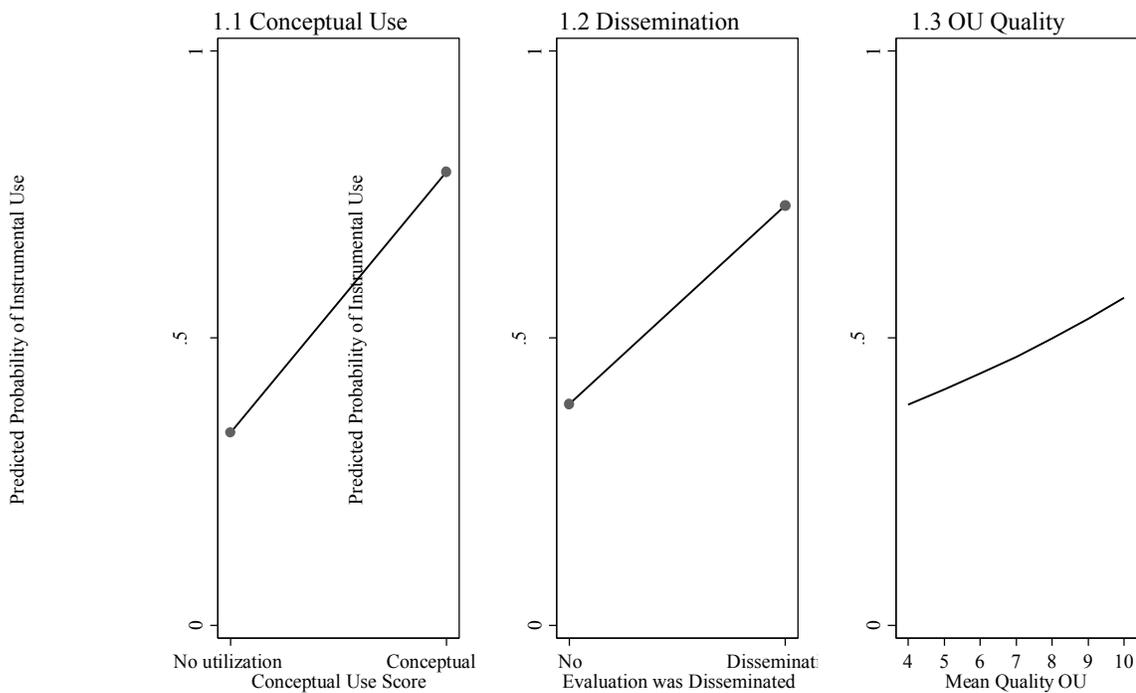


Figure 1. Predicted Probability of Instrumental Use (Based on Model 1.3)

Similarly, Figure 1.2 reflects the impact of Dissemination. In the absence of dissemination strategies, expected levels of Instrumental Use would drop to about 38%. By contrast, consistent dissemination of all findings would expand the predicted rate of Instrumental Use to 73%. Operating units with a Mean Quality score of 4 (the minimum value observed for an Operating Unit in the sample) have an expected rate of utilization of just 38%, while for those with a maximum Quality score of 10 the expected rate of utilization is 57%.

3.2. Conceptual Use

Table 2 below replicates the three model specifications used in Table 1, but treating **Conceptual Use** (eval_use.f) as the dependent variable. The estimates presented in Table 2 show that Dissemination has a very large impact on Conceptual Utilization, an effect visible in Figure 2.1 below. In the absence of Dissemination, the expected rate of Conceptual Use would drop to just 7%, while a universal practice of dissemination would expand the probability of Conceptual Use to 95%.

Table 2. Models of Conceptual Use (Pooled Sample)

	(2.1) Baseline	(2.2) Evaluation Features	(2.3) Other controls
Dissemination	5.79** (0.55)	5.90** (0.58)	6.35** (0.69)
Mean Quality OU	0.50+ (0.27)	0.48+ (0.27)	0.75* (0.30)
Timing (ref. = Mid-term)			
Final		0.16 (0.52)	0.14 (0.54)
Ex-post		-1.04 (1.10)	-1.12 (1.16)
Continuous		0.42 (2.24)	0.01 (2.23)
Region (ref. = Af/Pak)			
Africa			0.78 (1.07)
Asia			1.24 (1.12)
Latin America			1.49 (1.19)
Middle East			2.00 (1.38)
Europe and Eurasia			-0.10 (1.16)
Global			4.04* (1.76)
Evaluation Year			0.07 (0.26)
Constant	-6.45** (2.04)	-6.38** (2.06)	-147.54 (532.50)
N	308	308	308
Pseudo-R ²	.68	.68	.70
Chi ²	281.97	283.07	291.52
Log-L	-66.62	-66.07	-61.85

Note: Entries are logistic regression coefficients (standard errors). + p<0.1, * p<0.05, ** p<0.01

Table 2 also indicates that the Average Quality of evaluations commissioned by a unit influences its overall rates of conceptual use. Units with low-quality evaluations have expected rates of conceptual use close to 26%, while units with top-quality evaluations have expected rates of use around 56% (see Figure 2.3). As documented in the next section, this effect is not created by the quality of individual reports. It is not simply that USAID officials learn more from better reports, but rather that officials operating in a high-quality environment appear to be more willing (or able) to extract lessons from any evaluation.

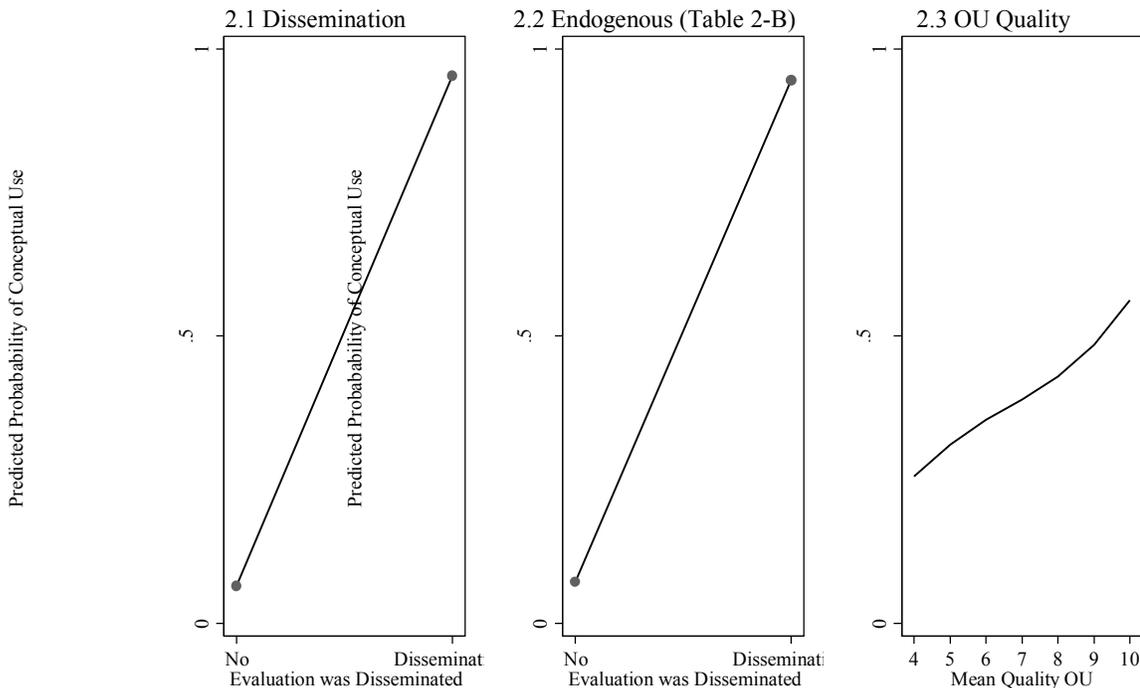


Figure 2. Predicted Probability of Conceptual Use (Based on Model 2.3 and Table 2-B)

Taken together, Figures 1.2 and 2.1 suggest that Dissemination promotes the Instrumental Use of evaluation findings in two ways: directly—presumably by creating interpersonal and collective pathways of influence—and indirectly, by promoting greater levels of Conceptual Use which in turn underpin the incorporation of evaluation lessons in further activities.

Reverse Causation. These results are potentially challenged by the possibility of reverse causation between Conceptual Use and Dissemination. While dissemination of findings naturally leads to greater learning, those who learn relevant lessons from an evaluation may also organize further events (including presentations, webinars, etc.) to disseminate these findings. Because regression models assume that causality is one-directional, this reciprocal effect could lead us to overestimate the effect of Dissemination on Conceptual Use in Table 2, or even misconstrue the correlation between Dissemination and Conceptual Use entirely, by assuming that causality flows from dissemination to learning when it may just flow in the opposite direction.

Unfortunately the cross-sectional nature of the data prevents a systematic exploration of this problem. Addressing this issue would require a more complex research design collecting data over the life cycle of particular evaluations, measuring the timing of dissemination and of related learning processes in real time. We can, however, employ econometric techniques to assess if reverse causality accounts exclusively for the strong effects of Dissemination documented in Table 2.

This issue can be conceptualized as a problem of “endogenous” dissemination. Dissemination may promote Conceptual Use, but evaluations more likely to inspire conceptual use will also be targeted for dissemination. Thus, any unmeasured factor (say, the presence of relevant findings) that makes Conceptual Use more likely will ultimately also make Dissemination more likely, potentially inflating our estimates for the one-directional effect of dissemination on learning.

To address this concern we need to decompose the overall variance of the independent variable (dissemination) into two components: a portion of the variance that cannot be attributed to reverse causation (i.e., a clearly exogenous component) and a residual (the remaining portion of the variance) which could in theory be driven by reverse causality. For example, there has been greater attention to issues of evaluation across USAID since 2011, following the release of its Evaluation Policy, and this has probably encouraged increasing rates of dissemination. Observed dissemination rates in this sample grew from 29% in 2011 to 41% in 2012, 33% in 2013, and 58% in 2014. By contrast, Model 2.3 (in Table 2) shows that there is no equivalent time trend for Conceptual Use—the coefficient for the Evaluation Year is statistically insignificant. Therefore, trends in conceptual use cannot explain this growing trend in dissemination through some form of reverse causation.

The Year of the Evaluation therefore is a good *instrument* to identify exogenous variance in dissemination rates, because Year is correlated with levels of dissemination but not with rates of learning. This instrument can be used to estimate a two-stage residual inclusion model (2SRI), presented in Table 2-B. This strategy is akin to other instrumental variable techniques such as two-stage least squares or treatment selection regression models.

Table 2-B. Model of Conceptual Use with Endogenous Dissemination

Dependent variable:	Stage1 Dissemination	Stage2 Conceptual Use
Dissemination		6.10** (0.92)
Mean Quality OU	0.02 (0.14)	0.77* (0.30)
Timing (ref. = Mid-term)		
Final	-0.30 (0.25)	0.14 (0.54)
Ex-post	0.91 (0.62)	-1.06 (1.17)
Continuous	1.85 (1.19)	0.08 (2.22)
Region (ref. = Af/Pak)		
Africa	-0.05 (0.44)	0.79 (1.04)
Asia	-0.26 (0.48)	1.25 (1.10)
Latin America	-0.36 (0.52)	1.50 (1.18)
Middle East	-0.28 (0.64)	2.01 (1.36)
Europe and Eurasia	0.44 (0.48)	-0.08 (1.14)
Global	1.21 (0.96)	4.04* (1.74)
Evaluation Year	0.32** (0.12)	
Control function [Residual of Stage 1]		0.18 (0.41)
Constant	-636.03** (245.76)	-9.49** (2.70)
N	308	308
Pseudo-R ²	.05	.70
Chi ²	20.90	291.63
Log-L	-194.54	-61.79

Note: Entries are logistic regression coefficients (standard errors). + p<0.1, * p<0.05, ** p<0.01
Year is the main instrument to identify the equation in Stage 1, and it is excluded in Stage 2.

The analysis presented in Table 2-B involves two related equations. The first one employs the instrument (the Evaluation Year) and the other exogenous predictors in Table 2 to estimate the probability of Dissemination, the endogenous variable. This model is presented in the column labeled “Stage 1”. The (standardized Pearson) residual of this equation represents the proportion of Dissemination that cannot be explained exogenously, and therefore could be (at least in principle) a product of reverse causation. We can use this residual as a crude approximation to the amount of reverse causation in our initial model of Conceptual Use. The second equation (Stage 2) therefore re-estimates Model 2.3 including a “control function” made of the residual from Stage 1. In this model, the effect of Dissemination remains very large and significant (in fact, the size of the coefficient is very similar to the one observed in Model 2.3), indicating that the effects of Dissemination remain sizeable even after we account for potential endogeneity. The middle panel of Figure 2 plots the marginal effect of Dissemination on Conceptual Use according to Table 2-B, showing that the substantive effect of this predictor is roughly equivalent

in Tables 2 and 2-B. Thus, although we cannot rule out the possibility of reverse causality with the existing data, the evidence suggests that the large effects of Dissemination shown in Table 2 are not a mere artifact of reverse causality.

4. The Quality Score Sample (N=241)

The quality score (QS) sample includes a larger number of variables coded from the scoring checklist for evaluations. Those items provide additional information on the nature and quality of the evaluation product gathered from the reports. The team also created an overall quality index based on 11 checklist items; observed values for this index range between 2 (low quality) and 10 (high quality).

This section incorporates the product quality indicators to the models for Instrumental Use and Conceptual Use developed in the previous segment of the report. The number of observations in the sample drops considerably, but the models offer the possibility of developing a more nuanced understanding of the utilization process.

4.1. Instrumental Use

Table 3 reports logistic regression estimates for models of **Instrumental Use** (eval_use) incorporating the quality indicators. Model 3.1 includes the main predictors from the previous section (Conceptual Use, Dissemination, and the Mean Quality score) plus the composite index based on 11 checklist items. The difference between the Quality Score and the Mean Quality score is that the first variable captures the quality of the specific evaluation under study, while the second one is an average for all evaluations coded for the Operating Unit.

Table 3. Models of Instrumental Use (QS Sample)

	(3.1) Quality Index	(3.2) Items	(3.3) Index & Controls
Conceptual use	2.42** (0.72)	2.66** (0.75)	2.82** (0.80)
Dissemination	1.58+ (0.85)	2.10* (0.94)	1.62+ (0.92)
Quality Score	-0.03 (0.12)		-0.02 (0.13)
Mean Quality OU	0.43+ (0.24)	0.55* (0.26)	0.25 (0.27)
Items			
Accurate Summary		0.14 (0.46)	
Method Fit		-0.35 (0.51)	
Evaluation Specialist		0.32 (0.45)	
Social Science Methods		-1.26* (0.64)	
Distinct Findings		-0.03 (0.47)	
Rec. for Specific Party		-0.63 (0.42)	
Timing (ref. = Mid-term)			
Final			-0.88* (0.41)
Ex-post			0.50 (1.26)
Continuous			-0.78 (2.26)
Region (ref. = Af/Pak)			
Africa			-1.49* (0.62)
Asia			-1.49* (0.67)
Latin America			-1.36+ (0.71)
Middle East			-3.58** (1.10)
Europe and Eurasia			-1.35+ (0.71)
Global			-0.79 (2.03)
Evaluation Year			0.07 (0.21)
Constant	-4.59** (1.53)	-4.39* (1.73)	-144.14 (425.76)
N	241	216	241
Pseudo-R ²	.36	.42	.42
Chi ²	112.03	119.49	130.96
Log-L	-99.80	-81.53	-90.34

Note: Entries are logistic regression coefficients (standard errors). + p< .10, * p<0.05, ** p<0.01

Model 3.2 replaces the aggregate index with a set of dichotomous indicators capturing the accuracy of the executive summary (eval_score1), the use of specific research methods to address particular evaluation questions (eval_score9), the presence of an evaluation specialist in the team (eval_score13), the consistency between the methods employed and the findings presented (eval_score20a), a clear distinction between findings and recommendations (eval_score23), and whether recommendations meet USAID expectations with respect to being directed to specific parties (eval_score31a). Evaluations coded as N/A for any of those items were dropped from the analysis.

Although Model 3.2 presents all variables at once, alternative models with different combinations of those dichotomous indicators generated equivalent results. Finally, Model 3.3 combines the aggregate index with additional controls for the timing of the evaluation, the region, and the year.

The results presented in Table 3 show that the findings discussed in the previous section, which underscore the role of Conceptual Use, Dissemination, and aggregate Quality at the unit level remain valid even when we estimate those effects with a much smaller sample. However, Table 3 shows no statistical evidence in support of the claim that formal quality features influence the Instrumental Use of particular evaluations. Neither the composite index nor the individual quality indicators display any significant association with the dependent variable.

This null finding does not prove that quality is irrelevant for the usefulness of evaluations, but it shows that the quality of specific reports does not have a systematic association with their observed rates of instrumental utilization. It is possible that officials are using the evaluations they receive as much as they can, even though their ability to adjust programs effectively is affected by the quality of the reports. It is clear, however, that officials are better able to engage in utilization when the *average* report in their units is of better quality. This point deserves further exploration with qualitative evidence.

