A Meta-Evaluation of Foreign Assistance Evaluations:

June 1, 2011

Office of the Director of U.S. Foreign Assistance

Department of State

Washington, DC

Acknowledgements

Many colleagues have helped in the preparation of the report. Krishna Kumar, the lead investigator, designed the study and discussed it with Peter Davis, who approved of the study. He then prepared the initial checklist and received valuable comments from Ruth Levine and her associates in the USAID Learning and Evaluation Office . Sarah Kohari helped in developing a coding scheme for the study. Kristina Hillman provided invaluable assistance in analyzing the data and preparing tables. John Eriksson, consultant, collaborated with Krishna Kumar in the final design, implementation and preparation of the study. The authors received thoughtful comments from Peter Davis and Richard Blue. Emily Crawford edited and formatted the report. The authors wish to convey their gratitude to all of them.

Krishna Kumar Ph.D	John Eriksson Ph.D
Office of the Director of Foreign Assistance	Consultant

Table of Contents

EXECUTIVE SUMMARY

Based on an analysis of the reports of 56 evaluations of U.S. Agency for International Development (USAID) projects issued in 2009, this study examines some of the concerns raised about the analytical and methodological quality of evaluations of foreign assistance projects and programs. The study makes a set of recommendations for improving both the rigor and usefulness of USAID and State evaluations. The reader is encouraged to read the main text, which includes quantitative data in tables and charts and narrative examples in boxes.

The focus of the study is solely on USAID evaluations. Although the Department of State has taken steps to promote evaluations, including the recent issuance of an evaluation policy, its share in the total number of evaluations conducted by the two agencies is negligible. Moreover, existing State evaluations are not presently located in one place. All completed USAID evaluation reports are expected to be sent to the USAID Development Experience Clearinghouse (DEC). USAID has also recently renewed its commitment to rigorous and relevant evaluations. For several reasons, only 56 evaluations could be reviewed of the 99 listed in FY2009 Performance Plan and Reports (PPRs) submitted to the Office of Director of U.S. Foreign Assistance.

After reviewing twenty evaluations, the lead investigator prepared a tentative list of the topics to be covered by the study and discussed it with Department of State and USAID colleagues. After considerable discussion, it was decided to focus on the following topics:

(a) The quality of the Statement of Work (SOW)
(b) Size and composition of evaluation teams
(c) Evaluation methodology
(d) Presentation of findings, conclusions, recommendations, and lessons as reflected in evaluation reports

A number of information items and evaluative questions were selected for each topic, which after pre-testing resulted in a checklist of 46 items that comprised the core of the evaluation (see Annex 1 for the checklist). Each evaluation report was carefully read by one of the two investigators, and checklist items were either coded or assessed with narrative comments. Evaluations were classified in two broad categories:

formative (or process) undertaken during the life of an intervention; and summative, conducted at or near its end. There were 30 formative and 26 summative evaluations.

Some limitations inherent in the present study should be noted. The investigators did not have the time and opportunity to interview the authors, managers, and other stakeholders of evaluations. As a result, they did not have in-depth knowledge of the context and expectations of the evaluations, the constraints that the evaluators faced, and their interactions with managers, host-country counterparts, and other stakeholders. Nor did they have any information about the utilization of evaluation findings. Many evaluation reports did not provide complete details for the variables covered in this study. Some did not provide any discussion of the methodology except a few brief lines, while others did not include relevant annexes. Therefore, the authors had to make inferences and estimates.

USAID evaluations were classified according to the five Program Objective categories for foreign assistance: Peace and Security, Democracy and Governance, Investing in People, Economic Growth, and Humanitarian Assistance. By far the largest number of evaluations was conducted in the third Program Objective, which covers many sectors, including disease control (e.g. HIV/AIDS, tuberculosis, malaria), maternal and child health, family planning, social services for vulnerable populations, education, and environment. The chief reasons for this preponderance include a long tradition of research and evaluation in these sectors, and relatively plentiful relevant data and program budget resources. The majority of evaluations were in Africa, which reflects the substantial volume of foreign assistance to African countries, a significant share of which is focused on the Investing in People Program Objective.

Statements of Work

The first main topic explored by the study is the evaluation's statement of work or SOW, intended to be a coherent plan for an evaluation. The SOW is prepared by the manager or Operating Unit commissioning the evaluation and covers such items as purpose and audience, evaluation questions, design and data collection methods, size and composition of evaluation team, level and duration of effort, and deliverables. This document is critical because a clear, comprehensive SOW ensures that the evaluation team will meet the information needs of the commissioning unit and reduces possible misunderstandings that might arise between evaluation managers and teams. The SOW

is also the core of a binding contract between USAID and the evaluation team or or contracted firm.

The main finding is that only 22 out of 56 evaluations included SOWs, which is low given that USAID guidelines for writing evaluation reports recommend that SOWs should be included. There could be several reasons for this low number, which are identified in the main report. However, the fact remains that SOWs were included in only 39 percent of the evaluations analyzed.

An important issue surrounding SOWs is whether adequate resources are budgeted for conducting evaluations. This issue is of paramount importance because it has a direct bearing on the nature and quality of data and findings. Evaluators of USAID projects and programs have often voiced their concern that they were not given sufficient time and resources to collect and analyze necessary data and to formulate empirically grounded recommendations. This study could only focus on some aspects of the adequacy of human resources (team size and time for fieldwork) because neither the SOWs nor the evaluation reports gave any information about the financial resources for conducting the field research nor the qualifications of evaluation teams. Using these indicators of resource adequacy relative to the demands implied by the SOW (and in some cases the political/security context in the country), only 19 (34 percent) of the evaluations were judged to have budgeted adequate resources, and 14 (25 percent) were judged to have budgeted only somewhat adequate resources, while it was not possible to assess resource adequacy for 23 (41 percent) of the evaluations (owing to lack of SOW and/or team size/duration). For example, one evaluation reported that time constraints "prevented the evaluation from being truly comprehensive" and resulted in the team being "unable to fully assess the quality of interventions."

Evaluation Teams

Along with SOWs and methodology, evaluation teams are critical to the quality, credibility, and utility of evaluations. The study focused on the size of teams and the participation of host-country evaluators and researchers. An "evaluation team" was defined as those professionals who have the requisite training and experience to formulate evaluation design; manage the collection of relevant and reliable data and evidence; identify and analyze findings; draw conclusions, lessons, and

recommendations; and prepare clear and objective reports. The required size of the evaluation team will depend on the scope and complexity of the evaluation and its duration.

USAID certainly does not follow a "one size fits all" approach to evaluation teams. While slightly less than half (45 percent) of the evaluation reports showed teams of 3 to 4 evaluators each, the rest of the teams varied from this modal size. There are outliers at the low end, with 8 evaluation teams (15 percent) comprised of only 1 to 2 evaluators, and at the high end, with 15 teams (27 percent) comprised of 5 to 9 evaluators each and 6 teams (12 percent) with 10 or more evaluators each. The review of evaluations suggests three factors that likely explain these large variations. Most important is the size and complexity of the project or program. A second factor is the conduct of surveys that obviously require more person power. Third, many evaluations listed the names of government and project officials who played an undefined role in the conduct of evaluation.

The study also examined the issue of the participation of host-country evaluators. Their participation is extremely desirable because they bring invaluable knowledge of a range of host-country factors relevant to the evaluation; they contribute to host-country ownership of the evaluation, lending it credibility; and their participation builds evaluation capacity within the country. The findings that emerge from the 53 evaluations containing data on teams show that, contrary to the general perception that host-country evaluators are not represented in evaluation teams, they in fact frequently participate in them. In only 40 percent of cases were host-country evaluators not engaged in the evaluations. In a few cases, they even led the evaluation teams. However, in general the host-country evaluators tend to play relatively secondary roles, ranging from assisting in designing research instruments, managing surveys, and translating key informant interviews, to providing feedback on draft reports.

Evaluation Methodology

There is a general perception in the evaluation and development community that foreign assistance evaluations lack methodological rigor. A common view is that evaluators often avoid using quasi-experimental and statistical designs, which involve the construction of a control/comparison group to examine what would have happened had the intervention not taken place.

Three aspects of methodology were examined in the study: evaluation designs, data collection methods, and conceptual framework. The main conclusion about evaluation designs to be drawn from the findings is that, contrary to popular perception, statistical and quasi-experimental designs are not uncommon in USAID evaluations. One in four evaluations conducted by USAID and its partners used them.

The reviewed evaluations used seven data collection methods: review of project and program documents, key informant interviews, site visits for direct observation, surveys, analysis of secondary data, focus group discussions, and group interviews. Apart from eight evaluations that relied solely on surveys, practically all evaluations used three or more data collection methods. This is important because multiple sources of data permit triangulation of findings and add to the credibility of conclusions and recommendations. Nearly half of the evaluations conducted surveys, thus utilizing quantitative data to supplement other methods. However, group interviews and focus group discussions were not widely used despite the fact they can generate useful information and insights. There was rarely any indication of the issues covered in interviews and discussions because most reports did not include interview protocols or lists of topics covered in focus group discussions. Nonetheless, nearly 75 percent of the reviewed evaluations were judged to have used appropriate data collection methods.

The study also investigated if evaluators examined the conceptual framework or the logic model underlying the evaluated development intervention in order to clarify the causal relationship between inputs, outputs, activities, outcomes, and impacts. Only 26 percent of the evaluations were found to have directly or indirectly used or referred to the underlying conceptual framework or model to evaluate performance or impacts.

Presentation of Data, Findings, Recommendations, and Lessons

A clear presentation of data, findings, recommendations, and lessons is critical for the credibility of evaluations and their use. Evidence- or data-based findings are the foundation of credible and useful evaluations. Findings may be quantitative or qualitative—ideally both—but they must be objective and representative. Their sources must be reliable and identifiable. Findings comprise the basis from which evaluation conclusions are drawn, which in turn provide the basis for setting out recommendations and drawing lessons learned.

The large majority of evaluation reports reviewed contained a clear presentation of the data and sources of the data. However, almost a third of the evaluations were deficient in some aspect of data presentation. The study findings are mixed regarding explanation of data limitations, which is important for credibility. About a third of evaluations provided only a partial or unclear explanation of data limitations and 40 percent provided no explanation at all. The inclusion of data collection instruments gives the reader an indication of the strengths and limitations of data collection. However, half the evaluations failed to include such instruments and 27 percent included some instruments. Only 23 percent included all instruments.

Recommendations drawn from evidence-based findings and conclusions are one of the hallmarks of evaluations that make them useful to stakeholders. In fact, of the 56 evaluations reviewed, 54 included recommendations. The study focused on whether recommendations followed from findings and whether they were actionable. On both these criteria, the evaluations fared well. In 75 percent of the reports (42), recommendations were judged to have followed logically from their findings; almost 25 percent followed only in part or not clearly from their findings. Over 50 percent of recommendations were actionable, while another 40 percent were judged as somewhat actionable.

Lessons are similar to recommendations in that they imply actions based on the findings and conclusions of evaluations, but they are not as immediate in their call for action as recommendations; nor are they necessarily addressed to specific stakeholder groups or action parties. The study found that only half the evaluations reviewed contained lessons. However, of these evaluations, less than half (16) contained fully adequate explanations of lessons.

Other important characteristics of evaluation reports were found to be satisfactory or better. Almost 80 percent of the 56 reports were written in clear language and style, and almost 90 percent contained executive summaries. Most answered all evaluation questions, although not always in a systematic order. Almost 90 percent of reviewed evaluation reports included a list of references.

Study Recommendations

Before outlining the study recommendations, one observation is warranted: the number of evaluations conducted by USAID is extremely small. It indicates that most of the interventions are not evaluated. Although USAID mandates that each major intervention should be evaluated at least once in its lifetime, the mandate seems not to have been followed.

1. **Issue: provision of sufficient time and person days for evaluation teams.**
   Recommendation: USAID/DOS should require that Operating Units budget sufficient time and person days for conducting summative evaluations and evaluations in war-torn societies.

2. **Issue: participation of host-country evaluators.**
   Recommendation: Guidelines should be issued which stipulate that all evaluations should have at least one local evaluator or researcher where feasible.

3. **Issue: involvement of government officials and project staff.**
   Recommendation: Guidelines should be issued to clarify the possible roles for host-country government officials and project/program staff in evaluations, and require that evaluations clearly spell out the nature and extent of their participation.

4. **Issue: documentation of qualitative/quantitative data collection instruments and data.**
   Recommendation: Guidelines should be issued on the presentation of qualitative and quantitative data in evaluation reports.

5. **Issue: discussion of conceptual framework/model**.
   <u>Recommendation:</u> Evaluators should be encouraged to examine and report on the underlying conceptual framework/model while evaluating the performance and impacts of an intervention.

6. **Issue: improving the quality of evaluation reports.**
   <u>Recommendation:</u> USAID should consider revising its guidelines and TIP for preparing evaluation reports, taking into consideration the above-mentioned shortcomings.

INTRODUCTION

1.1 Background and Purpose

Rigorous and timely evaluations are essential for improving the performance of foreign assistance projects and programs. Evaluations document organizational and programming experience, thereby facilitating learning by experience. They can also determine the cost-effectiveness of assistance programs, as well as the quality of their planning and implementation. Above all, evaluations can provide accountability information by determining if U.S. Government (USG) policies, programs, or partners have achieved what they were supposed to achieve, and if not, why not. Consequently, evaluation findings provide significant conclusions, lessons, and recommendations to different stakeholders: program managers, executive management, the Office of Management and Budget, the White House, Congress, the nongovernmental organization (NGO) community, and, above all, the American taxpayers. In the case of development assistance, evaluations should provide an accounting of projects and programs for similar categories of stakeholders in host countries, and provide lessons about sustainability for ongoing and new activities.

A list compiled by the Office of Director of U.S. Foreign Assistance shows that in 2009, USAID and the Department of State reported conducting about 100 evaluations. However, there is little information about the quality of these evaluations. A perception exists in some quarters that many, if not most, foreign assistance evaluations lack methodological rigor, that they are impressionistic rather than based on hard empirical evidence, and that their underlying conceptual framework is not fully articulated. Moreover, some host-country officials have argued that evaluations are conducted mainly by expatriate contractors, with minimal participation of the host-country researchers and evaluators. Some critics have also suggested that the recommendations made by evaluations are not derived from the data, and in many cases, are not actionable.

The objective of this study is to examine some of the concerns raised about the analytical and methodological quality of evaluations of foreign assistance projects and programs. Based on the findings, it also makes a set of recommendations for improving both the rigor and usefulness of the evaluations. Ideally, the findings and

recommendations of the study will facilitate a conversation on this topic, and help the Department of State and USAID in developing new or revised guidance for evaluations.