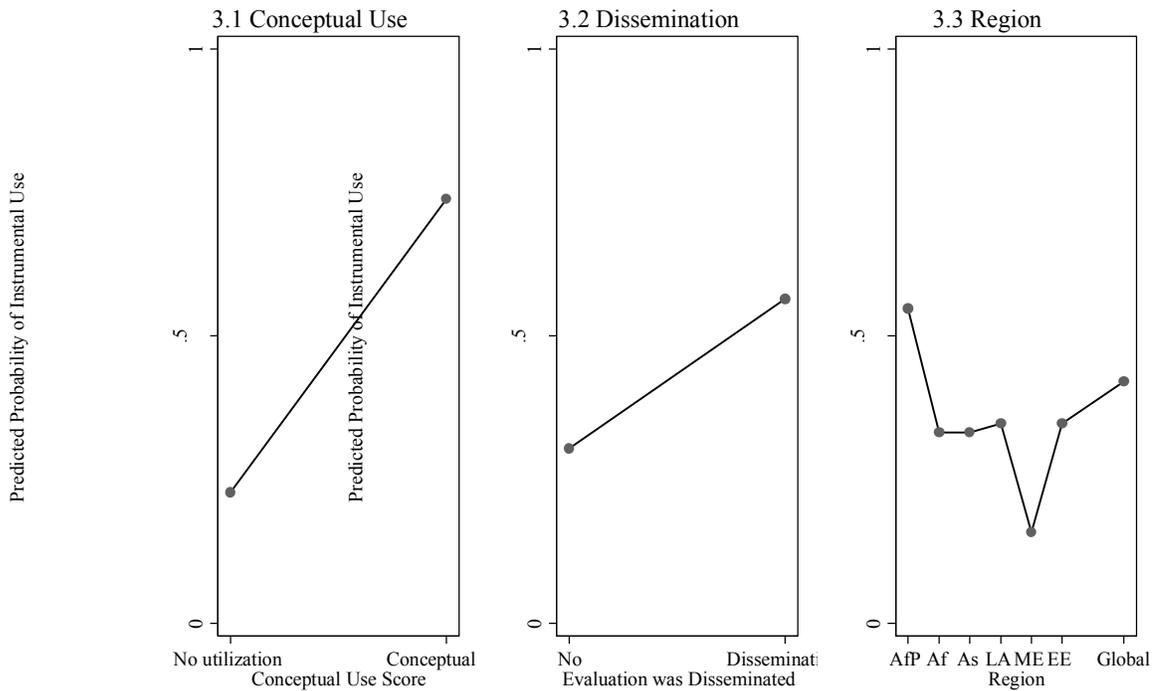


Figure 3. Predicted Probability of Instrumental Use (Based on Model 3.3)

Figure 3 presents graphical representations for the marginal effects of Conceptual Use and Dissemination, confirming the previous findings for the pooled sample summarized in Figure 1. The effect of Mean Quality is, however, insignificant in Model 3.3, after we include regional dummies in the equation. This probably reflects the fact that Afghanistan and Pakistan present the highest quality scores (7.9 on average) while the other regions present a somewhat lower average (7.2) and much greater variance. At the same time, Model 3.3 also suggests—in line with previous findings in Table 1—that evaluations for Afghanistan and Pakistan induce more instrumental use than evaluations for other regions. This is evident in Figure 3.3, which shows that only evaluations in this region present—assuming that other variables do not change—a predicted rate of instrumental use above 50%. Therefore, the effects of average quality at the Operating Unit level and regional effects are hard to disentangle in Table 3.

4.2. Conceptual Use and Dissemination

It is plausible that the quality of evaluations could influence Instrumental Use indirectly, by affecting learning and attitude change. To assess this possibility, Table 4 presents three models in which **Conceptual Use** (eval_use.f) is the dependent variable. The three models in this table replicate the specifications of Table 3, incorporating the composite quality index for specific evaluations (4.1), the individual quality items (4.2), and the index plus additional controls for the timing of the evaluation and the region (4.3). Unfortunately, the measure of Dissemination predicts Conceptual Use perfectly and therefore it is dropped from the analysis.

Findings for Conceptual Use in Table 4 are similar to the ones already presented for Instrumental Use in Table 3. The models fail to provide any evidence in support of an explanation based on

the quality of individual evaluations. Moreover, estimates for the effect of Mean OU Quality fail to achieve conventional levels of statistical significance (possibly due to the smaller sample size).

To assess alternative mechanisms of indirect influence, additional models used the quality index, the quality traits, and additional controls (scope of evaluation, timing, sector, region, year) as independent variables to estimate the probability of **Dissemination** (eval_dism.both). In line with null findings just reported in Tables 3 and 4, quality features (and almost every other explanatory variable) showed no connection with the probability of dissemination. The results of this ancillary analysis are not presented here to save space, but the table is available upon request.

To reiterate the results presented in Table 2-B, the only apparent trend in the dissemination model is an increasing propensity of USAID to disseminate evaluation results over time. Compared to 2011, evaluations conducted in 2012, 2013, and 2014 presented a significantly higher probability of diffusion.

Table 4. Models of Conceptual Use (QS Sample)

	(4.1) Quality Index	(4.2) Items	(4.3) Index & Controls
Quality Score	0.05 (0.10)		-0.03 (0.10)
Mean Quality OU	0.03 (0.19)	0.17 (0.19)	-0.01 (0.21)
Items			
Accurate Summary		-0.51 (0.35)	
Method Fit		1.02** (0.37)	
Evaluation Specialist		0.03 (0.35)	
Social Science Methods		0.30 (0.57)	
Distinct Findings		-0.47 (0.38)	
Rec. for Specific Party		-0.05 (0.33)	
Timing (ref. = Mid-term)			
Final			-0.05 (0.31)
Ex-post			-0.73 (1.14)
Continuous			1.61 (1.46)
Region (ref. = Af/Pak)			
Africa			-0.18 (0.54)
Asia			-0.57 (0.62)
Latin America			-0.09 (0.62)
Middle East			0.11 (0.75)
Europe and Eurasia			0.47 (0.58)
Global			1.93 (1.33)
Evaluation Year			0.36* (0.16)
Constant	-1.64 (1.19)	-2.25+ (1.32)	-718.85* (326.08)
N	241	216	241
Pseudo-R ²	.002	.042	.043
Chi ²	0.55	10.68	12.04
Log-L	-138.18	-121.32	-132.44

Note: Entries are logistic regression coefficients (standard errors). + p<0.10, * p<0.05, ** p<0.01

5. Analysis of the Survey Sample (N=118)

The smaller sub-sample of evaluations includes a very large number of independent variables generated by the survey. These capture information on the evaluation process (planning, participation, dissemination, post-evaluation activities, perceptions of influence), the evaluation product (purpose, questions, team, methods, evidence, recommendations, perceptions of the evaluation and its influence), and influence pathways variables (individual opinions, individual skill acquisition, interpersonal effects on justification and role of agents, collective decisions on program change, diffusion). More importantly, the survey also provides measures of Discernible Impacts, the highest outcome level in the *Evaluation Utilization Conceptual Framework*. The issue of substantive impacts will be addressed in section 5.3.

Given the large number of variables, and the many possible permutations of these variables in alternative models, the analysis focuses on three questions that build on the findings of the previous sections: (1) Are any specific forms of conceptual use the drivers for instrumental use? (2) Are any specific features of evaluations, not captured by the more general measures of product quality discussed in the previous section, relevant for conceptual or instrumental use? (3) Are any specific post-evaluation protocols, including post-evaluation reviews and forms of dissemination, the drivers for utilization?

These questions call for the inclusion of three groups of variables in the models for the survey sample. Reliability of the results is necessarily affected by the smaller sample size—for example, any attempt to incorporate the quality index for individual evaluations to the analysis reduces the sample size to 67 cases and makes any results insignificant—but the survey data allow us to qualify some of the previous findings in important ways.

1. The first cluster of variables captures the impact of “lower” forms of utilization on “higher” forms of utilization. For example, previous models showed that Conceptual Use is a driver of Instrumental Use. Similarly, Instrumental Use should be a driver of Discernible Impacts. To assess the role of different forms of Conceptual Use, the models employ three dichotomous indicators extracted from the survey: whether the respondent learned anything about the project (eval_learn.proj), whether he or she gained any knowledge about evaluations as such (eval_elab.eval), and whether the evaluation changed the participant’s opinion of a process (eval_opin.prcs). Ancillary tests including a fourth measure of whether the respondent gained any new skills (eval_skill) showed no significant effects for this variable and did not alter the results presented below. In addition, to assess the effects of Instrumental Use on Discernible Impacts, models for this outcome employ a six-point ordinal scale based on survey responses (eval_score.utilization.a) that captures the percentage of recommendations accepted and implemented by the team.

2. The second cluster of variables describes three characteristics of the “product” (the evaluation process and the related report) that transcend the formal measures of quality employed in the previous section: whether the evaluation process involved country partners (eval_sh.cp), whether results of the evaluation were presented in a timely manner that allowed for new decisions or adjustments to the program (eval_prcp.d), and whether the evaluation presented novel findings that were previously unknown to USAID (eval_prcp.finds.b).

3. The last cluster of variables captures post-evaluation protocols following the guidelines in ADS Chapter 203, section 3.1.9 (Responding to Evaluation Findings). The first item is a

dichotomous variable reflecting the organization of “proactive” review meetings. Proactive post-evaluation meetings are those in which respondents explicitly indicated that they (1) discussed specific findings, (2) accepted or rejected recommendations, (3) assigned responsibilities, (4) established timeframes, (5) determined how to follow up on the implementation, (6) adopted specific steps to learn from the evaluation, or (7) decided to monitor a post-evaluation Action Plan. Any of those actions qualified as “proactive” by contrast to evaluations followed by no meeting or by meetings in which no activity was acknowledged.⁵⁴

Given the importance of Dissemination in Tables 1 through 4, additional variables unpack the sixth step of ADS 203.3.1.9 (“Share and openly discuss evaluation findings...”) by identifying the targets of dissemination. Three dichotomous indicators capture whether any form of dissemination (e.g., briefings prior to completion of the evaluation, sharing the final report, invitation to a dissemination event) involved an Implementing Partner (eval_dism.ip), a Country Partner (eval_dism.cp) or Beneficiaries (eval_dism.db). Internal dissemination within USAID was not considered as an explanatory factor due to lack of variance: 98% of evaluations in the sample were shared internally.

5.1. Instrumental Use

Table 5 presents estimates for five logistic regressions in which, as in Tables 1 and 3, **Instrumental Use** (eval_use) is the dependent variable. Model 5.1 employs the three dichotomous indicators of conceptual use extracted from the survey: whether the respondent learned anything about the project, whether he or she gained any knowledge about evaluations as such, and whether the evaluation changed the participant’s opinion of a process. The results suggest that specific findings about the project, more than other forms of conceptual use, drive instrumental utilization of the results.

Model 5.2 compares the effects of different product features that were not captured by the quality indicators available in Model 3.2: whether country partners participated in at least one stage of the evaluation, whether the evaluation was received in time to make relevant decisions for the program, and whether the findings produced new information not known to USAID prior to the evaluation. The novelty of the findings appears to be a key factor driving innovations in USAID programs in Model 5.2, but this effect becomes only marginally significant (at the .10 level) in Model 5.4, for reasons discussed below.

More importantly, Model 5.2 confirms the importance of Mean Quality scores as a signal of the overall ability of an Operating Unit to engage in instrumental use. Because quality scores are not available for two Operating Units in the survey sample, the number of cases drops to 116 when we include this variable, but its effects remain robust and consistent with the findings for larger samples presented in Tables 1 and 3.

⁵⁴ The dichotomous indicator for proactive review meetings proved to be the predictor producing the most consistent results for this part of the analysis. The analysis explored other operational measures for the guidelines in ADS 203.3.1.9, including: (1) a simple dummy capturing evaluations that had post-evaluation meetings (eval_per.b), (2) a scale ranging between 0 and 10 to capture how many post-evaluation activities were held (eval_per.b, eval_per.dec.a-g, eval_per.mng, eval_per.rept), (3) an ordinal scale reflecting the steps in ADS 203.3.1.9 (meeting, review of findings, accept conclusions, assign responsibilities), and (4) ten dichotomous variables for specific actions in the post-evaluation actions sub-cluster. Most of these operational measures generated weak or inconsistent results, suggesting that it is not the number of actions, but their proactive nature, what induces instrumental use.

Table 5. Models of Instrumental Use (Survey)

	(5.1)	(5.2)	(5.3)	(5.4)
	Conceptual Use	Product	Post- Evaluation	Trimmed
Conceptual Use				
Learned from evaluation	3.26** (1.26)			2.96 (2.03)
Learned about evaluations	0.13 (0.99)			
Changed opinion of process	-0.38 (0.99)			
Product				
Country participated		1.20 (0.82)		
Timely		0.07 (1.22)		
New Information		2.85** (1.10)		2.28+ (1.32)
Mean Quality OU		1.20** (0.43)		1.89* (0.73)
Post-Evaluation				
Proactive meeting			3.64 (2.30)	5.68+ (3.18)
Dissemination to Partner			2.27** (0.83)	1.92+ (1.10)
Country			0.41 (0.80)	
Beneficiaries			-0.68 (1.06)	
Constant	-0.62 (1.37)	-7.83* (3.54)	-0.04 (0.59)	-16.53** (6.38)
N	118	116	118	116
Pseudo-R ²	.10	.29	.31	.52
Chi ²	7.16	19.75	23.61	36.02
Log-L	-34.02	-24.70	-25.80	-16.56

Note: Post-stratification weighted sample. Entries are logistic regression coefficients (standard errors).

+ p<0.10, * p<0.05, ** p<0.01

In turn, Model 5.3 incorporates the dichotomous indicator for proactive post-evaluation meetings as well as the measures of dissemination to implementing partners, country partners, and beneficiaries.

Model 5.3 suggests a more nuanced interpretation for the impact of Dissemination on Instrumental Use, documented with larger samples in Tables 1 and 3. Sharing results with implementing partners is the only practice that expands utilization significantly.

To summarize these results, the last column presents a “trimmed” model that excludes all independent variables with insignificant effects in the previous equations (but retains proactive post-evaluation meetings, a variable that shows significant effects in many alternative specifications). In this model, Dissemination of results to partners, the ability of the evaluation to generate New Information, the Mean Quality scores for the Operating Unit, and the organization of Proactive Post-Evaluation meetings retain strong and significant coefficients (albeit mostly at

the .10 level). The statistical effects of lessons learned from the evaluation become insignificant in the summary model.

Figure 4 provides a substantive interpretation of these effects by plotting the predicted probabilities of Instrumental Use under different conditions, based on Model 5.4. Panel 4.1 shows that the expected rate of instrumental use is close to 83% in the absence of any dissemination involving a partner and 94% when dissemination takes place. Similarly, predicted utilization is 87% when the report does not contain novel findings and 97% when it does (Figure 4.2).

Figure 4.3 highlights again the relevance of Mean Quality at the Operating Unit level for instrumental use: at low levels of quality (the minimum observed in the sample is 4), the predicted probability of instrumental use is 47%, while at maximum levels (10), the prediction is almost a 100% rate. In turn, Figure 4.4 shows that the adoption of Proactive Post-Evaluation protocols increases the expected rate of utilization from 87% to 99%.

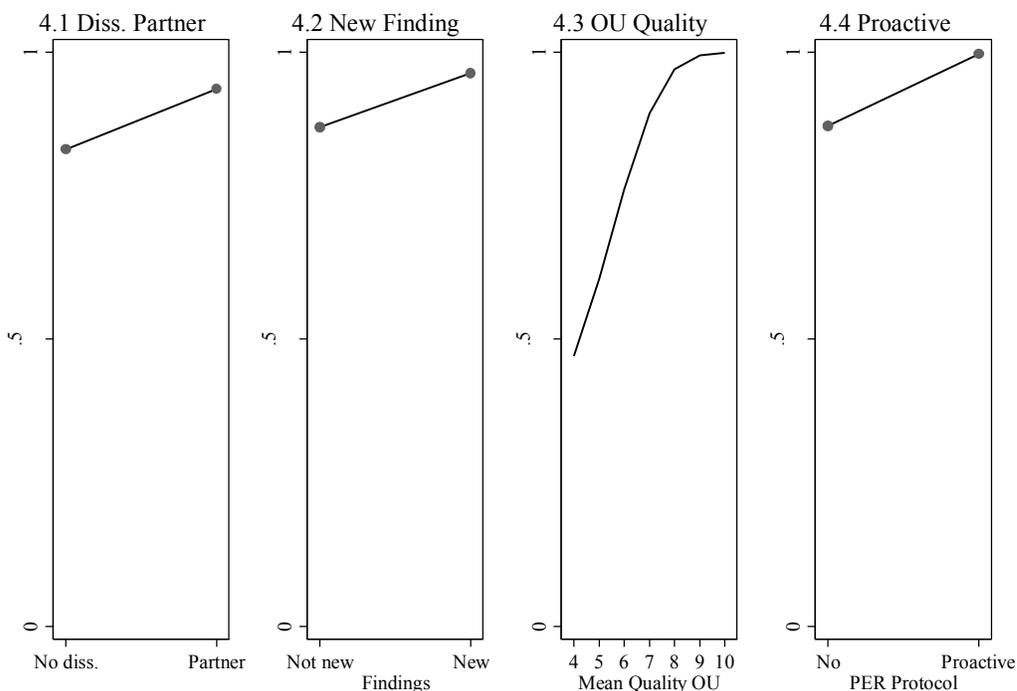


Figure 4. Predicted Probabilities for Instrumental Use (Model 5.4)

Robustness Tests for Proactive Post-Evaluation Meetings. To verify the effects of proactive post-evaluation meetings observed in Figure 4.4, Table 5-B replicates the analysis using an alternative measure of Instrumental Use. The dependent variable is a six-point scale based on the survey (eval_score.utilization.a) that captures the percentage of recommendations accepted and implemented by the team: None, Between 1 and 25%, 25-49%, 50-74%, 75-99%, and All (100%). Category “None” corresponds to situations when the dichotomous measure of

instrumental utilization (eval_use) is 0 (e.g., when no recommendation was accepted). The remaining categories of the ordinal scale “unpack” the range of situations when the dichotomous measure of instrumental use acquires a value of 1.