This study has focused solely on USAID evaluations for several reasons. Although the Department of State[1] has recently issued an evaluation policy and has taken steps to promote evaluations, its share in the total number of evaluations is negligible, and existing evaluations are not located in one place. Practically all major evaluations are done by USAID, which has an institutionalized evaluation system with guidelines for conducting evaluations. Its completed evaluation reports should be sent to the USAID Development Experience Clearinghouse (DEC), which is a depository of all its reports. Moreover, USAID has recently renewed its commitment to rigorous and relevant evaluations and has launched a major initiative to improve their quality and quantity, including the issuance of a new evaluation policy.[2]

1.2 Study Methodology

As a first step, the lead investigator carefully examined over twenty recent evaluations in the five USAID Program Objective areas to get an indication of their contents and to identify pertinent issues that need to be examined. Based on the review and his own experience, he prepared a tentative list of the topics to be covered and discussed it with colleagues in the Department of State and USAID. After considerable discussion, the investigators decided to focus on the following topics:

(a) The quality of the Statement of Work (SOW)

(b) Size and composition of evaluation teams

(c) Evaluation methodology

(d) Presentation of findings, conclusions, recommendations, and lessons

A number of information items and questions were selected for each topic, and a draft checklist with both informative and evaluative questions was prepared and

---

[1] The State Department issued its Program Evaluation Policy in 2010, which specifies evaluation requirements for its bureaus and offices.
[2] USAID's Bureau of Policy, Planning and Learning issued a new and comprehensive document entitled "USAID Evaluation Policy" in 2011.

pretested. Pretesting led to a slight revision of the checklist. The final version of the checklist is included as Annex 1 to this report.

The original plan was to focus on the evaluations listed in the FY2009 Performance Plan and Reports (PPRs) submitted to the Office of Director of U.S. Foreign Assistance. Although it listed 99 evaluations for USAID, that number was not correct. Many PPRs did not provide full titles of the evaluations, and it was therefore difficult to trace them. Some Operating Units listed evaluations that were not completed, while others did not send the reports to DEC. Under these circumstances, the lead investigator had no alternative but to focus on those evaluations completed in FY 2009 that were also available in DEC. The DEC listing was also not without problems. A large number of the studies listed as evaluations were sector or sub-sector assessments undertaken for planning purposes. Some evaluations were listed twice: both draft and final versions. In quite a few cases, listed evaluations were simply reviews, audits, and Government Accountability Office (GAO) evaluations. After careful examination, the lead investigator identified a total of 56 evaluations, which constituted the universe for the study.[3]

Each evaluation report was carefully read and its elements rated -by either the lead investigator or the consultant, and coded in Microsoft Excel files. (This rating was performed by two highly credentialed and experienced evaluators, thus constituting an expert review). In addition to coding the items covered in the checklist, both investigators took extensive notes that were also used in the preparation of the report. The size of the evaluation reports ranged from 40 to 200 pages. Evaluations were classified in two broad categories: formative (or process) and summative. The formative evaluations refer to the evaluations conducted during the life of a project or program, while summative evaluations are those done at or near the end of an intervention. There were 30 formative and 26 summative evaluations.

Table 1.1, which gives details of evaluations by Program Objective, indicates that by far the largest number of evaluations was conducted in the Program Objective of Investing in People. This Objective covers many sectors, including disease control (e.g. HIV/AIDS, tuberculosis, malaria), maternal and child health, family planning, social services for vulnerable pollutions, education, and environment. One plausible

---

[3] Two additional evaluations (one in French and the other in Spanish) had to be excluded from the list, as neither the lead investigator nor the consultant is proficient in these languages.

explanation for the preponderance of evaluations in this Objective is that many of these sectors have a long tradition of basic and applied research and evaluation. Relevant data are also more plentiful than is often the case in other sectors. Above all, the budgetary allocations for this Program Objective are the highest, with the exception of the Peace and Security Objective.

Table 1.1

Evaluations by Program Objective

Program Objective	Process	Summative	Total	Percentage
Peace and Security	2	2	4	7.1%
Democracy and Governance	2	4	6	10.7%
Investing in People	22	16	38	67.9%
Economic Growth	3	3	6	10.7%
Humanitarian Assistance	1	1	2	3.6%
Total	**30**	**26**	**56**	**100.0%**

Table 1.2 presents data about evaluations conducted in different regions. It indicates that the largest number of evaluations was conducted in Africa, as Africa includes the largest number of countries receiving U.S. foreign assistance. Moreover, assistance to Africa tends to be focused in the agriculture, health, and education sectors, which are most often the subject of evaluations.

Table 1.2

Evaluations in Different Regions

Regions	Number	Percentage
Africa	29	51.8%
Asia	12	21.4%
Europe and Eurasia	4	7.4%
Latin America	5	8.9%
Middle East	6	10.7%
Total	**56**	**99.9%**

1.3 Limitations of the Data and Information

A few limitations of the present study should be noted here. First and foremost is that it is based solely on review of evaluation reports. The investigators did not have time and opportunity to interview the authors, managers, and other stakeholders of evaluations. As a result, they did not have in-depth knowledge of the context of the evaluations, the expectations from evaluations, the constraints that the evaluators faced, and their interactions with managers, host-country counterparts, and other stakeholders. Neither did they have any information about the utilization of evaluation findings.

Second, as previously suggested, USAID is likely to have conducted more evaluations than are available in the DEC. Past experience indicates that despite the official requirement, Operating Units do not always send their evaluations to the DEC. Some Operating Units like to keep reports of internal evaluations to themselves, as they are not formal, not balanced, embarrassing to the commissioning organization, or not well-written. Still others are simply negligent. Consequently, some biases in the evaluations selected cannot be ruled out.

Third, many evaluation reports did not provide complete details for the variables covered in the study. Some did not provide any discussion of the methodology except a few brief lines, while others did not include relevant annexes. Therefore, authors had to make inferences. For example, when reports did not mention the size of the evaluation team, they looked at SOWs, acknowledgements, and title pages to try to estimate their number. Similarly, not all reports mentioned the precise number of expatriate and local researchers. The investigators had to examine SOWs, the names of the evaluators, or their institutional affiliation to determine their origin.

Finally, because many variables included in the checklist required subjective judgments, there is always a possibility that the two reviewers might have used different criteria. However, every effort was made to avoid such bias. The investigators discussed each item in detail, compared each other's data sets to discern any biases before merging them, and made suitable revisions when necessary. Therefore, while some bias cannot be ruled out, it is likely to insignificant.

<u>**EVALUATION STATEMENTS OF WORK**</u>

A statement of work (SOW) is essentially a blue print: a coherent plan for an evaluation. SOWs are prepared by Operating Units commissioning evaluations and cover items such as the purpose and audience, evaluation questions, designs and data collection methods, size and composition of evaluation team, the level and duration of effort, and deliverables. A well-written and thoughtful SOW ensures that the evaluation team will meet the information needs of the commissioning Operating Unit, and reduces possible misunderstandings that might arise between evaluation managers and evaluation teams. It also can become binding contract between USAID and the evaluators or the contracted firm. Therefore the first topic which the study examined was evaluation SOWs.

2.1 Contents and Quality of SOWs

Only 22 out of 56 evaluations included SOWs, despite the fact that the USAID guidelines for writing evaluation reports recommends that SOWs should be included in evaluation reports. Several explanations can be given for the non-inclusion of SOWs. Many of these evaluations were commissioned by partner organizations, and therefore it is quite possible that these organizations did not prepare detailed SOWs that could be included in evaluation reports, or that the evaluators were not aware that SOWs should be included. In a few cases, the evaluations were simply surveys, not written as formal evaluation reports. Moreover, even when SOWs were included in evaluations, they were not always attached to the main report submitted to DEC. Notwithstanding these exceptions, many evaluations commissioned by USAID directly or indirectly included SOWs. Even when they were not attached, they were referred to in the main body of the report.

The study also examined the contents of the available SOWs. Table 2.1 indicates that an overwhelming majority of SOWs included the purpose of the evaluation, evaluation questions, data-collection methods, and levels of effort. Only one out of four SOWs directly mentioned the audience of the evaluation, but this last variable is not required by the USAID guidance on SOWs.

Table 2.1

Items Covered in SOWs

Items	Percentage
Purpose	91. 1%
Audience	30.4%
Data Collection Methods	85.7%
Benchmarks	32.1%

The study also looked at the quality of SOWs. This assessment was based on three considerations: coverage of all relevant items, clear presentation of key evaluation questions, and language and style. The findings indicate that more than half were completely satisfactory, and the remaining ones partly satisfactory but could be improved.

2.2 Budgeting of Adequate Resources

An important issue surrounding SOWs is whether they budgeted adequate resources for conducting evaluations. This issue is of paramount importance because it has a direct bearing on the nature and quality of data and findings. Evaluators of USAID projects and programs have often voiced their concern that they were not given sufficient time and resources to collect and analyze necessary data and to formulate empirically grounded recommendations. While it is natural for evaluators to ask for more time and resources, the fact remains that gathering and analysis of data in host countries usually require more time and resources than are budgeted.[4] This study could only focus on the adequacy of human resources (team size and time for fieldwork) because neither SOWs nor evaluation reports gave any information about the financial resources for conducting field research.

To answer this question, the investigators considered several variables such as

---

[4] It should be noted that a study entitled "Beyond Success Stories: Monitoring and Evaluation for Foreign Assistance" by Richard Blue, Cynthia Clapp-Wincek and Holly Benner (2009: p iii noted that "SOWS for evaluation often reflect a mismatch between evaluation questions that must be answered and methodology, budget and time frame given for an evaluation."

the size of the evaluation team, the duration of fieldwork, the nature of project/program activities and their coverage, the political and economic environment of the country, and the nature and scope of evaluation questions. All these variables are related to the question of adequacy of resources. For example, if a project has multiple activities and covers a large country, it would require more person power than the one which has fewer activities and operates in a limited area. An evaluation that seeks to measure/assess outcomes and impacts would need more professional person power than an evaluation that primarily focused on performance, because performance data can be more easily collected than outcome and impact data. The findings are presented in Chart 2.1, and Box 2.1 gives selected examples from the evaluations examined.

Chart 2.1
Provision of Adequate Person Power Resources
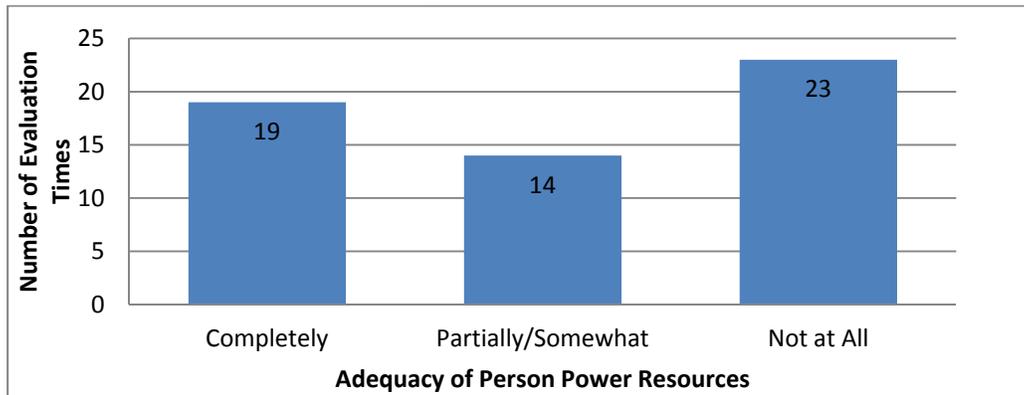
Box 2.1 Adequacy of Team Size and Times for Fieldwork

"Final Evaluation: The Private Sector Program in Ethiopia" was a summative evaluation with a team of two expatriate evaluators. Although thes team spent 23 days in the country and focused on public-private collaboration in the health sector, it was a small team for what appeared to be a relatively ambitious SOW for a summative evaluation in a large and populous country (no SOW was attached to the report).

"Assessment of USAID's Child Welfare Programs in Russia" had a team consisting of two expatriate evaluators. While the SOW of the team was not attached, gaps in the evaluation report make it clear that either the team was too small and/or the time in country of 16 days was insufficient to cover this vast and complex subject. The evaluation had both summative and process aspects.

"Teaching in the Window of Hope" had two expatriate evaluators for this relatively limited process evaluation in Zambia. Given the evaluation scope and the limited methodology, the 17 days in-country may have just been adequate.

2.3 Examples of Inadequate Budgeting of Evaluations

The investigators examined evaluation reports in detail to get additional insights on the complex issue of budgeting for evaluations. This examination suggested two categories of evaluations for which budgeted person power tended to be inadequate. The first category consisted of evaluations that involve assessing the outcomes and relevance of a project or program. In such cases, evaluators had to collect fresh quantitative or qualitative data, which are not easily available. This created problems for evaluators and might have compromised the reliability and validity of evaluation findings. Box 2.2 gives an example of an evaluation for an Ethiopia project (see basic characteristics in Box 2.1) for which the evaluation team noted that time constraints "prevented the evaluation from being truly comprehensive."

Box 2.2 *Report of Midterm Evaluation of Private Sector Program in Ethiopia for Public-Private Sector Collaboration in the Health Sector -(2009)*

The evaluation team noted that there were three principal constraints on the team's efforts, each of which may have resulted in gaps preventing the evaluation from being truly comprehensive. These were as follows:

- The time allocated to document review—The team was assigned only three days to review and analyze the significant amount of documentation associated with the project. (The team had to review more than 150 documents.) "The team believes that the assessment would have benefited from more time being allocated for this. While the team expended every possible effort to ensure that the assessment was truly evidence-based, it acknowledges that it may have missed a number of key points or issues."

- Time allocated for field visits—"While the team appreciated the time and effort of the project staff to arrange and coordinate the field visit to Amhara region, it felt that the brief time (4 days) available for field visits lessened its ability to fully assess the progress achieved by the project throughout the three regions."

- Quality of interventions—Because of the paucity of time and lack of additional expertise on specific issues, the team was unable to fully assess the quality of interventions.

Box 2.3 gives another example. In this case the USAID Office of Transition Initiatives (OTI) had commissioned an evaluation of its three year old program in Nepal, and its SOW listed lengthy, sweeping and complex sets of questions which cannot be realistically answered by a team of three expatriates who spent only 26 days

in Nepal. Excerpts from some of the evaluation questions in the SOW are shown in the box.

The second category of evaluations for which budgeting tended to be inadequate consisted of evaluations conducted in war-torn societies such as Afghanistan and Iraq. As compared to relatively peaceful societies, it takes much more time and effort to gather and analyze information. Often the secondary data are not available due to poor institutional infrastructure, and most of the information had to be gathered by evaluators themselves. Moreover, transportation creates additional problems, as it is often difficult to visit project sites without security support. Planned visits are many times cancelled for security and logistic reasons. It also takes time for evaluators to build trust with the respondents. A review of evaluation reports indicated that the vagaries of data collection in non-secure environments were often underestimated in SOWs. A good example is provided by the evaluation of the USAID's "Community Stabilization Program" in Iraq, which had an estimated outlay of $80 million.