Table 5-B. Models of Instrumental Use (Ordinal Measure)

	(5B.1) Conceptual Use	(5B.2) Product	(5B.3) Post- Evaluation	(5B.4) Trimmed
Conceptual Use				
Learned from evaluation	2.91* (1.25)			1.45 (1.37)
Learned about evaluations	0.47 (0.50)			
Changed opinion of process	-0.58 (0.48)			
Product				
Country participated		-0.22 (0.36)		
Timely		0.09 (0.63)		
New Information		0.75* (0.36)		0.58 (0.36)
Mean Quality OU		0.30 (0.19)		0.27 (0.18)
Post-Evaluation				
Proactive meeting			0.95** (0.35)	0.81* (0.36)
Dissemination to Partner			1.21+ (0.67)	0.84 (0.66)
Country			-0.17 (0.38)	
Beneficiaries			0.34 (0.41)	
Threshold 1	0.46 (1.26)	0.18 (1.61)	-0.77 (0.63)	2.45 (2.04)
Threshold 2	1.40 (1.27)	1.17 (1.61)	0.22 (0.64)	3.53+ (2.07)
Threshold 3	1.83 (1.28)	1.61 (1.62)	0.68 (0.65)	4.01+ (2.07)
Threshold 4	2.80* (1.29)	2.61 (1.62)	1.69** (0.65)	5.05* (2.08)
Threshold 5	4.17** (1.30)	3.94* (1.64)	3.06** (0.68)	6.39** (2.10)
N	115	113	115	113
Pseudo-R ²	.021	.017	.037	.043
Chi ²	8.11	6.44	14.30	16.08
Log-L	-189.51	-185.92	-186.41	-181.10

Note: Weighted sample. Entries are ordered logit coefficients (standard errors). + p<0.1, * p<0.05, ** p<0.01

Because the dependent variable constitutes an ordinal scale, models in Table 5-B are estimated using ordered logistic regression. This approach estimates the probability that any case (evaluation) in the sample will fall into each of the six categories in the scale, given the values observed for the independent variables. The approach assumes that independent variables have a

consistent effect across categories: positive coefficients indicate that an increase in the independent variable will “push” cases into higher categories, and negative coefficients indicate that an increase in the independent variable will “drag” cases into lower categories in the scale. By definition, the probabilities of falling into the six categories estimated for any given case must add up to 1. For example, if the model determines that an evaluation has a 95% chance of falling into the “None” group, the estimated probabilities of falling into the remaining five categories must add up to only 5%.

Although the use of this ordinal scale reduces the sample size slightly (n = 115) the results in Table 5-B confirm that post-evaluation protocols, in the form of proactive post-evaluation reviews, are an important driver of instrumental use. To assess the substantive meaning of these variables, Figure 5 presents the predicted probabilities of a typical evaluation falling in each of the six categories, based on Mean Quality scores and Post-Evaluation Protocols (estimates based on Model 5B.4). The figure indicates that the adoption of Proactive Post-Evaluation Protocols consistently *increases* the probability of observing *high levels* of instrumental use (plots for 75-99% and All), and *reduces* the probability of observing *low levels* of instrumental use (in the plots for None and Less than 25%).

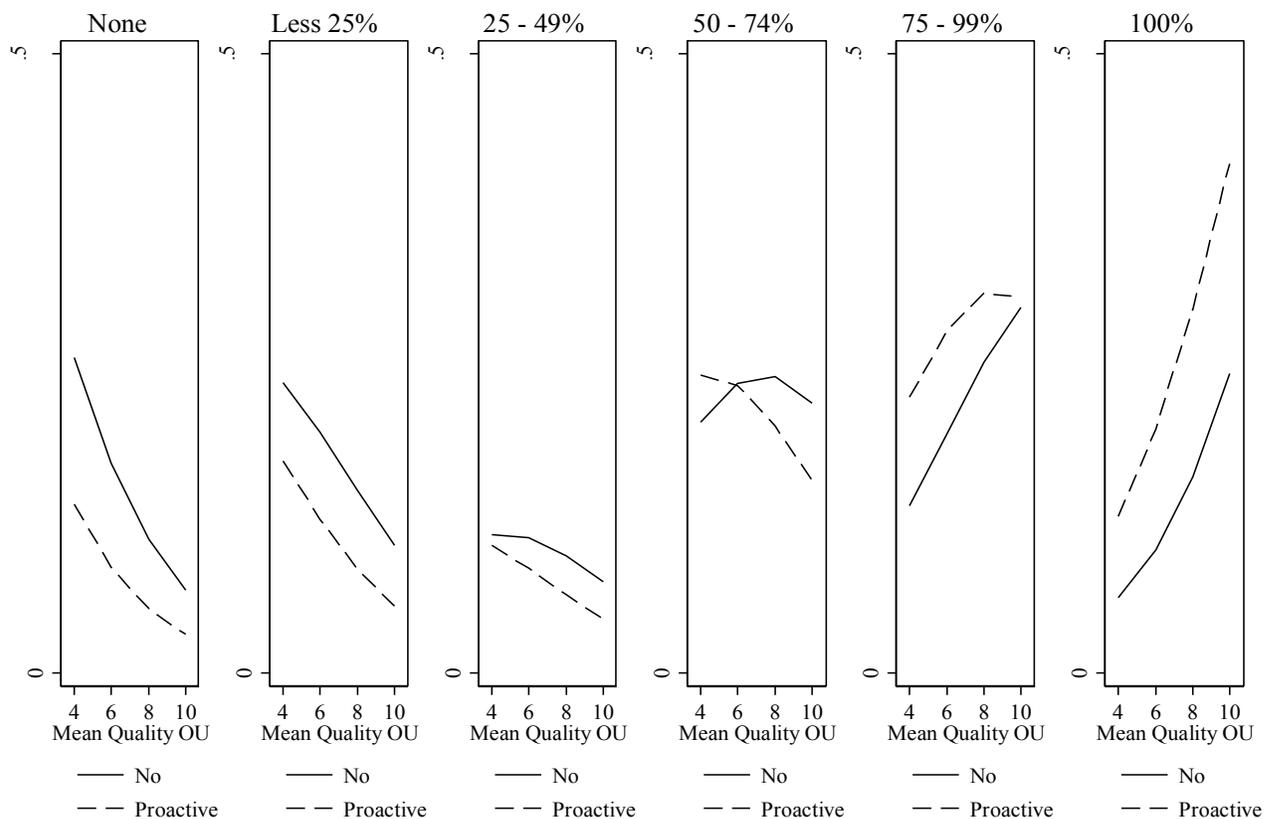


Figure 5. Predicted Probabilities for Different Levels of Instrumental Use (Model 5B.4): By Dissemination to Partners and Proactive Post-Evaluation Review

Patterns of Dissemination. Table 5 (Model 5.3) suggests that dissemination to implementing partners may be relevant to promote instrumental use. This finding qualifies in important ways the general effects of dissemination documented with larger samples in Tables 1 and 3. Efforts to probe further on this issue by asking whether some forms of dissemination to implementing partners are more effective than others were unsuccessful because the data do not allow for a definite answer to this question. The survey captured specific patterns of dissemination, including: (1) briefings to partners and beneficiaries prior to the completion of the evaluation report, (2) sharing the final report or executive summary, and (3) inviting partners and beneficiaries to dissemination events or providing them with materials on the evaluation. The survey also discriminated whether such activities were targeted at implementing partners, country partners, or beneficiaries. The analysis of these specific variables produced weak and inconsistent results, preventing any clear conclusions about more effective patterns of dissemination.

5.2. Conceptual Use

In line with the strategy pursued in previous sections, Table 6 moves to the analysis of **Conceptual Use** (*eval_use.f*). Narrower measures of conceptual use included as predictors in Table 5 are excluded from this analysis because they overlap with the dependent variable. For example, question 27 of the survey was used to code learning about the project (*eval_learn.proj*) which, combined with information from additional sources, was in turn employed to code the broader indicator of conceptual use that serves as the dependent variable in Table 6.

The results in Table 6 indicate that some features of the evaluation not captured by the general quality measures discussed in previous sections are crucial to promote conceptual use. Involvement of country partners in the evaluation process (*eval_sh.cp*) and the timeliness of the evaluation have positive and significant effects on the dependent variable (Model 6.1). Moreover, both items retain their levels of statistical significance in the trimmed model (6.4). The Mean Quality of the reports for the Operating Unit again emerges as an important predictor. Thus, a timely evaluation report, the involvement of local partners, and a high-quality environment seem to facilitate useful learning based on evaluation results.

The variable capturing proactive post-evaluation meetings has a positive but insignificant effect in Model 6.2. While proactive meetings consistently translate into more instrumental utilization, their effect on conceptual use is much more uncertain. Seeking to clarify the results in Tables 2 and 4, which showed that Dissemination is the main driver of Conceptual Use, Model 6.2 also tests the effects of different types of dissemination on the dependent variable. Surprisingly, no specific form of propagation presents a significant impact on the likelihood of conceptual utilization. The limited size of the sample may partly account for the null findings.

Table 6. Models of Conceptual Use (Survey)

	(6.1) Product	(6.2) Post- Evaluation	(6.3) Trimmed
Product			
Country participated	3.06* (1.48)		3.21* (1.45)
Timely	4.42** (1.43)		4.15** (1.38)
New Information	-1.15 (1.35)		
Mean Quality OU	1.27+ (0.71)		1.43* (0.68)
Post-Evaluation			
Proactive meeting		0.69 (0.86)	
Dissemination to:			
Partner		-0.21 (1.10)	
Country		0.82 (0.91)	
Beneficiaries		0.49 (1.20)	
Constant	-9.64+ (5.59)	2.18* (0.96)	-11.37* (5.29)
N	116	118	116
Pseudo-R ²	.46	.05	.45
Chi ²	21.55	2.49	20.72
Log-L	-12.44	-25.58	-12.85

Note: Weighted sample. Entries are logistic regression coefficients (standard errors). + p<0.1, * p<0.05, ** p<0.01

Figure 6 (based on Model 6.4) illustrates how evaluation features combine to promote or hinder conceptual use. An evaluation process not involving a country partner (in the left panel) has a predicted rate of conceptual use close to 88%, and this rate increases to 99% once a local partner becomes involved. Similarly, an untimely report (in the middle panel) has an expected probability of conceptual use close to 70%, while a timely report raises this probability to 98%. These values are above 70% in all cases because positive features may substitute for each other, offsetting the absence of an important characteristic. In turn, an evaluation produced in a low-quality environment presents an expected rate of conceptual use of 57%, while an evaluation produced in a high-quality environment is almost guaranteed to inspire conceptual use.

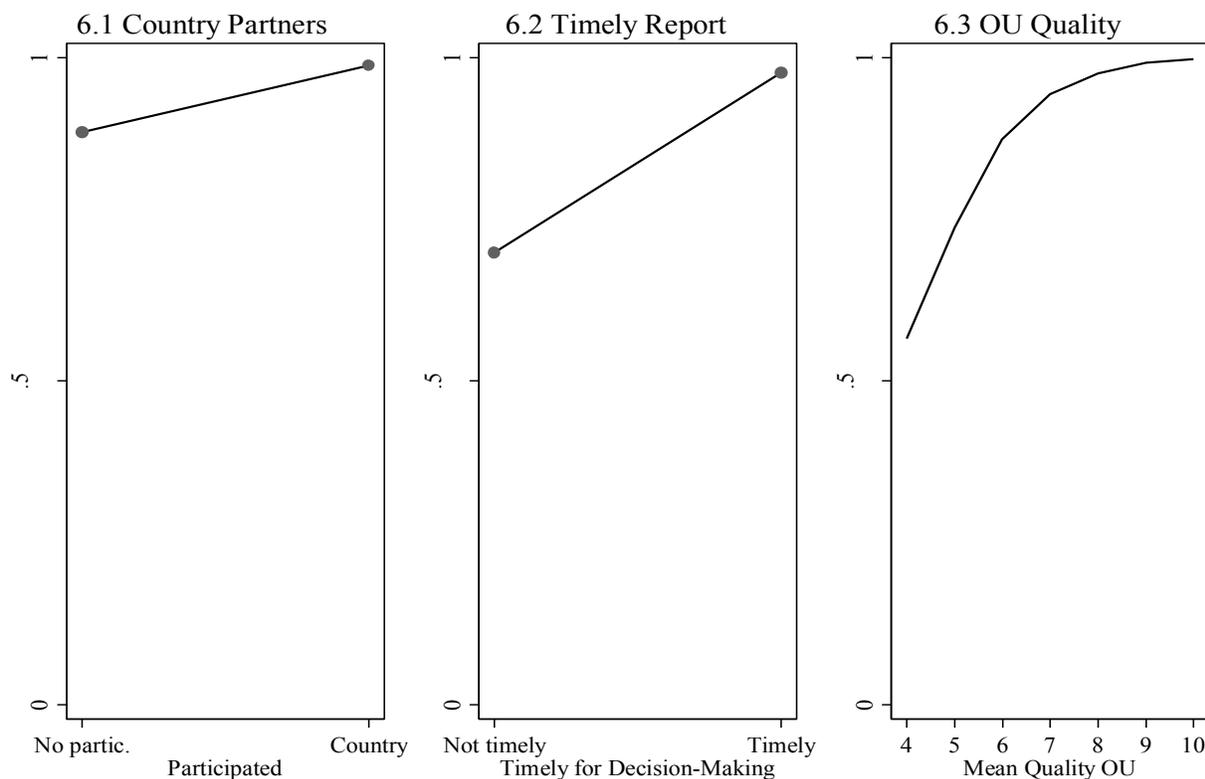


Figure 6. Predicted Probabilities of Conceptual Use (Model 6.3)

5.3. Discernible Development Impacts

The highest outcome level in the *Evaluation Utilization Conceptual Framework* corresponds to the presence of **Discernible Impacts**. Because there is no measure of discernible impacts available for the larger sample of evaluations, this outcome was not analyzed in previous sections. Based on survey responses (Q47), a dichotomous indicator (score_impact.a) captures when respondents claimed that development improvements occurred as a result of the evaluation process. The survey shows that—not surprisingly—discernible impacts are much harder to attain than conceptual or instrumental use: while 97% of respondents learned about their projects from evaluations (Q27) and 91% acknowledged some form of instrumental use (Q43), only 42% were ready to claim that better development outcomes occurred as a result of an evaluation (Q47).

The capacity of evaluations to produce substantive impacts is presumably influenced by their potential for Instrumental Use, the levels of Conceptual Use, and by additional background drivers such as processes of dissemination.

Table 7. Models of Discernible Impacts (Survey)

	(7.1) Forms of Use	(7.2) Product	(7.3) Post- Evaluation	(7.4) Trimmed
Utilization				
Instrumental (ordinal)	0.34* (0.14)			0.27+ (0.14)
Learned from evaluation	-0.26 (1.33)			
Learned about evaluations	0.51 (0.60)			
Changed opinion of process	0.61 (0.56)			
Product				
Country participated		0.42 (0.41)		
Timely		-0.03 (0.67)		
New Information		0.60 (0.39)		
Mean Quality OU		0.02 (0.19)		
Post-Evaluation				
Proactive meeting			-0.31 (0.41)	
Dissemination to:				
Partner			2.46* (1.01)	1.56 (1.00)
Country			1.24** (0.46)	0.91* (0.42)
Beneficiaries			-0.70 (0.49)	
Constant	-2.01 (1.40)	-1.03 (1.67)	-2.93** (0.99)	-3.11** (1.03)
N	115	116	118	115
Pseudo-R ²	.06	.02	.11	.11
Chi ²	9.89	3.21	17.94	16.80
Log-L	-73.39	-77.17	-71.15	-69.94

Note: Weighted sample. Entries are logistic regression coefficients (standard errors). + p<0.1, * p<0.05, ** p<0.01

Table 7 presents four logistic models for Discernible Impacts. Equation 7.1 includes four independent variables. The first one measures prior instrumental use through a six-point scale; the other three variables are dichotomous items that capture conceptual use: learning about the specific project, the evaluation process, or a change in opinions. The results of this model indicate that instrumental utilization (the more proximate cause of program outcomes), rather than just learning, contributes to substantive development impacts.