Box 2.4 *Community Stabilization Program in Iraq*

 In 2005, the USG Interagency in Iraq created a nonmilitary program to complement security and stabilization efforts. USAID launched the Community Stabilization Program (CSP) to meet this urgent need. The project was designed to reduce the incentives for young people to participate through creating employment and engaging them in development activities. This $80 million program undertook a variety of activities in 18 geographic locations throughout Iraq. The program had four major components: short-term employment generation through short-term development initiatives; long-term job creation by supporting micro, small, and medium-sized private enterprises; education and vocational training; and sport and cultural activities to engage the youthful population.

        The objectives of the evaluation were to answer the following questions: Did CSP achieve overall program objectives to reduce the incentives for participating in the insurgency? Did CSP employment and engagement activities for young men contribute toward short- and long-term stability? How effective was CSP in reducing the number of security incidences in different localities? How effective was CSP in increasing Iraqi citizens' acceptance of the legitimacy of the Government of Iraq? To what extent did CSP support the surge and overall coalition efforts in Iraq, including secondary impacts of saving U.S. soldiers' lives? What positive and negative lessons were learned about the CSP project design and implementation that could be applied toward future USAID or USG projects in other conflict or post-conflict environments?
        USAID hired an international consulting firm to conduct an evaluation of the program. The evaluation team, which consisted of seven evaluators, spent only seven weeks in a highly hostile environment. The team noted:

   "The duration provided for this evaluation… was too short to provide a   comprehensive
      evaluation of pertinent issues, a number of which were
      identified during the conduct of key informant interviews or as a result
      of polling. To provide a comprehensive analysis of a program of this size
      and importance and its integration with other USG stakeholders in
      counterinsurgency, the evaluation team believes an in-depth study is in
      order, which should likely take 4-6 months." (IBTC; 2009, p. 44)

While the evaluation reports could not shed any light on why adequate resources were not budgeted, the investigators' own experience suggests several explanations. The most likely is that project/program planning documents usually do not allocate sufficient resources to conduct empirically grounded evaluations. Although USAID has taken steps to correct the situation by issuing new policy and guidelines, it will take time before these measures bear fruit. Another explanation is that evaluation managers tend to have limited knowledge and little or no experience in conducting evaluative research, and therefore underestimate the time and resources needed for it. Still another explanation is that evaluation managers tend to be risk averse and therefore want to include many questions without any regard to the constraints of data collection.

A common practice is that the draft SOW is circulated among interested parties, who generally add new questions rather than suggesting the deleting the existing questions. Whatever the reason, the quality suffers.

2.4 Conclusions

Three major conclusions emerge from the above data and findings. First, a majority (almost 60 percent) of evaluation reports submitted to DEC do not include SOWs, which indicates that the guidance provided in the TIP for writing evaluation reports is not followed. This is hardly satisfactory. Second, most SOWs included in evaluation reports are generally comprehensive and cover all important topics, which indicates that USAID evaluation managers have become quite proficient in preparing SOWs and writing clear SOWs. Third, and which is rather disturbing, is that SOWs tend to be overly ambitious and/or that insufficient time is budgeted for undertaking rigorous data collection and analysis. This is particularly true in the case of summative evaluations, which examine the impacts of development interventions and the evaluations that are conducted in high-threat environments.

EVALUATION TEAMS

This chapter focuses on evaluation teams, particularly their size, and the participation of host-country evaluators and researchers. For the purposes of this study, an "evaluation team" is defined as only those professionals who have the requisite training and experience to formulate evaluation design; manage the collection of relevant and reliable data and evidence; identify and analyze findings; draw conclusions, lessons, and recommendations; and prepare clear and objective reports.[5] It does not include auxiliary team members such as data collectors, drivers, clerical and administrative help, and other support staff. The required size of the evaluation team will depend on the scope and complexity of the evaluation and its duration. Wherever possible, an evaluation team should include two evaluators in order to crosscheck observations from key informant and group interviews, compare emerging findings, and draw conclusions, lessons, and recommendations.

3.1 Size of Evaluation Teams

Chart 3.1 presents data about the size of evaluation teams for the 53 reports for which information on evaluation team members was available.

---

[5] The qualifications of the evaluation team leader are particularly critical. Unfortunately the evaluation reports did not provide sufficient evidence to assess the qualifications of team leaders.

Chart 3.1
Evaluation Team Size
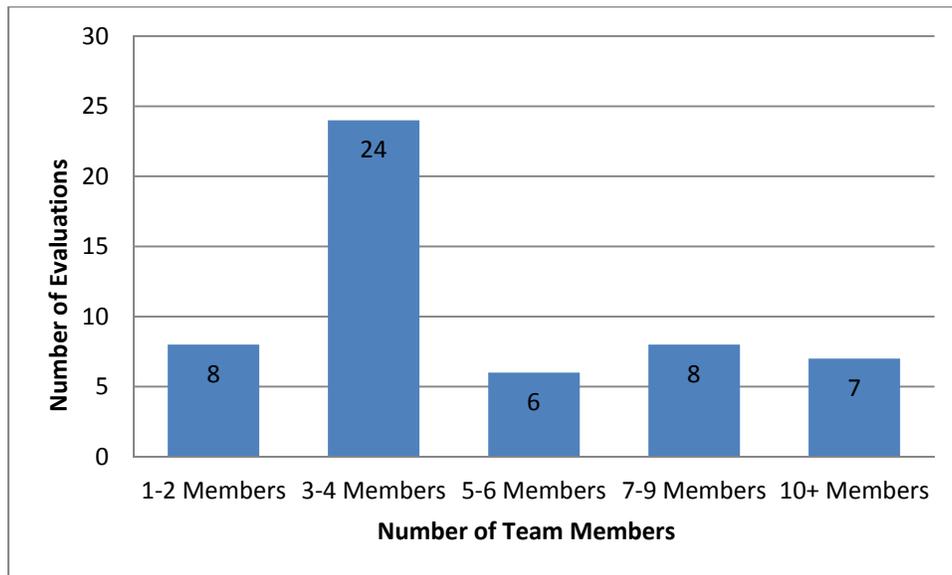
The chart shows that USAID certainly does not follow a "one size fits all" approach. While slightly less than half of the evaluations (45 percent) had teams of 3 to 4 evaluators each, the rest of the teams varied from this modal size. There are outliers at the low end, with 8 evaluation teams (15 percent) comprised of only 1 to 2 evaluators, and at the high end, with 15 teams (28 percent) comprised of 5 to 9 evaluators each and 6 teams (11 percent) with 10 or more evaluators each.

A word of caution is necessary about the quality of the data. Because evaluation reports did not follow a consistent approach to reporting, the investigators faced a major problem in identifying the size of evaluation teams. Many reported them in different places, ranging from the title page to an acknowledgements section, to the purpose or methodology sections, or to a special annex at the end of the report. Moreover, evaluation reports often did not distinguish evaluators from those who provided support but were not evaluators themselves. Government officials were often listed as part of evaluation teams but their role as evaluators was not clear. Similarly, USAID staff or staffs of implementing agencies such as NGOs were listed as evaluation team members, but it was not clear what role they played. Although investigators carefully reviewed each evaluation to report only those team members who, from their titles and positions, seemed to play an evaluation role as described above, some errors cannot be ruled out.

What explains such large variation in the size of the evaluation teams? A review of evaluations points to three factors. First and most important is the size and complexity of the intervention. As expected, relatively small projects did not require a large team and vice versa. The second factor is the conduct of surveys. The evaluation teams that conducted surveys obviously required more person power. Often, such evaluations listed researchers who managed or conducted them as evaluators. Although the study tried to exclude enumerators while calculating the size of evaluation teams, it could not exclude the names of researchers who managed them. Third, as indicated earlier, many evaluations listed the names of the government and project officials who played some role in the conduct of evaluations. Box 3.1 illustrates these considerations by mentioning several of the "outlier" evaluation teams with relatively small and relatively large sizes.