The measure for instrumental use employed in 7.1 is the ordinal, six-point scale used as dependent variable in Table 5-B (eval_score.utilization.a). This scale captures the percentage of recommendations implemented (None, 1-25%, 25-49%, 50-74%, 75-99%, and 100%). The primary measure of instrumental use (eval_use) employed as dependent variable in previous sections cannot be included as a predictor in Model 7.1 because *all* evaluations with discernible

impacts are also instances of instrumental use according to this measure. This alignment of the data suggests that instrumental use is possibly a necessary condition to achieve discernible impacts. Unfortunately the absence of any “deviant” evaluations with reported impacts but without instrumental use prevents the inclusion of this indicator in the estimation (technically, it produces perfect separation of the data). Thus the ordinal measure is employed in the analysis. The fact that the ordinal measure, which reduces sample size to 115, retains significance at the .05 level suggests that the dichotomous measure would have a significant effect if it could be employed.

Model 7.2 suggests that no features of the evaluation product (timeliness, participation of local partners, or the novelty of the findings) are consistently related to discernible impacts. This is not surprising if we consider that these features operate early in the causal chain: they facilitate conceptual use, which promotes instrumental use, which in turn may lead to development outcomes. In turn, Models 7.3 and 7.4 show that post-evaluation activities have diverse effects. Proactive reviews may influence instrumental use, but they have no direct influence on development outcomes. By contrast, sharing the findings of an evaluation with implementing partners and (especially) with country partners constitute important ways to promote discernible outcomes.⁵⁵

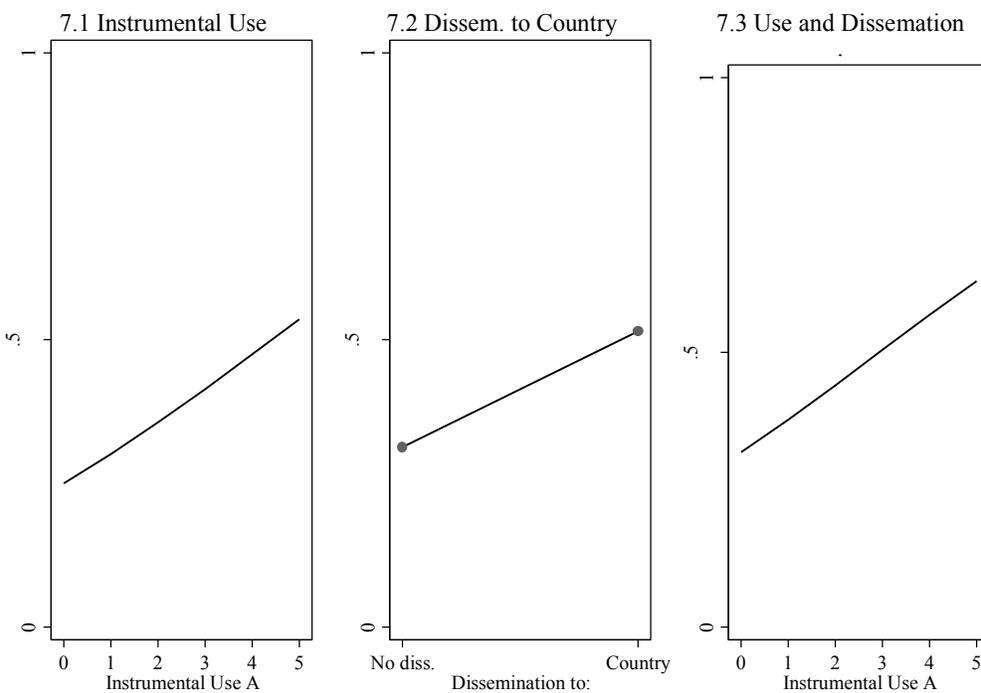


Figure 7. Predicted Probability of Discernible Impacts (Model 7.4)

⁵⁵ In order to pursue a further analysis of discernible development outcomes, a set of QCA models was tested for the smaller subsample of evaluations that combine survey data and quality scores. These QCA models were unable to identify any alternative “paths” or configurations of causal conditions able to guarantee the production of discernible developing outcomes. The QCA analysis, however, suggested that instrumental use (eval_use) and overall dissemination (eval_dism.both) are both necessary conditions for the translation of evaluation lessons into actual developing outcomes. The QCA output is not reported here due to the limited usefulness of the results, but this information is available on request.

The trimmed model (7.4) confirms that the main mechanism promoting discernible development outcomes is an active pattern of instrumental use reinforced by dissemination of results among local country partners. However, the explanatory capacity of the model (reflected by the Pseudo-R) is low when compared to summary models for previous outcomes. This indicates that development outcomes are also driven by powerful external factors unrelated to the evaluation process, not captured by the statistical model.

Figure 7 plots the predicted probabilities of discernible outcomes under the three different conditions (based on Model 7.4). If we assume that cases in the sample retain the other two variables at their observed values, the adoption of any favorable condition (instrumental use or dissemination to country) by itself is unlikely to push the expected probability of discernible outcomes above 54%. A hypothetical sample combining the simultaneous benefits of instrumental use and dissemination to country partners (depicted in Figure 7.3) would increase the expected rate of discernible outcomes to 63%.

5.4. Summary Lessons and Conclusions

These results should not be construed as evidence of the ineffectiveness of evaluations. On the contrary, statistical analyses presented in previous tables have shown that officers participating in the survey display an ability to process (i.e., make conceptual use of) recommendations when reports are presented in a timely manner and incorporate the knowledge of local country partners; and that the same officers are able to transform these lessons into actionable plans by adopting proactive post-evaluation protocols, disseminating results to implementing partners, and producing new information. Regular access to quality evaluations (more than the quality of any particular study) is crucial to facilitate conceptual and instrumental use within Operating Units. The translation of those plans into discernible development outcomes is facilitated by effective instrumental use and by further dissemination of the results to country partners, but it is also mediated by external conditions beyond the possibility of direct control. Therefore, it should not be surprising that observed rates of success are lower when we assess utilization outcomes at the highest level.

Table 8 compares the predictors with significant effects in the summary models for Conceptual Use, Instrumental Use, and Discernible Impacts (Tables 5 through 7). The independent variables are re-arranged in three blocks: (1) factors over which USAID officials have direct control (incorporating country partners, organizing proactive post-evaluation meetings, dissemination to country partners), (2) factors over which officials have limited or only indirect control, via planning and SOW design (the timeliness of reports, the novelty of findings, and the overall quality environment), and (3) variables that represent the causal sequence of utilization (for example, instrumental use promotes discernible impacts).

Four factors under the direct control of Operating Units appear distinctively relevant:

- Involving country partners in the evaluation process facilitates Conceptual Use;
- Sharing the results with implementing partners and Proactive post-evaluation meetings promote Instrumental Use, and
- Dissemination to country partners is key to facilitate Discernible Impacts on development.

Table 8. Main Drivers of Evaluation Utilization

	Conceptual Utilization	Instrumental Utilization	Discernible Impacts
Direct control	Involvement of country partners in the evaluation.	Dissemination to implementing partners. Proactive post-evaluation meetings.	Dissemination to country partners.
Indirect control	Timeliness of the evaluation. Mean quality scores for Operating Unit.	Evaluation generates new information. Mean quality scores for Operating <u>Unit</u> .	
Indirect effects			Instrumental utilization.

In addition, three other factors under the indirect control of Operating Units facilitate evaluation use:

- Timely Evaluations and the Mean Quality environment for the Operating Unit are important conditions to facilitate Conceptual Use;
- Novel Findings and the Quality environment for the Operating Unit emerge as the main factors promoting Instrumental Use.

It is important to note that some of these variables also have indirect effects on higher-level outcomes. For example, Proactive Meetings encourage Instrumental Use, and Instrumental Use in turn promotes Discernible development outcomes.

ANNEX 5 – STUDY INSTRUMENTS

- 5.1 Policy Document Instrument
- 5.2 CDCS Instrument
- 5.3 PPR Instrument
- 5.4 Mission Order Instrument
- 5.5 Purposes, Questions, and Recommendations Instrument
- 5.6 Interview Guide for Operating Unit Group Discussions
- 5.7 Evaluation Report Quality Review Checklist

5.1 – Policy Document Review Template

Policy Document Evaluation Utilization Analysis

The purpose of this exercise is to identify instances where USAID used evidence from USAID evaluations in the development of Policy Documents.

Instructions: Perform a keyword search in the document (ctrl + f) for the term “evaluation” to see if any evaluations are directly cited as evidence to justify policy decisions or the theory of change. Then, skim foot notes, in-text citations, and bibliographies for the different types of evidence used.

Question	Response	
Policy Document Title		
Year Published		
1. What types of sources did the Policy document claim to have used as evidence?	Answer: Check all that apply	
	Claimed evidence from a USAID evaluation	<input type="checkbox"/>
	Claimed evidence from other USAID documents or experience	<input type="checkbox"/>
	Claimed evidence from USAID Evidence Summits	<input type="checkbox"/>
	Claimed evidence from non-USAID evaluations	<input type="checkbox"/>
	Claimed evidence from published research	<input type="checkbox"/>
	Claimed evidence from other non-USAID documents	<input type="checkbox"/>

CONTINUE TO QUESTION 2 ONLY IF POLICY DOCUMENT CITES EVALUATIONS OR IMPLIES THAT ACTIVITY/PROJECT/PROGRAM RESULTS PROVIDED EVIDENCE FOR POLICY DECISIONS.

2. If the Policy document cites evaluations, enter the quotes and page numbers in the boxes to the right. Add additional rows as needed.	Answer: Enter All Quotes Referencing Evaluations or Implying Possible Use of Evaluations				
	#	Quotes		Page #s	
	1				
	2				
	3				
	4				
3. If the Policy document cites evaluations, please identify if it was a USAID evaluation. Please provide as much information as possible for each evaluation. Use the same numbering from evaluations as in #2 above.	Answer: Enter all evaluations referenced				
	#	USAID Evaluation (Y/N)	Evaluation Title or Description	Sponsoring Operating Unit or Organization	URL, DEC ID, or other location
	1				
	2				
	3				
	4				
4. If the Policy document cites USAID evaluations, enter the section title each quote was found in. Use the same numbering from evaluations as in #2 above.	Answer: Enter the Section Title for Each Quote				
	1				
	2				
	3				
	4				
	5				

5.2 – CDCS Review Template

CDCS Evaluation Utilization Analysis

The purpose of this exercise is to identify instances where Missions used evidence from USAID evaluations or other sources in the development of their CDCS or RDCS.

Instructions: Perform a keyword search in the document (ctrl + f) for the terms “evidence” and “evaluation” to see if any examples of use of evidence are identified. Then, look for in-text citations and footnotes in the report for additional use of evidence.

Question	Response	
Operating Unit Name		
Year Published		
1. Did the CDCS claim to have used evidence from USAID evaluations or other sources when developing the CDCS? If so, what types of sources were claimed?	Answer: Check all that apply	
	No claim that evidence was used	
	Claimed use of evidence, but source not provided	
	Claimed evidence from a USAID evaluation	
	Claimed evidence from other USAID documents	
	Claimed evidence from non-USAID evaluations	
	Claimed evidence from published research	
	Claimed evidence from other non-USAID documents	

CONTINUE TO QUESTION 2 ONLY IF THE CDCS SPECIFICALLY CITES AN EVALUATION (USAID OR NOT). IF NO SPECIFIC CITATION OR REFERENCE, CONTINUE TO QUESTION 8.

2. If the CDCS references evidence from evaluations, please copy and paste the surrounding quote into the space provided, including page numbers. Add additional rows as needed.	Answer: Enter All Quotes Referencing Evaluations				
	#	Quotes		Page #s	
	1				
	2				
	3				
	4				
3. Count the number of quotes referencing evaluations and the number of quotes which specifically reference USAID evaluations from Question 2. Enter the totals in the cells on the right.	Answer:				
				Total	
	Total Number of Quotes Referencing Evaluations				
	Total Number of Quotes Referencing USAID Evaluations				
4. If the CDCS specifically cited evidence from evaluations, please identify if it was a USAID evaluation. Please provide as much information as possible for each evaluation. Use the same numbering from evaluations as in #2 above.	Answer: Enter all evaluations referenced				
	#	USAID Evaluation (Y/N)	Evaluation Title or Description	Sponsoring Operating Unit or Organization	URL, DEC ID, or location
	1				
	2				
	3				
	4				
5. Count the number of distinct evaluations and USAID	Answer:				
				Total	

Question	Response			
evaluations specifically cited for evidence in the document from Question 4. Enter the totals in the cells on the right.	Total Number of Distinct Evaluations Cited			
	Total Number of Distinct USAID Evaluations Cited			
6. If the CDCS cites any evaluations, please indicate which section the reference was found in by checking the appropriate box. If "Other" is checked, also provide the name of the section. Use the same numbering from evaluations as in #2 above.	Answer:			
	#	Development Hypothesis	Results Framework	Other
	1			
	2			
	3			
	4			
7. If, in the process of presenting evidence from evaluations, the CDCS discussed the data quality behind the evidence, please provide language in the space provided. Data quality may be positive or negative, such as "they were from a rigorous evaluation," "came from a representative sample survey," or "weren't corroborated by a second source."	Answer: Use Evaluation #s from #2 above; Add rows as needed			
	#	Data Quality Issues	Page #s	
	1			
	2			
	3			
	4			
5				

THE FOLLOWING SECTION IS FOR ALL CDCS DOCUMENTS

8. If, in the section of the CDCS focused on Learning (described in ADS 201.3.3.5), and specific references to evaluation use or dissemination are included, please summarize or provide those specific references, including page numbers, in the space provided.	Answer:	
	Use or Dissemination Language	Page #s
9. In the References section of the CDCS, look for any additional evaluations that were not explicitly cited in the body of the document. Please paste the full references in the cell to the right and note whether or not each is USAID evaluation. Add additional rows as needed.	Answer: Enter All Evaluation References	
	Reference	USAID Evaluation (Y/N)
10. Count the total number of evaluations and USAID Evaluations listed in the References section of the CDCS. Enter the totals in the cells to the right.	Answer:	
		Total
	Total Number of Evaluations in References	
	Total Number of USAID Evaluations in References	

Text Box I. ADS 201 Guidance on Evaluation Utilization for a CDCS

201.3.3.3 Results Framework

- Is the DO based on a clear development hypothesis and strong evidence, including **evaluations** conducted by the Mission?

201.3.4.1 Phase I – Initial Consultations

There are two deliverables under Phase I. The first is a digital video conference (DVC). The second is a Consultation Note.

(1) The DVC ... includes the following key elements:

- A summary of the analyses, assessments, **evaluations**, and other evidence that will be used to inform the strategy process, including those that need to be initiated or completed;

(2) **Consultation Note**

Analysis: A CDCS must be grounded in evidence and analysis. During the Initial Consultations Phase, the Mission determines what research, assessments, and **evaluations** are needed to inform the CDCS process... Missions are encouraged to draw evidence from third-party assessments and/or **evaluations**, to complement Mission assessments, including from government sources, civil society, the private sector, and other donors.

201.3.4.2 Phase 2 – Results Framework Development

This phase includes...

(1) **Conduct Analyses:** The Mission must review, analyze, and draw evidence-based conclusions from assessments and **evaluations** to produce the RF and full CDCS, including an analysis of what has worked or not worked in achieving results through past programs, projects, and activities..... The completed assessments and **evaluations** provide the evidence and information needed to establish a development hypothesis that describes the causal linkages between the CDCS Goal, DOs, IRs, and sub-I....The Mission must reference the assessments and **evaluations** used to reach significant conclusions in the CDCS.

201.3.4.4 Phase 3 - Full CDCS and Abbreviated CDCS Preparation, Review, and Approval

This phase includes ...

(1) **Finalize Analysis and Consultations:** The Mission completes ongoing assessments, **evaluations**, and discussions with local stakeholders, the State Department and other USG agencies, other donors, and partners to inform the full CDCS.

5.3 – PPR Entry Review Template

PPR Evaluation Use Analysis

The purpose of this exercise is to identify examples of evaluation utilization that were described in the PPR Evaluation Registry.

Instructions: In the PPR Evaluation Registry, look first at the “Report Status” column. Only evaluations that are indicated as “Completed” should be considered for this exercise. For each completed evaluation, read the content of the “Evaluation Use.” If the Evaluation Use column describes evaluation utilization that already occurred, paste the use statement into the table below, followed by the other identifying information in the rows below it. Not all Evaluation Use statements describe utilization, and some statements of utilization discuss only future plans for utilization.

Evaluation Registry Column	Paste the information given in the PPR
Evaluation 1	
Evaluation Use	
Operating Unit	
Evaluation Title	
Link To DEC/EMS #1	
Link To DEC/EMS #2	
Evaluation 2	
Evaluation Use	
Operating Unit	
Evaluation Title	
Link To DEC/EMS #1	
Link To DEC/EMS #2	

Add additional rows as necessary.

5.4 – Evaluation Mission Order Review Template

Evaluation Mission Order Changes Analysis

The purpose of this exercise is to identify instances where Missions made changes to the standard Evaluation Mission Order template that could affect Evaluation utilization at their Mission.

The Standardized Mission Order can be found at:

<https://usaidelearninglab.org/sites/default/files/resource/files/Exercise%2018-1%20Adapting%20the%20Standardized%20Mission%20Order%20on%20Evaluation.pdf>

Question	Response	
Operating Unit Name		
Year Published		
Is the Mission Order based on the standard Evaluation Mission Order template?	Answer: Check the appropriate response	
	Yes	
	No	
If the Mission Order was not based on the standardized template, STOP. It cannot be analyzed using these methods. Otherwise, continue to the next section.		

Compare the Mission Order to the standardized template. Search the sections listed below for any significant changes from the standardized template that could affect evaluation utilization at the Mission.

Definitions:

Addition – Added content only

Deletion – Deleted content only

Both – Deletions and additions occurred but in different sub-topics

Modification – A single topic had additions and deletions

None – No substantial changes affecting utilization were made

Check the type of changes, if any, that were made in the sections listed below.					
Section	Additions	Deletions	Both	Modifications	None
Definitions					
Roles & Responsibilities					
Mission Evaluation Plan					
Determination of Evaluations to Include					
Evaluation Plan Maintenance					
Budgeting					
Planning Individual Evaluations					

Section	Additions	Deletions	Both	Modifications	None
Implementation					
Reports					
Learning & Reporting					
Responding to Findings					
Sharing with Stakeholders					
Appendices					

For each instance of a significant change, paste the changed quote into the table below. Identify the section that the quote was taken from.

For additions and modifications, BOLD the changed language within the quote. For deletions, describe what was deleted. In cases where an image, chart, or annex was added, describe the addition instead of pasting it in.

Add additional rows as necessary.

Section	Quote

5.5 – Evaluation Purposes, Questions, and Recommendations Review Template

Purpose, Questions, and Recommendations Analysis

The purpose of this exercise is to perform an in-depth analysis of a sample of evaluations to determine what types of purpose statements and recommendations are found in evaluations. In addition, how many questions and recommendations do evaluations typically have? This analysis also seeks to determine the extent to which evaluation purposes, questions, and recommendations adhere to USAID guidance.

Instructions: Locate the purpose statement, evaluation questions, and recommendations in the evaluation report and answer the questions below. If the purpose, questions, or recommendations listed in the executive summary differ from those found in the report body or the attached statement of work, refer to the report body or statement of work.

Question	Response
Evaluation Report Title	
Sponsoring Operating Unit	
Report URL	

PURPOSE STATEMENT

1. What is the purpose of the evaluation? Paste the purpose statement into the space at the right. Indicate the page number on which it was found.	Paste or write in the purpose statement	Page #

EVALUATION QUESTIONS

2. In what section of the report or annexes were the evaluation questions found? Indicate the page number(s) where the questions were found.	Answer: Section Title	Page #
3. How many evaluation questions are there? Count the number of questions based upon the numbering or bullets given by the report, ignoring sub-questions (i.e., 2a. and 2b. count as only one question combined). Then, count the total number of question marks within the evaluation questions.	Question	Answer
	How many evaluation questions are there, based upon the numbering in the report?	
	How many question marks are there total within the evaluation questions?	
	Is the number of questions and the number of question marks the same or different?	

RECOMMENDATIONS

4. How many recommendations	Question	Number
-----------------------------	----------	--------

Question	Response	
<p>are there in the report? Count formal recommendations based on the numbering or bullets given in the recommendation section(s) of the report. Then, count informal recommendations or “should” statements found in other sections of the report.</p>	<p>How many formal recommendations are there in the recommendation section(s) of the report? Ignore sub-recommendations (i.e., 2a. and 2b. count as only one recommendation combined). If a recommendations section exists, but recommendations are included in narrative and are not countable, write “uncountable.”</p>	
	<p>How many informal recommendations are there outside of the recommendations section of the report? Perform a word search (ctl + f) on “recommend” and “should” in the other sections of the report. Count the number of these informal recommendations and indicate the total at the left.</p>	
<p>5. If the evaluation report includes formal recommendations, paste the first five formal recommendations found in the recommendations section(s) of the evaluation report.</p>	<p>Enter the first five formal recommendations</p>	
	#	Recommendation
	1	
	2	
	3	
	4	
5		

5.6 – Interview Guide for Operating Unit Group Discussions

This instrument served as the basis for all interviews with Operating Units. Instruments were customized slightly depending on the group being interviewed (i.e., Mid-level, Sr. Level, Regional Mission, etc.).

Interviewers please start as promptly as possible and try to finish the introductions not later than 5-7 minutes after the hour. Note that you are expected to cover all seven main questions. Moderate your use of sub-questions and prompts within an allowance of 6-7 minutes per main question. Keep the amount of time per question fairly equal, except perhaps for Questions 6 and 7 which may go faster.

[Introduction]

Thank you for joining us today for this interview. This interview, and others like it, is being carried out with USAID staff as part of an Agency-wide review of how and to what extent evaluations are being utilized as to better understand the effectiveness and impact of ongoing and completed USAID projects and activities as well as Agency strategies and policies. Other aspects of this study include a survey to follow-up on a sample of evaluations that have been completed since USAID issued its Evaluation Policy in 2011 and a number of desk reviews, including reviews of the degree to which Mission CDCSs and Agency policies reflect evaluation findings.

We would like to begin with brief introductions – our and your names and offices. [Interviewers please introduce yourselves and other on the study team.]

We have about half a dozen questions to cover in this interview. We will be allocating about 5 minutes to each question, leaving some time in between questions and at the end for you to share with us ideas about utilization that we may not have asked about. We are going to start with what happens to evaluations once you receive them and work backwards to what you do or might do early in an evaluation to foster the utilization of evaluations. It is our understanding that you are aware that this interview is being taped, that you know your names will not be cited in our report, and that you have signed the consent form we shared with your Program Office. Is that correct? So let's get started:

1. What happens in this Mission/Office/Bureau when evaluations are completed – what are your evaluation follow-up processes and experiences?
 - Are there standardized post-evaluation processes you all use or are follow-up steps different for different offices or unique to each evaluation? [Interviewers, please be sure that this discussion covers all the post evaluation steps listed in ADS 203.3.2.9, items 2-5]
 - Are the post-evaluation processes your Mission/Office/Bureau uses relatively new – perhaps started after the 2011 Evaluation Policy came out – or have you been using them for a much longer time?
 - How important are individuals – “people who are champions for doing something about evaluation findings” – for ensuring that your Operating Unit learns from and acts on an evaluation?
2. Whether, how and to what degree have your Operating Unit's evaluations reached and affect USAID's partners in this country/region/technical field? By this we mean implementing partners, but also country government partners, stakeholders in civil society organizations or other donors that are working on the same issues. [Interviewers be aware of ADS 203.3.9, item 6]
 - Which types of partners are most/least likely to pay serious attention to USAID evaluations, and why?

- Are written evaluation dissemination plans prepared and implemented for evaluations this unit conducts? Are they something new to this Mission/Office/Bureau – or are they a long standing practice?
 - Have any of you every prepared and used an evaluation dissemination plan? How did that work out?
3. In addition to improving current projects and direct follow-ons to existing activities, USAID (in ADS 201) expects that earlier evaluations (undertaken by this Operating Unit or elsewhere) will be reviewed and cited in relation to proposed development hypotheses when a new project PAD is developed or a DO strategy for a new CDCS is prepared. Have you been involved in efforts to use evaluations to frame or defend development hypotheses on which new programming depend? [Interviewers be aware of ADS 201.3.3.2.and 3.3.3 as well as ADS.3.15.2 and other sections]
- Are there any particular procedures your Mission/Office/Bureau uses to locate relevant evaluations or facilitate this type of use of evaluations?
 - When you cite evaluations in PADS or CDCSs, do you write only about what they said, or also do you also write about the strength of the evidence they provide?
4. Whether and how have evaluations this Operating Unit has carried out “made a difference” at the level of the development outcomes? This is a “so what” question. We know from the survey responses and other interviews that lots of evaluations lead to improved activity work plans and are factored into new project or activity designs – but what does that mean in terms of development effectiveness – can you see specific effects on education, or health, or food security outcomes in target areas or on national policies that trace back to what you have learned from evaluations?
- Prompt, if needed: One of the big stories we have heard about evaluation results has to do with the wide adoption by developing countries of deworming programs for school age children and education benefits that brings. In PPL, we were also told an old story about a road evaluation that drew such clear lessons about operations and maintenance that it pretty much changed the way all USAID roads projects are structured worldwide. Are there any stories that are known around your Mission/Office/Bureau about important results or changes that have emerged from evaluations?

ASK FOR THE NAMES OF ANY EVALUATIONS THAT INTERVIEWEES DESCRIBE AS HAVING IMPORTANT DEVELOPMENT EFFECTS.

5. How have steps you have taken when evaluations are first being planned, or when a SOW is being prepared, or during the course of an evaluation, affected the degree to which findings are utilized?
- Has utilization been greater when intended USAID and non-USAID evaluation users actively participate in the evaluation planning/design process or serve in some way as evaluation team members?
 - Are there other pre-evaluation steps you can think of that have resulted in some evaluations being better or more widely utilized than others?
 - What impediments, if any, make it difficult to implement recommendations from evaluations or otherwise utilize evaluation results?
6. Are there any tools or techniques this Mission/Office/Bureau uses to facilitate evaluation utilization that you think are particularly effective and would help others?
7. Are there any types of tools or aides or other type of support for fostering evaluation utilization that you would like to have but do not currently have?

Stop asking questions 5 to 7 minutes before the hour is up and ask if there is anything else the group would like to tell us about evaluation utilization. Allow up to five minutes to discuss new topics.

Stop at 1 minute before the end of the hour and provide our study contact email in case they have more they would like to tell us: Jenna at [e-mail]. Remember to say thank you before signing off.

5.7 – Evaluation Report Quality Review Checklist, Rater’s Guide and Overall Evaluation Quality Review Score

A. Evaluation Report Review Checklist⁵⁶

Keyed to USAID ADS 203.3.1.8 (11/2/12) and USAID How-To Note Preparing Evaluation Reports

(Includes a guide for comparing ratings to USAID 2009-2012 meta-evaluation averages)

Evaluation Report Review Checklist		Yes	No	N/A ⁵⁷
Executive Summary				
1.	Does the Executive Summary accurately reflect the most critical elements of the report?			
Program/Project Background				
2.	Are the basic characteristics of the program, project or activity described (title, dates, funding organization, budget, implementing organization, location/map, target group, contextual information)?			
3.	Is the program or project’s “theory of change” described (intended results (in particular the project purpose); development hypotheses; assumptions)			
Evaluation Purpose				
4.	Does the evaluation purpose identify the management reason(s) for undertaking the evaluation?			
Evaluation Questions				
How many evaluation questions does the evaluation report state that the evaluation addressed (in the body of the report, not the SOW)? ⁵⁸ Count the number of visible question marks.		Enter a number below		
5.	Are the evaluation questions stated in the body of the report clearly related to the evaluation purpose?			
6.	Are the evaluation questions in the report identical to the evaluation questions in the evaluation SOW?			
7.	If the questions in the body of the report and those found in the SOW differ, does the report (or annexes) state that there was written approval for changes in the evaluation questions?			
Methodology				
8.	Does the report (or methods annex) describe <u>specific</u> data collection methods the team used?			
9.	Are the data collection methods presented (in the report or methods annex) in a manner that makes it clear which specific methods are used to address <u>each</u> evaluation question? (e.g., matrix of questions by methods)			
10.	Does the report (or methods annex) describe <u>specific</u> data analysis methods the team used? (frequency distributions; cross-tabulations; correlation; reanalysis of secondary data)			
11.	Are the data analysis methods presented (in the report or methods annex) in a manner that makes it clear how they are associated with the evaluation questions or specific data collection methods?			
Team Composition				
12.	Did the report (or methods annex) indicate that the evaluation team leader was external to USAID?			
13.	Did the report (or methods annex) identify at least one evaluation specialist on the team?			
14.	Did the report (or methods annex) identify local evaluation team members?			

⁵⁶ 11/23/13 version

⁵⁷ In this instrument we define N/A as “the conditions required to answer the question are not all present.”

⁵⁸ This question is not a numbered checklist question as it cannot be answered yes or no, but it nevertheless provides important information about the evaluation report.

Evaluation Report Review Checklist	Yes	No	N/A ⁵⁷
15. Did the report indicate that team members had signed Conflict of Interest forms or letters? <i>(check if the report says this or the COI forms are included in an annex)</i>			
Study Limitations			
16. Does the report include a description of study limitations (lack of baseline data; selection bias as to sites, interviewees, comparison groups; seasonal unavailability of key informants)?			
Responsiveness to Evaluation Questions			
17. Is the evaluation report structured to present findings in relation to evaluation questions, as opposed to presenting information in relation to program/project objectives or in some other format?			
18. Are <u>all</u> of the evaluation questions, including sub-questions, answered primarily in the body of the report (as opposed to in an annex)?			
19. If any questions were not answered, did the report provide a reason why?			
Findings			
20. Did the findings presented appear to be drawn from social science data collection and analysis methods the team described in its study methodology (including secondary data it assembled or reanalyzed)?			
21. For findings presented within the evaluation report is there a transparent connection to the source(s) of the data? <i>(60% of the beneficiaries' interviews reported that...)</i>			
22. In the presentation of findings, did the team draw on data from the range of methods they used rather than answer using data from primarily one method?			
23. Are findings clearly distinguished from conclusions and recommendations in the report, at least by the use of language that signals transitions (“the evaluation found that....” “the team concluded that”)?			
24. Are quantitative findings reported precisely, i.e., as specific numbers or percentages rather than general statements like “some”, “many”, or “most”?			
25. Does the report present findings about unplanned/unanticipated results?			
26. Does the report discuss alternative possible causes of results/outcomes it documents?			
27. Are evaluation findings disaggregated by sex at all levels (activity, outputs, outcomes) when data are person-focused?			
28. Does the report explain whether access/participation and/or outcomes/benefits were different for men and women when data are person-focused?			
Recommendations			
29. Is the report’s presentation of recommendations limited to recommendations? <i>(free from repetition of information already presented or new findings not previously revealed)</i>			
30. Do evaluation recommendations meet USAID policy expectations with respect to being specific? <i>(states clearly what is to be done, and possibly how?)</i>			
31. Do evaluation recommendations meet USAID policy expectations with respect to being directed to a specific party? <i>(identifies who should do it)</i>			
32. Are all the recommendations supported by the findings and conclusions presented? <i>(Can a reader follow a transparent path from findings to conclusions to recommendations?)</i>			
Annexes			
33. Is the evaluation SOW included as an annex to the evaluation report?			
34. Are sources of information that the evaluators used listed in annexes?			
35. Are data collection instruments provided as evaluation report annexes?			
36. Is there a matching instrument for <u>each</u> and <u>every</u> data collection method the team reported that they used?			
37. Were any “Statements of Differences” included as evaluation annexes (prepared by team members, the Mission, the implementing partner, or other stakeholder)?			
Evaluation Data Warehousing			

Evaluation Report Review Checklist	Yes	No	N/A ⁵⁷
38. Does the evaluation report explain how/in what form the evaluation data will be transferred to USAID (survey data, focus group transcripts)?			
Link to Evaluation Policy quality standards (proxy for evaluation team awareness of expectations)			
39. Does the evaluation SOW include a copy or the equivalent of Appendix I of the Evaluation Policy?			

B. Evaluation Report Review Checklist – Rater’s Guide

B. Evaluation Report Review Checklist - Rater’s Guide ⁵⁹	
Executive Summary	
1. Does the executive summary present an accurate reflection of the most critical elements of the report?	An executive summary must provide an accurate representation of the gist of the evaluation report without adding any new “material” information or contradicting the evaluation report in any way. “Critical” implies that not all information included in the evaluation report needs to be present in the executive summary, but that critical information from <u>all major elements should be discussed (i.e., evaluation purpose, questions, background information, methods, study limitations, findings, and recommendations)</u> . If an executive summary is not present, mark “N/A.”
Program/Project Background	
2. Are the basic characteristics of the project or program described (title, dates, funding organization, budget, implementing organization, location/map, target group)?	The project description plays a critical role in enabling the reader to understand the context of the evaluation, and involves several characteristics such as the title, dates, funding organization, budget, implementing organization, location/map, and target group. All of these characteristics play an important role and virtually all should be present to receive credit for this item in order to take a holistic view of whether the project is sufficiently well-described. If one or two characteristics are missing or weak but you get the gist of the project and can answer all future questions, then check “yes.”
3. Is the project or program’s “theory of change” described (intended results (in particular the project Purpose); development hypotheses; assumptions)	The “theory of change” describes, via narrative and/or graphic depiction of the intended results and causal logic, how anticipated results will be achieved. You may see this described as the development hypotheses and assumptions underlying the project or program. We expect that a clear explanation of the theory of change/development hypotheses will be presented in the evaluation report <u>before</u> the evaluation’s finding are presented.
Evaluation Purpose	
4. Does the evaluation purpose identify the management reason(s) for undertaking the evaluation?	Evaluation policy states that USAID is conducting evaluations for learning and accountability purposes. Beyond that, it is important that the evaluation purpose identifies the specific decisions or actions the evaluation is expected to inform (e.g., continue, terminate, expand, or redesign an intervention). If a statement of the evaluation purpose is not found, or is only present in the SOW, mark “N/A.”

⁵⁹ For this checklist the term N/A means that the conditions needed to rate a particular item are not present. For example, if no evaluation questions were included in the evaluation report, then later items that ask about characteristics of the evaluation questions cannot be answered and should be rated N/A. Shading on the checklist response column indicates with N/A is an allowable answer.