Box 3.1 Selected Examples of Evaluations of Varying Evaluator Team Size

One to two evaluators per evaluation

- *"Agricultural Recovery in the Commune of Gros-Morne Artibonite Department"* (Haiti) was a limited summative evaluation of a one-year project undertaken by only one expatriate and no host-country evaluators. The evaluation employed a statistical design and sampled 10 percent of project beneficiaries. Given the limited nature of the project and the evaluation, the very small team size was probably appropriate, although it would have been helpful to include a host-country evaluator.

- *"Sustaining the Lives and Dignity of IDPs in Purnea District – Bihar"* (India) involved only one evaluator. This was a very small project managed by a local NGO. The project had detailed information about the project beneficiaries, which provided useful and relevant input to the evaluation.

- *"Impact Assessment Report: Liberia Energy Assistance Program (LEAP)"* involved two expatriate evaluators. The evaluators spent a relatively short time in the country, but the scope of the evaluation was extremely limited, involving surveys of four stakeholder groups on their experience with prepaid urban electric meters.

Ten or more evaluators per evaluation

- *"Provincial Strengthening in Northern Afghanistan: Capacity Building and Innovation to Support Basic Package of Health Services"* included thirteen expatriate evaluators. This summative evaluation entailed 24 days in-country. A major factor which explains the large number of expatriates is that most of the expatriates were working on the project and the

---

concerned government agencies were drafted for the evaluation.

- *"Synergy and Action for Nutrition+ (SAN+) - Child Survival Project – Koulikoro Region, Mali"* was a summative evaluation led by an expatriate evaluator plus two other expatriate evaluators and 14 host-country evaluators. This included a number of Government officials for whom it was impossible to identify their roles in the evaluation. Not included are an additional 12 Malians, who played obvious non-evaluator roles, including data collection, administration, and other support.

- *"Evaluation of the Home Based Management of Malaria Strategy in Rwanda 2008"* involved nine (estimated) Rwandan and three expatriate evaluators. The 12 team members for this process evaluation were drawn from the Ministry of Health, "USAID/BASICS," and a group called "Strengthening Pharmaceutical Systems," which may partly explain the relatively large size of the team. The roles of the Rwandan team members are not clear.

     Does the USAID Program Objective or sector make a difference to the size of evaluation teams? Given the classification of reviewed evaluations into five Program Objectives, it is possible to determine evaluator team size by Objective. Objective 3, "Investing in People," dominates each team size category, including the largest category: ten or more members. Six of seven evaluator teams with ten or more members fall into Objective 3. Details are given in Table 3.1 below. Even though averages mask diversity, for purposes of inter-sectoral comparisons they can be useful, and are shown in the last column of the table.

Table 3.1

Evaluation Team Size by Program Objective Areas

Objective	1-2 evaluators	3-4 evaluators	5-6 evaluators	7-9 evaluators	10 + evaluators	Average no. of evaluators
1. Peace and Security	---	2	---	2	---	6.7
2. Governing Justly and Democratically	1	4	1	---	---	3.3
3. Investing in People	4	15	5	6	6	5.9
4. Economic Growth	2	2	---	---	1	4.0
5. Humanitarian Assistance	1	1	---	---		2.0
Total No. of Evaluations	8	24	6	8	7	

3.2 Participation of Host-Country Evaluators/Researchers

The study also examined the issue of the participation of host-country evaluators. There is a broad consensus among the evaluation community that the active participation of host-country evaluators is extremely desirable for three reasons. First, they bring invaluable knowledge of a range of factors relevant to the evaluation, including familiarity with data sources and their reliability; knowledge of important local stakeholders and the best way to make contacts; and awareness of the local political and cultural contexts, including language, that can improve the relevance and utility of data-collection instruments. Second, their participation contributes to host-country ownership of findings, conclusions, lessons, and recommendations, thereby lending credibility to the evaluation in the political and economic environments of the host country. This helps to ensure sustainability to those USAID-assisted activities intended to continue beyond the life of project with host-country resources. Finally,

their participation builds evaluation capacity within the country. As in other fields, "learning by doing" in evaluation is as important as formal training, if not more important. Sustainability of development depends in part on developing a "culture of evaluation" in society—not just among professional evaluators, but also among policymakers, politicians, and the public. However, caution is necessary to assure that host-country evaluators have the necessary qualifications and independence.

Table 3.2 gives details about the participation of the host-country evaluators. It divides evaluations in three categories: those conducted only by expatriates, those conducted only by host-country evaluators, and those conducted jointly.

Table 3.2
Composition of the Evaluation Teams

| Composition of Teams	Number of Teams	Percentage
Expatriates Only	21	39.6
Host-Country Only	2	3.8
Joint Expatriate and Host-Country	30	56.6
Total	**53**	**100%**

The table shows that close to forty percent of evaluations were exclusively done by expatriates. This is problematic. During the past decade, availability of local evaluators and researchers who can actively participate in evaluations has increased all over the developing world, and their participation would not only improve the quality of evaluations but may also reduce the overall costs.[6]

Chart 3.2 further explores the issue by dividing evaluation teams in three categories: evaluation teams in which the majority or all evaluators were expatriates, those in which the majority or all were host-country evaluators, and those in which the number of expatriates and host-country evaluators was equal.

---

[6] Depending on the nature of the SOW, including the budget.

Chart 3.2
Joint Evaluation Distribution
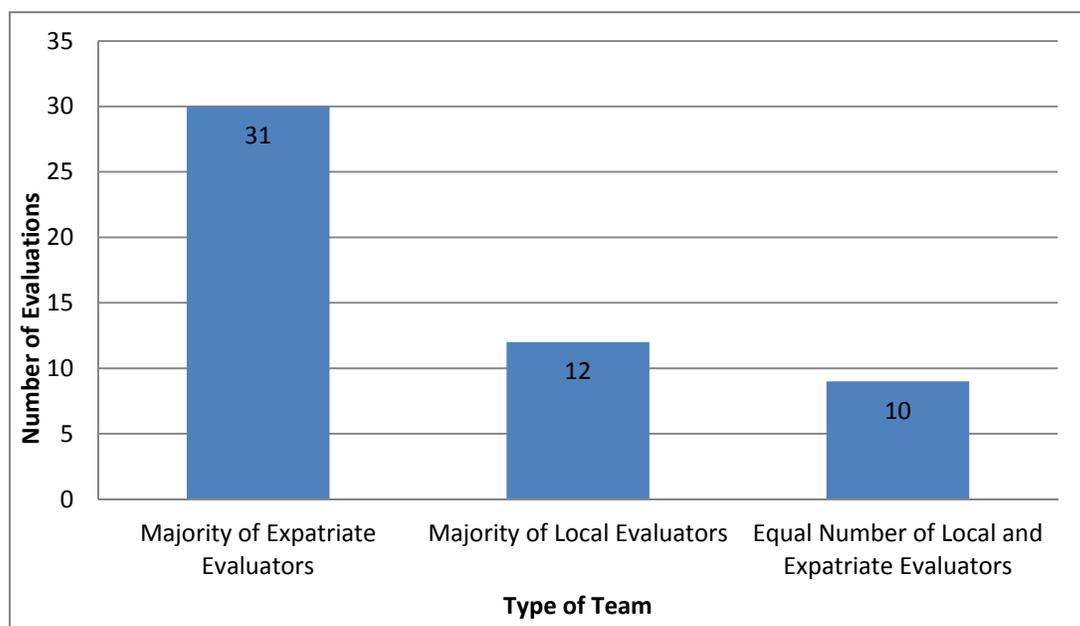
To reap the benefits of participation of host-country evaluators, it is important that they play meaningful professional roles. These could include evaluation design, implementation, report preparation, dissemination, and utilization (at least having an opportunity to indicate their views about utilization). Unfortunately, it was not possible in most cases to determine from the reports the specific roles the host-country evaluators played. Some examples of reports that showed host-country evaluators are provided in Box 3.2, but in many cases it was not possible to determine the *kinds* of roles they played.