B. Evaluation Report Review Checklist - Rater's Guide⁵⁹

Evaluation Questions	
5. Are the evaluation questions clearly related to the evaluation purpose?	The evaluation questions, as stated in the evaluation report, should have a direct and clear relationship to the stated evaluation purpose. If no evaluation questions are provided in the body of the report before the findings, or in the SOW, check "N/A." Even if questions are provided, this question cannot be answered if no evaluation purpose was included. Thus if item (4) above indicated that there was no purpose stated, then this question must be marked "N/A."
6. Are the evaluation questions in the report identical to the evaluation questions in the SOW?	This question is about evaluation questions found in the body of the report and in the SOW. There must be questions in both places in order address this question. If questions are present in only one of these two places, mark "N/A."
7. If the questions in the body of the report and those found in the SOW differ, does the report (or annexes) state that there was written approval for changes in the evaluation questions?	The evaluation SOW is the contract evaluators work from, so it is imperative that the questions/issues in the body of the evaluation report match those included in the SOW word for word. If the evaluation team changed, removed, or added evaluation questions/issues, USAID policy states that they should only have done so with written approval from USAID. While this written approval does not need to be included in an annex, it does need to be mentioned in the body of the report. If the answer to 6 is "yes" or "N/A" then mark 7 as "N/A." If the answer to 6 is "no" then answer 7 with a "yes" or "no."
Methodology	
8. Does the report (or methods annex) describe <u>specific</u> data collection methods the team used?	USAID requires that an evaluation report identify the data collection methods used, but does not indicate where this information must be presented. It is common to include the methodology description in the body of the report with a longer and more detailed methods annex, so be sure and check the annex. To receive credit, the methods description must be specific on how and from whom data will be collected. It is insufficient to say, "interviews will be conducted." To be adequate a description of methods must indicate what types of interviews, estimated numbers, and with whom they will be conducted (e.g., key informant interviews, individual interviews with beneficiaries, group interviews).
9. Are the data collection methods presented (in the report or methods annex) in a manner that makes it clear which specific methods are used to address <u>each</u> evaluation question (e.g., matrix of questions by methods)?	USAID How-To guidance on evaluations advises that data collection methods should be explained in relation to each evaluation question/issue the evaluation team addressed. This information may be found within the body of the report or may be presented in a methods or design annex. While the methods can be associated to questions in a variety of ways, some evaluations use a matrix for this purpose that lists an evaluation question and then describes the data sources, data collection methods, sampling strategies, and data analysis methods. If no data collection methods are provided, or if

B. Evaluation Report Review Checklist - Rater's Guide⁵⁹

	no questions/issues exist, check the box for "N/A."
10. Does the report (or methods annex) describe <u>specific</u> data analysis methods the team used? (frequency distributions; cross-tabulations; correlation; reanalysis of secondary data)	USAID requires that an evaluation report identify the data analysis methods used, but does not indicate where this information must be presented. It is common to include the methodology description in the body of the report with a longer and more detailed methods annex. To receive credit, the data analysis methods description must be <u>specific about how, or through what method, data will be analyzed</u> . It is insufficient to say, "qualitative and quantitative analyses will be conducted" and instead must provide detailed information on the kinds of analyses to be conducted (e.g., frequency distributions, cross-tabs, correlations, content analysis, pattern analysis).
11. Are the data analysis methods presented (in the report or methods annex) in a manner that makes it clear how they are associated with the evaluation questions or specific data collection methods?	The evaluation report should make it clear which data analysis methods described were used to analyze data to answer specific evaluation questions/issues. [The question parallels #9 above for data collection methods.] Information on data analysis methods may be available within the body of the report or may be found in a methods or design annex. As indicated under item (9), some reports include a matrix that describes data analysis approaches as well as data collection methods in relation to each evaluation question. Note that wherever a discussion of data analysis methods takes place, it is acceptable for this description to relate data analysis methods to data collection methods, instead of directly to evaluation questions. If no data analysis methods are provided (marked "no" for previous question, #9), or if no questions exist, check the box for "N/A."
Team Composition	
12. Did the report (or methods annex) indicate that the evaluation team leader was external to USAID?	USAID counts an evaluation as being external if the team leader is external, meaning that the team leader is an independent expert from outside of USAID who has no fiduciary relationship with the implementing partner. If the evaluation is a self-evaluation (USAID or its implementing partner is evaluating their own project/activity) then this answer must be no. To receive credit, the evaluation must indicate the team leader in either the body of the report (including cover or title page) or in the methods section. A search for the term "team leader" may expedite this process. <u>If the report is not explicit in stating the team leader was external, it may be inferred from a description of the team leader or the organization with which they are associated (e.g., university professor or evaluation firm that is not the project implementer)</u> . Independence may also be confirmed via a "no-conflict of interest" statement often included as an annex. If the report identifies that the team was independent, but there is no designated team leader, check "N/A."
13. Did the report (or methods annex) identify at least one evaluation specialist on the	At least one member of the evaluation team must be an evaluation specialist and clearly indicated as such in either the body of the

B. Evaluation Report Review Checklist - Rater's Guide⁵⁹

team?	report or in the methods annex. The term “evaluation specialist” must be explicit and not implied.
14. Did the report (or methods annex) identify local evaluation team members?	USAID encourages the participation of country nationals on evaluation teams. The report need not use the word “local” specifically, but can be referred to by designation such as “Brazilian education specialist,” if in Brazil. This person could be any country national, including a foreign service national (FSN). Simply guessing a person’s country of origin based on their name is insufficient. Do not guess.
15. Did the report indicate that team members had signed Conflict of Interest forms or letters (check if the report says this or the COI forms are included in an annex)?	USAID requires that evaluation team members certify their independence by signing statements indicating that they have no conflict of interest or fiduciary involvement with the project or program they will evaluate. USAID guidance includes a sample Conflict of Interest form. It is expected that an evaluation will indicate that such forms, or their equivalent, are on file and available or are provided in an evaluation annex.
Study Limitations	
16. Does the report include a description of study limitations (lack of baseline data; selection bias as to sites, interviewees, comparison groups; seasonal unavailability of key informants)?	It is common for evaluators to encounter unexpected interferences with anticipated study designs such as unavailability of key informants or lack of access to activity sites. In other instances, stakeholder preferences may introduce selection biases. In any such instance, evaluators are obligated to include these “study limitations” and a description of the impact they have had on the evaluation. Study limitations may only be included for this item if they directly impact the evaluator’s ability to credibly and effectively answer an evaluation question (i.e., if all data can still be collected, even if inconveniently or at a higher cost, it is not a limitation). Limitations do not need to have their own distinct section provided they are located towards the end of the methodology description and before the introduction of findings.
Report Structure Responsiveness to Evaluation Questions	
17. Is the evaluation report structured to present findings in relation to evaluation questions, as opposed to presenting information in relation to project objectives or in some other format?	The most straightforward way to meet USAID’s requirement that every evaluation question/issue be addressed, is a question-by-question (or issue-by-issue) report structure. Historically, evaluations have not always taken this approach, and instead structured the report around such things as project objectives, or locations. If no evaluation questions/issues exist around which a report could be structured, check “N/A.” If the evaluation questions/issues and the team’s answers to those questions/issues are the dominant structure of the report, check “yes.”
18. Are <u>all</u> of the evaluation questions, including sub-questions, answered	The purpose of an evaluation report is to provide the evaluators’ findings and recommendations on <u>each</u> and <u>every</u> evaluation question. Accordingly, USAID expects that the answers to all

B. Evaluation Report Review Checklist - Rater's Guide⁵⁹

<p>primarily in the body of the report (as opposed to in an annex)</p>	<p>evaluation questions/issues, including any sub-questions/issues, will be provided primarily in the body of the report. Answering main questions/issues in the body and sub-questions/issues in an annex is not consistent with USAID expectations. If no evaluation questions/issues are provided (either in the body of the report or in an annex) to which a team could respond, check "N/A."</p>
<p>19. If any questions were not answered, did the report provide a reason why?</p>	<p>If the answer to question 18 is "yes," mark this answer as "N/A." If the answer to question 18 is "no," does the evaluation report provide an explanation as to why specific questions were not answered or were answered somewhere other than in the body of the report?</p>
<p>Findings</p>	
<p>20. Did the findings presented appear to be drawn from social science data collection and analysis methods the team described in study methodology (including secondary data assembled or reanalyzed)?</p>	<p>USAID's commitment to evidence-based decision-making is necessitating a shift to stronger and more replicable approaches to gathering data and presenting action recommendations to the Agency. The more consistent use of credible social science data collection and analysis methods in evaluations is an important step in that direction (e.g., structured and well documented interviews, observation protocols, survey research methods). If the report did not describe the data collection and analysis methods used, check "N/A."</p>
<p>21. For the findings presented within the evaluation report is there a transparent connection to the source(s) of the data? (60% of the beneficiaries interviews reported that...; reanalysis of school records shows....; responses from mayors indicate that...)</p>	<p>While most evaluation reports present sets of findings, it is not always clear where those findings came from. It is helpful to the reader to connect the sources of data to the findings those data are being used to support. For example, "children's consumption of protein increased" does not indicate where that finding came from. Alternatively, "60% of mothers who participated in the survey stated that their children's consumption of protein had increased" does a good job of connecting the finding to the source. This is true for both qualitative and quantitative findings. If the findings in the report were connected to sources of data as indicated above, check "yes." If findings are generally presented without reference to their source, check "no."</p>
<p>22. In the presentation of findings, did the team draw on data from the range of methods they used rather than answer using data from or primarily one method?</p>	<p>In addressing this question, only include those methods specifically referenced in the methods section of the report or in the methods annex. Of the methods actually used, the evaluation should demonstrate a balanced use of data from all data collection methods. If no methodologies were introduced from which they could later be drawn on, check "N/A."</p>
<p>23. Are findings clearly distinguished from conclusions and recommendations in the report, at least by the use of</p>	<p>As defined by the Evaluation Policy, evaluation findings are "based on facts, evidence, and data...[and] should be specific, concise, and supported by quantitative and qualitative information that is reliable, valid, and generalizable." The presence of opinions, conclusions,</p>

B. Evaluation Report Review Checklist - Rater's Guide⁵⁹

language that signals transitions (“the evaluation found that...” or “the team concluded that...”)?	and/or recommendations mixed in with the descriptions of findings reduces a finding’s ability to meet USAID’s definition.
24. Are quantitative findings reported precisely, i.e., as specific numbers or percentages rather than general statements like “some,” “many,” or “most”?	When presenting quantitative findings it is important to be precise so that the reader knows exactly how to interpret the findings and is able to determine the accuracy of the conclusions drawn by the evaluators. Precision implies the use of specific numbers and/or percentages as opposed to general statements like “some,” “many,” or “most.” If no potentially quantitative findings are provided, check “N/A.”
25. Does the report present findings about unplanned/unanticipated results?	While evaluators may be asked to look for unplanned or unanticipated results in an evaluation question, it is common to come across such results unexpectedly. If such results are found, by request or unexpectedly, they should be included in the report.
26. Does the report discuss alternative possible causes of results/outcomes it documents?	Though evaluators may be asked to look for alternative causes of documented results or outcomes in an evaluation question, it is possible for evaluators to come across such potential alternative causes unexpectedly. If any such causes are found, it is important that the evaluators bring such information to the attention of USAID.
27. Are evaluation findings disaggregated by sex at all levels (activity, outputs, outcomes) when data are person-focused?	The Evaluation Policy and USAID in general are making a big push for gathering sex-disaggregated data whenever possible. To support this focus, it is valuable for evaluators to include data collection and analysis methods that enable sex-disaggregation whenever the data they anticipate working with will be person-focused. Such data should be represented at all project levels from activities to outputs to outcomes to the extent possible. If no person-focused data was collected and therefore there was no data that could be disaggregated by sex, check “N/A.”
28. Does the report explain whether access/ participation and/or outcomes/benefits were different for men and women when data are person-focused?	USAID expects that evaluations will identify/discuss/explain how men and women have participated in, and/or benefited from, the programs and projects it evaluates. This involves more than simply collecting data on a sex-disaggregated basis. Addressing this issue can be presented in one general section or on a question-by-question basis; either is acceptable. If data was not collected in a person-focused manner for the evaluation, check “N/A.”
Recommendations	
29. Is the report’s presentation of recommendations limited to recommendations (free from repetition of information already	Presentation of recommendations in an evaluation report affects the usability of the report. Recommendations build on information previously introduced through findings and conclusions. Therefore, the presentation of recommendations does not need supporting findings and conclusions repeated or any new supporting findings or

B. Evaluation Report Review Checklist - Rater's Guide⁵⁹	
presented or new findings not previously revealed)?	conclusions introduced. The presence of any information other than the specific, practical, and action-oriented recommendations could have a diminishing effect on report usability. If no recommendations are present in the report, check "N/A."
30. Do evaluation recommendations meet USAID policy expectations with respect to being specific (states what exactly is to be done, and possibly how)?	Recommendations that are specific are inherently more actionable than those which are not. The recommendation, "improve management of the project," is much less specific than one that says "streamline the process for identifying and responding to clinic needs for supplies in order to reduce gaps in service delivery." If no recommendations are presented in the evaluation report, check "N/A."
31. Do evaluation recommendations meet USAID policy expectations with respect to being directed to a specific party?	USAID encourages evaluation teams to identify the parties who need to take action on each recommendation. Doing so makes it easier for USAID staff to understand and act on and evaluations implications. If no recommendations are presented in the evaluation report, check "N/A."
32. Are all the recommendations supported by the findings and conclusions presented (can a reader follow a transparent path from findings to conclusions to recommendations)?	Managers are more likely to adopt evaluation recommendations when those evaluations are based on credible empirical evidence and an analysis that transparently demonstrates why a specific recommendation is the soundest course of action. To this end, USAID encourages evaluators to present a clear progression from Findings → Conclusions → Recommendations in their reports, such that none of a report's recommendations appear to lack grounding, or appear out of "thin air." If no recommendations are presented in the evaluation report, check "N/A."

Annexes	
33. Is the evaluation SOW included as an annex to the evaluation report?	This question checks on evaluation team responsiveness to USAID's Evaluation Policy, Appendix I, requirement for including an evaluation SOW as an evaluation report annex.
34. Are sources of information that the evaluators used listed in annexes?	USAID's Evaluation Policy, Appendix I, requires sources of information to be included as an evaluation report annex. Sources include both documents reviewed and individuals who have been interviewed. Generally it is not expected that names of survey respondents or focus group participants will be individually provided, as these individuals are generally exempted based on common/shared expectations about maintaining confidentiality with respect to individual respondents.
35. Are data collection instruments provided as evaluation report annexes?	This question focuses on the inclusion of data collection instruments in an evaluation annex including interview guides or survey

B. Evaluation Report Review Checklist - Rater's Guide⁵⁹	
	questionnaires.
36. Is there a matching instrument for each and every data collection method the team reported that they used?	This question examines how comprehensive a set of the instruments used for collecting data for a USAID evaluation a report provides. USAID's standard in its Evaluation Policy is "all" tools.
37. Were any "Statements of Differences" included as evaluation annexes (prepared by team members, or the Mission, or implementing partner, or other stakeholders)	Including "Statements of Differences" has long been a USAID evaluation report option. This question determines how frequently "Statements of Differences" are actually included in USAID evaluations. Statements are often written by evaluation team members, or alternatively by the Mission, a stakeholder, or implementing partner. If one or more "Statements of Differences" are included, check "yes."
Evaluation Data Warehousing	
38. Does the evaluation report explain how the evaluation data will be transferred to USAID (survey data, focus group transcripts)?	USAID Evaluation Policy (p. 10) calls for the transfer of data sets from evaluations to USAID, so that, when appropriate, they can be reused in other assessment and evaluations. Given this requirement, it is helpful if an evaluation report indicates how and when that transfer was made.
SOW Leading Indicator of Evaluation Quality (answer if SOW is a report annex)	
39. Does the evaluation SOW include a copy or the equivalent of Appendix I of the Evaluation Policy?	USAID policy requires that statements of work (SOWs) for evaluations include the language of Appendix I of the USAID Evaluation Policy. If no SOW is included as an annex to the evaluation report, check "N/A."

C. Overall Evaluation Report Review Score

Creating the Overall Evaluation Report Quality Score

The overall evaluation quality "score" is a calculation based on eleven of the factors included in this checklist. To calculate the "overall score" for the evaluation, award 1 point for "yes" on items 1, 8, 10, 16, 20, 23, 32, 34 and 35. If both item 2 and 3 are "yes", award one more point, for a total of 10 maximum points.

ANNEX 6 – BIBLIOGRAPHY

- Alkin, Marvin C. and Sandy M. Taut. "Unbundling Evaluation Use." *Studies in Educational Evaluation* 29 (2003): 1-12. Accessed March 30, 2015. <http://escholarship.org/uc/item/2q69w6z6>.
- Allen, Marisa. "Steeping the organization's tea: Examining the relationship between evaluation use, organizational context, and evaluator characteristics." PhD diss., Case Western Reserve University, Mandel School of Applied Social Sciences, 2010. Accessed July 15, 2015. http://msass.case.edu/downloads/vgroza/Allen_Dissertation.pdf.
- Allen, Susan, Mattias Kjaer, Kelly Smith, and Gabrielle Plotkin (Social Impact), and David Hunsicker and Rebekah Krimmel (USAID/CMM). "Evaluative Learning Review Synthesis Report: USAID/CMM's People-to-People Reconciliation Fund, Annual Program Statement (APS)." Social Impact, March 2014. http://pdf.usaid.gov/pdf_docs/pbaaa370.pdf
- Amo, Courtney. "Investigating the relationship between process use and use of evaluation findings in a government context." MA thesis, University of Ottawa, Faculty of Education, 2009. Accessed June 26, 2015. <http://www.ruor.uottawa.ca/handle/10393/28077>.
- Brandon, Paul R. and J. Malkeet Singh. "The Strength of the Methodological Warrants for the Findings of Research on Program Evaluation Use." *American Journal of Evaluation* 30, no. 2 (2009): 123-157. Accessed June 26, 2015. <http://aje.sagepub.com/content/30/2/123.short>.
- Bollen, Kenneth, Pamela Paxton and Rumi Morishima. "Assessing International Evaluations, An Example From USAID's Democracy and Governance Program." *American Journal of Evaluation* 26 no. 2 (2005):189-203.
- Burr, Erin Mehalic. "Evaluation Use and Influence among Project Directors of State GEAR UP Grants." PhD diss., University of Tennessee, Knoxville, 2009. Accessed March 24, 2015. http://trace.tennessee.edu/utk_graddiss/15/.
- Center for Global Development. *When Will We Ever Learn? Improving Lives Through Impact Evaluation. A Report of the Evaluation Gap Working Group.* Washington, D.C.: Center for Global Development (2006)
- Christie, Christina A. "Reported Influence of Evaluation Data on Decision Makers' Actions." *American Journal of Evaluation* 28, no. 1 (2007): 8-25. Accessed June 29, 2010. <http://aje.sagepub.com/content/28/1/8.short>.
- Cousins, J. B. (2003) "Utilization Effects of Participatory Evaluation," in T. Kelligan and D. L. Stufflebeam (eds.) *International Handbook of Educational Evaluation*, pp. 245–66. Dordrecht: Kluwer Academic Press.
- Cousins, J. Bradley and Kenneth A. Leithwood. "Current Empirical Research on Evaluation Utilization." *Review of Educational Research* 56, no. 3 (1986): 331-364. Accessed March 26, 2015. <http://rer.sagepub.com/content/56/3/331.short>.
- Daigneault, Pierre-Marc. "Taking stock of four decades of quantitative research on stakeholder participation and evaluation use: A systematic map." *Evaluation and Program Planning* 45 (2014): 171-181. Accessed June 26, 2015. <http://www.sciencedirect.com/science/article/pii/S0149718914000421>.

- Feinstein, Osvaldo. "Use of Evaluations and Evaluation of their Use." *Evaluation* 8, No. 4 (2002): 433-439. <http://evi.sagepub.com/content/8/4/433.abstract>.
- Fleischer, Dreolin N. and Christina A. Christie. "Evaluation Use: Results From a Survey of U.S. American Evaluation Association Members." *American Journal of Evaluation* 30, no. 2 (2009): 158-175. Accessed http://www.managingforimpact.org/sites/default/files/resource/evaluation_use_results_from_a_survey.pdf
- Freudenberger, Elizabeth, "Sectoral Synthesis of 2013-2014 Evaluation Findings." *Management Systems International* (August 2015). http://pdf.usaid.gov/pdf_docs/PA00KQT8.pdf
- Frietas, Sheizi Chalheira de. "An Exploratory Study on ENADE Evaluation Report Utilization and its Impact on Undergraduate Accounting Program Performance in Brazil." Master's thesis, University of Sao Paulo, College of Economics, Business and Accounting, 2012. Accessed June 9, 2015. <http://www.teses.usp.br/teses/disponiveis/12/12136/tde-15012013-134543/publico/SheiziCalheiradeFreitasVC.pdf>.
- GAO. "Program Evaluation: Strategies to Facilitate Agencies' Use of Evaluation in Program Management and Policy Making." United States Government Accountability Office, Report to Congressional Committees, GAO-13-570 (June 2013). Accessed June 9, 2015. <http://www.gao.gov/products/GAO-13-570>.
- Hageboeck, Molly and Gregory Norfleet. "Mobiles for Education Evaluation Abstracts." *Management Systems International, Inc. & Development & Training Services, E3 Analytics and Evaluation Project*. (June 2015). http://www.meducationalliance.org/sites/default/files/meducation_evaluation_abstracts_6-30-15.pdf
- Hageboeck, Molly. "Trends in International Development Evaluation Theory, Policy and Practices." *Management Systems International* (August 2009). http://pdf.usaid.gov/pdf_docs/PNADQ464.pdf.
- Hageboeck, Molly, Micah Frumkin, and Stephanie Monschein. "Meta-Evaluation of Quality and Coverage of USAID Evaluations: 2009-2012." *Management Systems International* (August 2013). http://pdf.usaid.gov/pdf_docs/PDACX771.pdf.
- Henry, Gary T. and Melvin M. Mark. "Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions." *American Journal of Evaluation* 24, no. 3 (2003): 293-314. Accessed December 12, 2014. <http://aje.sagepub.com/content/24/3/293.short>.
- Herbert, James Leslie. "The Influence of evaluation on human service practitioner learning: An analysis of two case studies of practitioners' experiences of evaluation." PhD diss., University of Western Sydney, School of Social Sciences and Psychology, 2013. Accessed March 23, 2015. <http://uwsprod.uws.dgicloud.com/islandora/object/uws%3A21258/datastream/PDF/download/citation.pdf>.
- Herbert, James Leslie. "Researching Evaluation Influence: A Review of the Literature." *Evaluation Review* 38, No. 5 (2014): 388-419. Accessed March 18, 2015. <http://erx.sagepub.com/content/38/5/388.short>.
- Johnson, Kelli, Liji O. Greenseid, Stacie A. Toal, Jean A. King, Frances Lawrenz, and Boris Volkov. "Research on Evaluation Use: A Review of Empirical Literature From 1986 to 2005." *American*

- Journal of Evaluation* 30, No. 3 (2009). Accessed June 29, 2010.
http://www.wmich.edu/evalphd/wp-content/uploads/2010/05/Research_on_Evaluation_Use.pdf.
- Justice Research and Statistics Association, United States, Bureau of Justice Assistance. "Is This a Good Quality Outcome Evaluation Report? A Guide for Practitioners." The Bureau of Justice Assistance Center for Program Evaluation and Performance Measurement (2011).
- Kirkhart, K. E. "Reconceptualizing evaluation use: An integrated theory of influence." *New Directions in Evaluation* 88, (2000): 5-23. <http://onlinelibrary.wiley.com/doi/10.1002/ev.1188/abstract>.
- Ledermann, Simone. "Exploring the Necessary Conditions for Evaluation Use in Program Change." *American Journal of Evaluation* 33, no. 2 (2012): 159-178. Accessed June 26, 2015.
<http://aje.sagepub.com/content/33/2/159.short>.
- Leviton, Laura C. "Evaluation Uses: Advances, Challenges and Applications." *American Journal of Evaluation* 24, no. 4 (2003): 525-535. Accessed March 28, 2015.
<http://www.sciencedirect.com/science/article/pii/S1098214003000870?np=y>.
- Leviton, Laura C. and Edward F. X. Hughes. "Research on the Utilization of Evaluations." *Evaluation Review* 5, no. 4 (1981): 525-548. Accessed March 26, 2015.
<http://erx.sagepub.com/content/5/4/525.short>.
- Loud, Marlène Läubli and John Mayne (ed.). *Enhancing Evaluation Use: Insights from Internal Evaluation Units*. Los Angeles, CA: Sage (2014).
- Mark, Melvin M. and Gary T. Henry. "The Mechanisms and Outcomes of Evaluation Influence." *Evaluation* 10, no. 1 (2004): 35-57. Accessed March 24, 2015.
<http://evi.sagepub.com/content/10/1/35.abstract>.
- Mayhew, Fred. "Integrating the Funder-Fundee Relationship into a Model of Evaluation Utilization." *Journal of MultiDisciplinary Evaluation* 7, no. 16 (2011): 1-19. Accessed July 15, 2015.
http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/315.
- OECD/DAC. "Evaluation in Development Agencies" DAC Guidelines and Reference Series.
http://www.oecd-ilibrary.org/development/evaluation-in-development-agencies_9789264094857-en.
- Oliver, Monica LaBelle, "Evaluation of Emergency Response: Humanitarian Aid Agencies and Evaluation Influence." PhD diss., Georgia State University, 2008. Accessed March 24, 2015.
http://scholarworks.gsu.edu/pmap_diss/23.
- Reboot with FHI 360 and USAID's Global Development Lab. "Mission data utilization project: Uganda Research Summary." Internal Report.
- Spencer, Liz, Jane Ritchie, Jane Lewis, and Lucy Dillon. "Quality in Qualitative Evaluation: A framework for assessing research evidence." Cabinet Office, Government Chief Social Researcher's Office. August, 2013.
- Sridharan, Sanjeev and April Nakaima. "Ten steps to making evaluation matter." *Evaluation and Program Planning* 34, no. 2 (2011): 135-146. Accessed December 30, 2014.
<http://www.sciencedirect.com/science/article/pii/S0149718910000819>.
- Stake, Robert E. and Thomas A. Schwant. "On Discerning Quality in Evaluation." In Ian Shaw, Mel Mark, and Jennifer Greene (eds.), *Handbook of Evaluation* (2006): 419-438.

Vargas Banuelos, Noelle. "Tracing the Pathways of Evaluation Influence: From Theory to Practice." PhD diss., University of California, Los Angeles, 2012. <http://gradworks.umi.com/35/06/3506171.html>.

The World Bank. "Making Smart Policy: Using Impact Evaluation for Policy Making: Case Studies on Evaluations that Influenced Policy." *Doing Impact Evaluations*, no. 14 (June 2009). http://siteresources.worldbank.org/INTISPMA/Resources/383704-1146752240884/Doing_ie_series_14.pdf.

The World Bank. "Influential Evaluations: Evaluations that Improved Performance and Impacts of Development Programs." Washington D.C.: International Bank for Reconstruction and Development, 2004. <http://documents.worldbank.org/curated/en/2004/01/6051317/influential-evaluations-evaluations-improved-performance-impacts-development-programs>.

The World Bank. 2008a. Annual Report 2008. Washington, D.C.: World Bank. [http://siteresources.worldbank.org/EXTANNREP2K8/Resources/YR00 Year in Review English.pdf](http://siteresources.worldbank.org/EXTANNREP2K8/Resources/YR00%20Year%20in%20Review%20English.pdf)

Yin, Robert K., Carol H. Weiss, Oona M. Cheung, and Joanne Capper. "Preliminary Study of the Utilization of A.I.D.'s Evaluation Reports." Cosmos Corporation. December 31, 1988.

ANNEX 7 – STORIES OF EVALUATION USE RESULTS

This annex to MSI's report for PPL/LER owes a debt of gratitude and inspiration to the World Bank team, led by Keith Mackay, Michael Bamberger, and Patrick Grasso, for their 2004 volume, *Influential Evaluations*. Those stories of improved effectiveness, efficiency, and development impact that flowed from the utilization of World Bank evaluations made us certain that such stories would also be found in USAID's experience, if only we made an effort to look for them.

This annex, together with the report it accompanies, could not have been assembled without the generous contribution of time and memories from dozens of USAID staff members around the world. They told us their stories about the results of actions taken on the findings and recommendations in evaluations their USAID Operating Units in Washington and overseas commissioned. The study team also wishes to thank USAID's team in the Office of Learning, Evaluation, and Research (LER), specifically Tania Alfonso, Jerome Gallagher, and Negar Akhavi, for their continuous support and guidance throughout the process. Outside of USAID, Jennifer Dahnke at the Feed the Future Knowledge-Driven Agriculture Development Project (KDAD) provided valuable substantive and procedural support. In addition, USAID's Development Experience Clearinghouse (DEC) staff provided us with invaluable assistance in locating evaluations and deserves credit for the improvements it has made in this collection over the past few years.

MSI would also like to acknowledge our evaluation research team for this study, and for their diligent efforts to follow up on stories identified through interviews, survey research, crowdsourcing calls, and USAID document reviews. Among the MSI team whose inputs on this stories volume were invaluable are Micah Frumkin, Jenna Lindeke Heavenrich, and Lala Kasimova.

Molly Hageboeck
Team Leader/MSI Technical Director

INTRODUCTION

This annex is a companion to MSI's 2015 study report on *Evaluation Utilization at USAID*. That study was undertaken to determine how and in what ways USAID evaluations are being used and how the application of evaluation findings might be improved. In addition to being of use to USAID's own staff, it was anticipated that the main report and this companion volume would be of interest to a range of stakeholders including the Department of State and other USG colleagues on country teams; Congress; the Office of Management and Budget (OMB); and country partners and other donors with whom USAID collaborates overseas.

Ten stories about the results of evaluation utilization are included in this volume. They are presented in a brief format designed to encourage their use either individually, or in small groups, and not necessarily as the entire volume.

These 10 were developed from a larger set of 57 stories that emerged from interviews, survey research, a crowdsourcing effort, and study team reviews of USAID documents such as Country Development Cooperation Strategies (CDCSs). For each story included in this volume, MSI's study team undertook online follow-up research to validate story elements and bring them up to date, where available resources permitted.

As it researched evaluation utilization in USAID, the study team found that most evaluation utilization stories involved a progression along which several types of results emerged. As in the main volume, we characterize those results here as:

- **Direct actions or changes** that USAID or one of its partners makes based on evaluation findings or recommendations, such as making adjustments in the priorities, work plans, or timelines of ongoing activities;
- **Discernible effects of direct action or changes**, such as improvements in the efficiency or effectiveness of an ongoing activity; and
- **Changes in development outcomes**, such as more beneficiaries whose lives are improved because activities and projects became more effective, efficient, or sustainable.

The remainder of this volume is dedicated to the stories told by USAID staff and expanded through follow-up research. While categorizing actual evaluation utilization stories into the results stages above is invariably imprecise, the 10 stories are clustered to illustrate results at each of these stages.

The Value of Evaluation

“Evaluation is useful only insofar as it provides evidence to inform real-world decision making.”

–USAID Evaluation Policy

STORIES OF EVALUATION USE RESULTS:

Direct actions or changes introduced by USAID or its partners

1. Afghanistan – Improving Training Under the Workforce Development Program (found in the body of the report)
2. Indonesia – Focusing USAID’s Forestry Strategy
3. Morocco – Improving Capacity in Civil Society Organizations (CSOs)
4. Somalia – Strengthening Local Presence in Support of Program Outcomes

Discernible effects of initial actions taken based on evaluation results

5. East Africa – Raising Awareness on Countering Violent Extremism
6. Afghanistan – Refining USAID’s Construction Management Policy

Changes in Development Outcomes that Appear to Flow from Evaluation Utilization

7. Ethiopia – Scaling Up HIV Testing and Treatment (found in the body of the report)
8. Central America – Improved Municipal Crime Prevention
9. Armenia – Advancing Least Cost Energy Solutions
10. Nigeria – Scaling Up Improved Agricultural Productivity

FOCUSING USAID'S FORESTRY STRATEGY IN INDONESIA BASED ON EVALUATION

In 2013, USAID/Indonesia commissioned an evaluation of its Forest Resource Sustainability Program (FOREST), including its USAID/Indonesia Forest and Climate Support (IFACS) activity. This evaluation found that USAID is spread out among too many landscapes to be effective and recommends USAID/Indonesia “select a sub-set of focal IFACS landscapes for intensification of efforts to achieve meaningful results ... [and] redefine IFACS strategies based on realistic goals.”

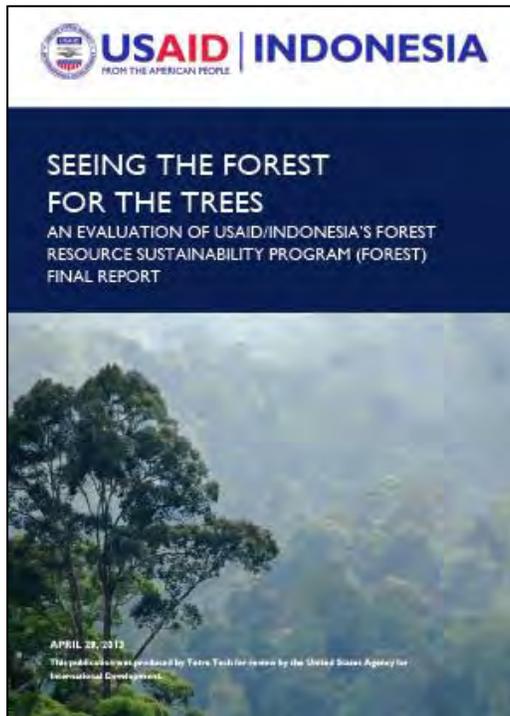


Photo:
TetraTech, <http://www.tetrattech.com/en/projects/indonesia-forestry-and-climate-support-project>

**Excerpt from: Seeing the Forest for the Trees:
An Evaluation of USAID/Indonesia's Forest Resource Sustainability Program (FOREST)**
 (http://pdf.usaid.gov/pdf_docs/PA00JP2G.pdf)

“Within IFACS the landscape approach has not always been well used. While the landscape approach is meant to be a problem-based approach for planning to identify interventions that would positively impact the focus landscape, IFACS field staff have often taken it to mean that interventions are to be focused within the limits of the landscape. Rather than working to have impact on target landscapes, IFACS regional teams have often limited themselves to working within the confines of landscapes. This has constrained strategic engagement and affected overall results achievement. Present IFACS leadership understands the issues and is taking steps to help regional managers take appropriate measures.