Box 3.2 The Roles of Host-Country Evaluators/Researchers

"International Human Rights Law Outreach Program." This is an example of an evaluation led by a host-country evaluator. Since the evaluation was led by an Egyptian evaluator, Egyptian research assistants may have been employed to gather data and information, but this not specified.

"Summary: PEPFAR Public Health Evaluation – Care and Support– Phase I: Kenya." The report identifies ten evaluators but it was possible to conclude that only three were likely Kenyan and six likely expatriate evaluators, since the tenth evaluator participated in at least two PEPFAR evaluations. All ten individuals are shown as "authors," and no specific roles (e.g., team leader) or institutional affiliations are given. The investigators estimated origins by examining the surnames of team members. This was a first phase

---

process evaluation that was to be followed by an impact evaluation in a second phase.

"Comparative Study of Recent Literacy Programs Piloted in Malawi and Mid-Term Evaluation of the Beginning Literacy Program of Malawi." This combined research study and evaluation was conducted by a team of seven Malawian researchers led by Dr. Grace Chiuye of the Centre for Educational Research and Training, University of Malawi. However, the report identifies six of the seven team members as "research assistants." A strict interpretation of "professional evaluator" would imply only one team member, the leader, Dr. Chiuye, as an evaluator.

"The evaluation of "Synergy and Action for Nutrition+ (SAN+) - Child Survival Project – Koulikoro Region, MALI" reported a relatively large team of 14, with 3 expatriate evaluators, including the idenitified team leader, and 11 Malian evaluators. Some of the latter included Government officials and it was not possible to tell whether they had performed evaluation or support/facilitation roles.

"PEPFAR Public Health Evaluation Care and Support." This report listed a team of eleven evaluators, four likely Ugandan and six likely expatriate. The nationality of the eleventh evaluator could not be determined. As in the Kenya PEPFAR evaluation, all eleven individuals are shown as "authors" with no specific roles or institutional affiliations shown.

The roles of host-country evaluators were obvious in those cases in which they acted as team leaders, as in the Malawi and Egypt examples in Box 3.2. However, the Malawi evaluation still raises the question of who should be identified as evaluators, since six members of the seven-member team are identified by the report as "research assistants." A more common pattern is for a team consisting mainly of host-country evaluators to be led by an expatriate evaluator, as was the case for the Mali evaluation described in Box 3.2.

3.3 Conclusions

Three observations can be made. First, the size of evaluation teams significantly varies; slightly less than half of the evaluations have a 3-4 members per team. This means that USAID does not follow a "one size fits all" approach to the conduct of evaluations. Second, contrary to the general perception that host-country evaluators are not represented in evaluation teams, the data show that they frequently participate in them. In only 40 percent of cases, host-country evaluators were not engaged in evaluations. In a few cases, they even led the evaluation teams. Third, host-country evaluators/researchers tend to play relatively secondary roles. Such roles range from

assisting in designing research instruments, managing surveys, translating key informant interviews, and providing feedback on draft reports.[7]

---

[7] Assuming they have the requisite technical qualifications, unless English is a first language or they have mastered considerable proficiency in writing English, it may not be practical to use host-country evaluators for significant report writing or they may require intensive guidance in clear writing and formatting.

EVALUATION METHODOLOGY

A major issue that has invited considerable discussion in recent years concerns evaluation methodology. There is a general perception in the evaluation and development community that foreign assistance evaluations lack methodological rigor. Evaluators often avoid using quasi-experimental and statistical designs, which involve the construction of a control/comparison group to examine what would have happened had the intervention not taken place. Instead, the assessment of the outcomes and impacts of development interventions is usually based on anecdotal evidence drawn from key informant interviews and other qualitative information that cannot be verified. An influential report entitled "When We Will Ever Learn" by the Center for Global Development (2006: p.1) noted: "…it is deeply disappointing to recognize that we know very little about the net impact of most of these social programs." It has made a powerful case for rigorous evaluation designs that can measure the net impacts of development interventions. Therefore, the present study examined evaluation designs and data collection methods used in evaluations.

4.1 Evaluation Designs

The study originally categorized evaluation designs in four categories: case study, multi-case study, quasi-experimental, and statistical. However, the evaluations using multi-case study and quasi-experimental designs were few, and the investigators decided to group them with the other two designs. Multi-case study design was combined with the single case study, and the quasi-experimental design was combined with the statistical design. A case study design situates a project or program in its distinctive social, political, and institutional environment and evaluates its performance and impacts by using both qualitative and quantitative data. Quasi-experimental and statistical designs, on the other hand, seek to isolate the effects of an intervention from exogenous factors. Using control groups or statistical controls, they endeavor to measure the net outputs, outcomes, and effects of the intervention. Boxes 4.1 and 4.2 give examples of both types of designs.

Box 4.1 Examples of Statistical/Quasi-Experimental Designs

"Provincial Strengthening in Northern Afghanistan: Capacity Building and Innovation to Support the Basic Package of Health Services" is a good example of an evaluation that used statistical data to examine the overall outputs and outcomes of the project. This was largely possible because the evaluators had access to monitoring data as well as the data gathered by the host-country government and another project in the region. The evaluation did not solely rely on statistical data alone, and conducted site visits and key informant interviews.

The process evaluation of *"Extending Social Insurance to Informal Sectors Workers in Nicaragua via Microfinance Institutions"* is solely based on a randomized sample survey. The objectives of the evaluation were to assess and compare the effectiveness of delivering health insurance to the informal sector with and without the assistance of microfinance institutions, and to measure the impact of insurance on access to affordable health care, including reproductive health and family planning, for the targeted population. The study introduced an experimental component in which individuals were allocated health insurance subsidies of varying amounts by a "lottery." Individuals were also randomly assigned to sign up for the insurance.

"The Impact of the Second National Kenya Civic Education Programs on Democratic Attitudes, Values, and Behaviors" is also an example of an impact evaluation based on rigorous statistical design using statistical control. The primary source of data for the evaluation was a survey of 3,600 individuals conducted across the country between December 10, 2008, and January 30, 2009. The survey teams interviewed 1,800 individuals who had been exposed directly to the activities of the program, along with 1,800 "control group" individuals who were similar to the treatment group but who had no direct exposure. Treatment-group individuals were selected based on a two-stage random sampling process.

Box 4.2 Examples of Case Study Designs

"Midterm Assessment of the Liberian Teachers Training Program" (LTTP) provides a good example of a case study method in which evaluators used both qualitative and quantitative methods to examine the program's performance. The team collected monitoring data from five rural teacher-training institutes about trainees and trainers. It conducted interviews with the staff of the three in-service cluster program offices. In addition, the team visited seven primary or primary-junior high schools, systematically observed classes, and interviewed the principals and samples of teachers and students. Wherever possible in the school community, the team met with at least one or more members of the local parent-teacher associations. At the national level, the team conducted in-depth interviews and discussions with senior Ministry of Education officials, LTTP staff and consultants.