USAID experience has shown that a landscape approach can be a powerful tool to identify entry points and opportunities to achieve climate change and conservation results. This experience further suggests that this tool should be considered a core approach in the forest sector moving forward.”



According to an interview with staff from USAID/Indonesia's Environment Office:

“There were three or four key findings [from the evaluation] that really made us switch gears in terms of how we implemented our entire approach to the forestry sector. The first was that the evaluation found that we were just working in too many places and suggested that we focus and concentrate our geographic locations. It was a very, very difficult thing to do, but we had to actually consolidate down from about fifteen different landscapes to five. Then we also increased our budget to work in those five landscapes, and we had to increase our budget to add specific activities that were lacking before in the work that we were doing. We realized that we were, I guess I would say, a mile wide and an inch deep and that we really needed to go deeper and have more impactful interventions. It was only because of that evaluation that we made that decision.”

Further evidence of the use of this evaluation and its effects on USAID/Indonesia's strategy can be found in references to the evaluation in the Mission's Country Development Cooperation Strategy (CDCS), where the evaluation is cited as being influential in identifying areas necessary to achieve significant conservation progress.

BUILDING BETTER CSO CAPACITY AS A RESULT OF USAID EVALUATIONS

In 2012, an evaluation of USAID/Morocco’s Civil Society Advocacy Project (SANAD) revealed that the project’s “emphasis on quantity [of Civil Society Organizations (CSOs) reached] may have led to a reduced impact on each organization.” The evaluation also found that not all CSOs received the full suite of integrated training, funding, and technical assistance available, but those that did were able to achieve more sustainable results.



Evaluation Report Cover

Excerpt from: Evaluation of USAID/Morocco Civil Society Advocacy Project (SANAD)

(http://pdf.usaid.gov/pdf_docs/PDACW258.pdf)

“Over the three years of the project, SANAD interventions reached a total of around 700 organizations, of which 450 organizations benefited from at least one capacity reinforcement activity. The number of beneficiaries is impressive and reflects the project design’s emphasis on reaching a large number of organizations. It is obvious that such an emphasis may reduce the program’s impact on each individual organization that is targeted, and this outcome is indeed reflected in the findings of the evaluation... The SANAD Grantees that benefited from [an integrated package of training, funding and technical assistance] showed the greatest improvement in their organizational and advocacy capacities.”

Following the receipt of the evaluation report, USAID acted on the evaluation’s recommendations, incorporating the results into project design documents for upcoming awards and even incorporating them into the Country Development Cooperation Strategy. According to staff at USAID, they are actively using this new approach in the current civil society strengthening program: They are “very targeted, very focused on a select number of leading NGOs, and ... not going to go and try to work with hundreds, but just with a few CSOs on key themes and topics.” Similarly, that same staff member indicated that they were also incorporating the recommendation of an integrated package of “a grant, technical assistance on the themes and areas on which the CSOs work, and coaching and training along the way, along the implementation process.” The USAID/Morocco Mission’s Senior Management further endorsed the approach, indicating that they intended to apply this new focus to work being conducted in the sectors of education and economic growth activities as well.

Evidence of use of this evaluation can also be found in USAID/Morocco’s CDCS, where it states:



Photo: MSI,
<http://www.msiworldwide.com/wp-content/uploads/SANAD1.pdf>

“Building on its experience, USAID will enhance CSO organizational capacity and financial integrity by improving their ability to develop sustainable financial plans, maintain transparent organizational structures and increase technical capacity for more focused advocacy or governance oversight. Moroccan CSOs recognize their deficiencies and seek continued capacity building support. Evaluations of the most recent USAID civil society project noted that in order for this effort to be successful, CSO capacity building should be delivered through an integrated package of training, funding and technical assistance, and culminate in a specific advocacy effort. Accordingly, USAID will provide capacity development and mentoring to CSOs working in targeted reform areas, such as government accountability and women’s empowerment, and when appropriate, award direct grants to suitable projects.”

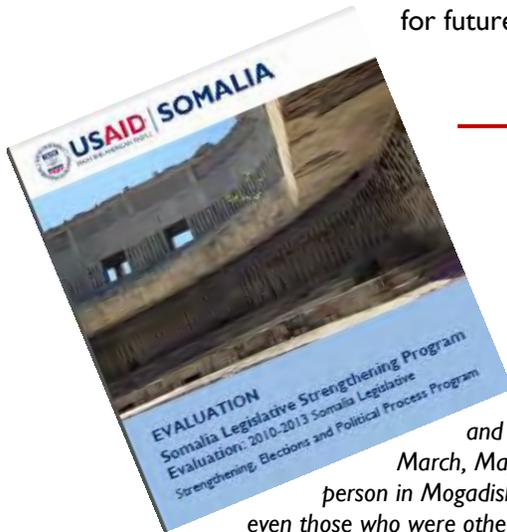
EVALUATION ENCOURAGES A STRONGER A LOCAL PRESENCE

An evaluation of USAID’s Somalia Legislative Strengthening Program in 2013 found that while “USAID’s partners had established good will and strong relationships with Somali counterparts,” the lack of constant presence of the partners in Somalia was reportedly presented a “challenge in the [implementing] partner’s ability to understand the nuances of the [local] partners to the extent possible” and recommended that “USAID make full-time presence in Mogadishu a requirement for future programs.”



Sheikh Aden Sheikh Mohamed, chair of the House Budget and Finance Committee, follows proceedings during a workshop in Nairobi.

cerpt from: Somalia Legislative Strengthening, 2010-2013 Somalia Legislative Strengthening, Elections and Political Process Program
 (http://pdf.usaid.gov/pdf_docs/PA00JSXC.pdf)



“The implementing partner’s (IP) staff has had a limited presence in Mogadishu. It describes its approach as ‘an incremental and phased approach to re-establishing an on-the-ground presence in Somalia that prioritizes the safety and security of its staff and partners.’ IP staff visited Mogadishu for meetings and to assess security context in March, May and October 2013. There is currently no IP Office in Mogadishu and no IP contact person in Mogadishu. IP trainings and seminars are conducted outside of Somalia. Interview respondents, even those who were otherwise very positive about the IP, said that this was a significant handicap to the IP’s capacity. ‘Imagine,’ said one, ‘imagine how effective they could be if they were here.’ Respondents noted that this policy reduced the IP’s grasp of on-the-ground political and contextual nuances; it limited the number of MPs they could train; and created incentives for MPs to participate only for the per diems and comfortable hotels. Evaluators concluded that a substantive presence in Mogadishu is a requirement for effective IP operations in the future. They also noted a significant divergence between the IP’s assessment of the security situation in Mogadishu and the costs of sufficient security and the assessments of USAID and other organizations active in Somalia. **Evaluators recommend that USAID make full-time presence in Mogadishu a requirement for future programs.**”

The evaluation USAID commissioned was to be used to inform USAID decisions regarding future Somalia governance programs. The Agency’s May 2014 solicitation for a new governance activity, known as the Strengthening Somalia Governance (SSG), was awarded in September 2014. It initiated project



Participants attending a Human Resources Workshop

activities in November 2014, with a new implementing partner, with in-country offices, consistent with the solicitation’s requirement that “the Contractor must be based in Somalia and must utilize local presence in Mogadishu.” In correspondence, USAID/East Africa staff reported that shifting the implementing partner’s base of operations “is making a huge difference in their operations.”

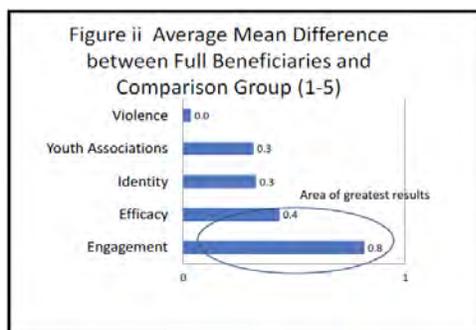
EVALUATIONS SUPPORT EFFORTS TO COUNTER VIOLENT EXTREMISM

In 2012, USAID/East Africa commissioned an evaluation of three activities focused on countering violent extremism in the region, and more specifically on violent extremism in Somalia and areas of Kenya with large Somali populations. The evaluation, which attempted to determine whether USAID's current violent extremism work is having the intended effect, found that activities did have an impact on people's stated perspectives and views on a range of issues related to violent extremism, in particular dampening support for violence and violent extremism.



Source: BBC News, Dec 5, 2012

Excerpt from: Mid-term Evaluation of Three Countering Violent Extremism Projects
http://pdf.usaid.gov/pdf_docs/PDACX479.pdf



“Beneficiaries of USAID’s Countering Violent Extremism (CVE) projects in East Africa have a demonstrated advantage over comparison groups on a host of variables known to be drivers of violent extremism. In a survey of almost 1,500 ethnic Somali youths in Somalia and Kenya administered in November and December 2012, full beneficiaries of three USAID CVE projects were compared to similar numbers of partial beneficiaries (mostly program drop outs or less involved participants) and a comparison group of non-beneficiaries.”

Upon receiving this evaluation report, USAID immediately began sharing the results, claiming the report as being a powerful tool in explaining the benefits of addressing this challenge prior to people becoming radicalized. They also used this report to justify USAID's role in these efforts to other U.S. Government Agencies, such as the Department of Defense; in July 2015 the White House released a statement that they intend to provide at least \$40 million in assistance related to countering violent extremism in East Africa.

USAID also took the report findings to other international donors such as Canada, Britain, and France, even engaging a center of excellence on countering violent extremism that is hosted by the Government of the United Arab Emirates in Abu Dhabi. According to USAID, that center of excellence has now taken some of the best practices described in the evaluation and is applying them in other areas of the globe experiencing similar challenges to those of Somalia.



Photo: CVE, project photo of Peace March in Eastleigh

EVALUATIONS AFFECTING USAID POLICIES AND STRATEGIES

In 2011, USAID/Afghanistan commissioned an evaluation of its Strategic Provincial Roads Southern and Eastern Afghanistan SPR-SEA Program. The evaluation found that cooperative agreements and grants are not as effective implementing mechanisms as contracts in terms of the levels of implementing partner accountability to USAID, particularly in regards to infrastructure activities.



Excerpt from: Final Report Evaluation of USAID's Strategic Provincial Roads Southern and Eastern Afghanistan SPR-SEA Program

(http://pdf.usaid.gov/pdf_docs/pdact897.pdf)

“USAID selected the Cooperative Agreement as the appropriate implementing mechanism...stating that [these agreements] carry certain disadvantages, such as (critically) less accountability. The alternative implementing mechanism – a Contract – has a different set of advantages including contractor responsiveness to USAID, high accountability, strong technical skills and a high USAID design input. But as with Cooperative Agreements, Contracts also have disadvantages such as mixed capacity building, minimal NGO innovation, greater USAID management and the possibility that higher overhead will equal higher program cost.” Based on these differences and what it learned about the project’s experiences, the evaluation team concluded that “the greater responsiveness found in contracts (typically through details reporting requirement) and accountability would have better served SPR.”

When this evaluation report was completed, and circulated in the Mission and in Washington, it garnered a good deal of attention and, while it is somewhat unusual for a single project evaluation to catalyze changes in Agency policy, the findings of the evaluation resonated with decision makers at the policy level. In 2013 USAID released a new operating policy, entitled “USAID Implementation of Construction Activities,” that mandates the use of contracts rather than other mechanisms for projects that involve construction.



Source: www.devex.com

USING IMPACT EVALUATION TO PREVENT CRIME IN CENTRAL AMERICA

In 2014, USAID completed a three-year impact evaluation it had commissioned to examine the effects of crime prevention efforts under the Central America Regional Security Initiative (CARSI). The evaluation, a randomized control trial, found that over the 2010 to 2014 period, a community-based policing approach to crime and violence prevention helped citizens feel safer, perceive less crime and murders, and express greater trust in police, with up to a 50 percent improvement from those communities that did not receive treatment.

Countries included in the Impact Evaluation



Source: http://pdf.usaid.gov/pdf_docs/PBAAB431.pdf

Excerpt from: Impact Evaluation of USAID’s Community-Based Crime and Prevention Approach in Central America
http://pdf.usaid.gov/pdf_docs/PBAAB431.pdf

“The study included 127 neighborhoods in El Salvador, Guatemala, Honduras, and Panama that met criteria for shared characteristics making them vulnerable to crime, and residents were surveyed to gather their perceptions, both before and after programming efforts, on a number of crime and violence related factors. Neighborhoods were split between treatment communities and control group communities in order to establish a counterfactual. In the end, the study found numerous positive results from community-based prevention efforts, including:

- 51% fewer residents reported awareness of murders
- 51% fewer residents reported being aware of extortion and blackmail
- 35% fewer residents reported avoiding dangerous areas because of fear of crime”

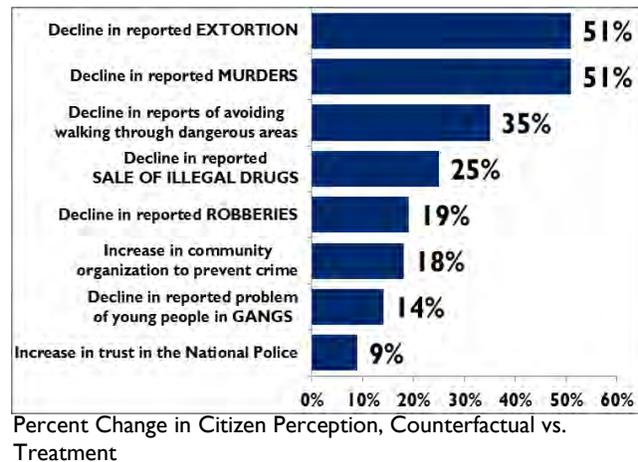


Photo: U.S. Department of State. Salvadoran police receive instructions during the start of presidential-election security operations in February 2014.

Recognizing the importance of these results, USAID staff disseminated the findings of the evaluation more widely than usual. In Central America, USAID presented findings at conferences, including one in Guatemala. In the United States, USAID hosted an event at the Woodrow Wilson Center to make the evaluation’s findings known, and it shared them with Congress and the vice president. USAID staff have characterized the CARSI model as one that has “changed the conversation” around the dangerous issue of crime and violence in Central America, moving its focus from law enforcement and suppression to crime prevention. In El Salvador, USAID has committed to rolling out the CARSI model in additional high crime areas its new (2013-2017) country strategy. Further, the government of El Salvador has adopted the model as national policy and plans to mandate municipal crime prevention committees. In the United States, USAID is using the evaluation’s results, on a “proof of concept” basis, to justify requests for additional funds to

further replicate this model.



EVALUATION RECOMMENDATIONS MOVE ARMENIA TOWARD ENERGY SECURITY

An end-of-project performance evaluation of USAID/Armenia’s Energy Security and Regional Integration (ESRI) Project recommended that USAID continue its efforts in the country’s energy sector, given its past success and strategic positioning in the country. More specifically, the evaluation team recommended the development of a least-cost power expansion plan for the country as well as supporting the development of policies, laws, and regulations to promote renewable energy technologies.



Excerpt from: Performance Evaluation of the Energy Security and Regional Integration (ESRI) Project
(http://pdf.usaid.gov/pdf_docs/PA00JR2M.pdf)

“For the nuclear task, all stakeholders agreed that the work produced by the contractor was valuable and, thus, relevant to their ongoing work. Since 2007, the GOA has been committed to building a new nuclear plant as a replacement to the existing Metsamor nuclear plant, which had been targeted for retirement in 2016. This commitment has remained in place for five years, even after the Fukushima disaster prompted some countries to abandon their nuclear power programs. However, it has been estimated that the current lead time to new NPP commissioning is nine years – and that is if detailed design begins immediately. Many possible developments could intercede during this time period that can result in abandonment of the project. Further, nuclear power plant capital costs have escalated rapidly – to roughly \$7 billion – since the original \$5 billion estimate for a 1,000 MW plant. Moreover, the last least-cost power sector expansion plan for Armenia is now six years old.”

According to USAID staff in Armenia, immediately following the conclusion of the evaluation they began to implement the recommendations, which they expressed as being “very valuable,” in the design and implementation of other energy projects or activities. The USAID CDCS states that “Evaluations [including ESRI] have informed the design of the strategy including approaches to cross-border initiatives, political processes, rule of law and health.” One recommendation in particular—to develop a least-cost energy development plan—has ushered in a new era for Armenia’s energy security as the newly developed plan is considered the foundation for Armenia’s new Energy Strategy, with senior Armenian government officials publicly thanking USAID for the quality of analyses and advice provided.



LEADS Project team presents the Least-Cost Energy Development plan to government, power company, bank, and donor representatives.

[This page is left intentionally blank.]

U.S. Agency for International Development
Bureau for Policy, Planning and Learning
Office of Learning, Evaluation and Research
Ronald Reagan Building
1300 Pennsylvania Avenue, NW
Washington, D.C. 20523