"Evaluation of the Economic Management for Stability and Growth Program" for Kosovo is a process evaluation that followed a case study design. The evaluators mostly relied on interviews with the officials of the government, international donor agencies and NGOs, national and local experts, and other stakeholders. A major limitation of the evaluation is that it did not use other data collection methods such as surveys or focus group discussions, which would have provided additional information and insights.

The midterm Evaluation of *"Support for Accelerated Growth and Increased Competitiveness Program"* in Senegal also used case study design to evaluate program performance. The team reviewed

the various program documents, which included program plans, annual reports, training material, local consulting contracts, value-chain analysis, success stories, technical manuals and monitoring, and evaluations manuals and data. It also interviewed the officials of USAID, the Government of Senegal ,and private sector organizations/enterprises using interview guides. Finally, it conducted a focus group discussion with the members of a local cooperative.

The number of case study and statistical/quasi-experimental designs was 42 and 14, respectively. Table 4.1 gives the number of evaluation designs by Program Objectives.

Table 4.1
Evaluation Designs by Program Objectives

Objective	Case Study	Statistical/Quasi-Experimental	Total
Peace and Security	4	1	5
Democracy and Governance	5	1	6
Investing in People	28	9	37
Economic Growth	4	2	6
Humanitarian Assistance	1	1	2
Total	**42**	**14**	**56**

Two general conclusions can be drawn from the above table. First, contrary to popular perception, statistical and quasi-experimental designs are not uncommon in USAID evaluations. One in four evaluations conducted by USAID and its partners used them. Second, evaluations in the Program Objective "Investing in People" used these methods in the same proportion as all evaluations. A third of evaluations of Economic Growth interventions (although the number was very small) also fell in this category. A plausible explanation is that that education, health, and economic sectors rely to a greater extent on quantitative data than do evaluations in other sectors, and are therefore more inclined to statistical/quasi-experimental designs.

4.2 Data Collection Methods

The study also examined the use of data-collection methods. The evaluation reports indicate that evaluations used seven data-collection methods: review of project and program documents, key informant interviews, site visits for direct observation, surveys, analysis of secondary data, focus group discussions, and group interviews. Table 4.2 gives the number and percentages for each of these methods.

Table 4.2
Data Collection Methods

| Method	Frequency	Percentage of Evaluation Reports
Review of Project/Program Documents	47	83.9
Key Informant Interviews	45	80.4
Direct Observation	42	75.0
Surveys	29	51.8
Analysis of Secondary Data	28	50.0
Focus-Group Discussion	14	25.0
Group Interviews	12	21.4

A caution is necessary in interpreting the above table. Eight evaluations were conducted as surveys, but their reports do not mention the review of project and program documents as sources of data, when in fact teams would normally review project and program documents before launching surveys. It is also likely that they used key informant interviews in interpreting the survey data.

With this caveat, a few findings can be presented. First, excluding the eight survey-based evaluations, which did not mention other survey collection methods, practically all evaluations used three data collection methods: review of project or program documents, key informant interviews, and site visits for direct observation and interviews. The rationale mentioned for collecting data from multiple sources include triangulation of findings, thereby adding to the credibility of conclusions and recommendations. Second, nearly half of the evaluations conducted surveys, which ranged from national sample surveys to mini-surveys of project beneficiaries. This is quite important, as it indicates that many evaluation teams, even when they adopt a case study design, gather and utilize quantitative data. Third, half of the evaluations

also utilized secondary data such as monitoring data when available, statistics provided by governmental organizations, and in a few cases, data and information gathered by universities and other research institutions. Fourth, group interviews and focus group discussions are not widely used, despite the fact they can generate useful information and insights. One possible explanation is that evaluators do not use these methods because they are not proficient in their use. Finally, while evaluation reports generally documented the research instruments used in surveys, most of them did not include interview protocols for key informant interviews or group interviews. Neither did they provide the lists of topics covered in focus group discussions. As such, one does not get any idea of what issues were covered in interviews and discussions.

The study also tried to assess evaluations based on appropriateness of the data collection methods used for answering evaluation questions. In making such judgments, the investigators not only focused on methodological norms of social science inquiry, but also considered the constraints of time, resources, and availability of data that evaluation teams faced. Chart 4.1 shows that in nearly 75 percent of the cases, evaluation teams were judged to have used appropriate data collection methods.

Chart 4.1
Appropriateness of Data Methods
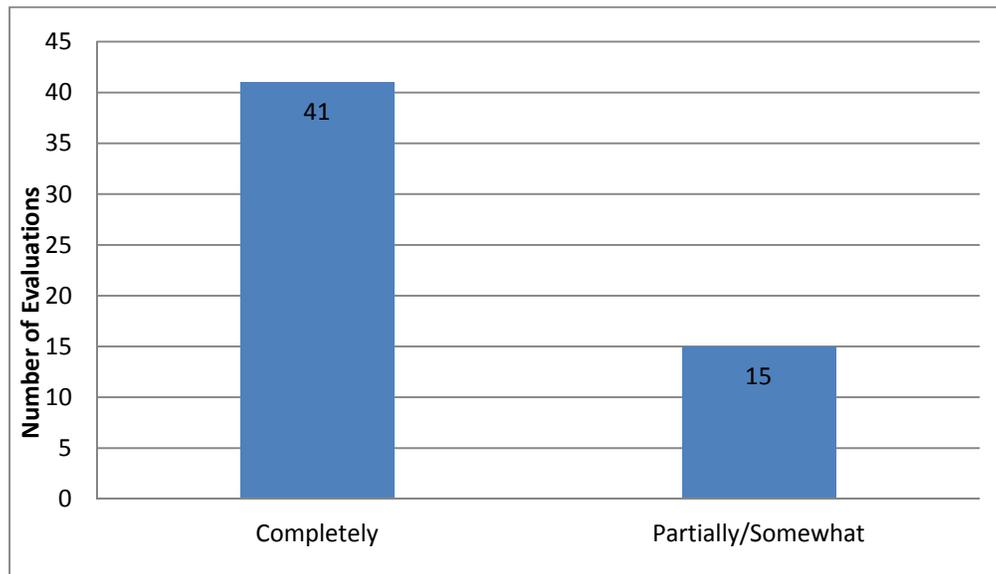
4.3 Conceptual Framework/Model

Finally, the study investigated if evaluators examined the conceptual framework or the logic model that underlies each development intervention and should clarify the causal relationship between inputs, outputs, activities, outcomes, and impacts. A conceptual framework also identifies the set of assumptions about the context, which are critical to the successful implementation of a project and program. A clear articulation of the conceptual framework/logic model enables evaluators to determine if a project or program is performing as originally conceptualized and is likely to produce the intended results. The data indicates that only 26 percent of the evaluations directly or indirectly used or referred to the underlying conceptual framework/model to evaluate the performance or impacts of an intervention.

4.4 Conclusions

The above discussion shows that a quarter of the evaluations follow what may be called "quasi-experimental" or "statistical" designs to test the effectiveness and/or impacts of development interventions. Most of these evaluations have been conducted in the Program Objective "Investing in People." Exponents of rigorous quantitative evaluation designs may argue that this is not satisfactory, and that more quasi-experimental and statistical designs are needed to measure the outcomes of projects and programs funded by USAID. However, the data and findings presented above also indicate that most evaluations used multiple data collection methods. From this, one can conclude that a large majority of the reviewed USAID evaluations employed triangulation among findings from different data sources to enhance the validity of their findings and recommendations. Finally, only a quarter of evaluation reports discussed or referred to the conceptual framework/logic model, which underlies the evaluated interventions. This is hardly encouraging, as the explicit articulation of the underlying conceptual framework contributes to the analytical rigor of evaluations.

PRESENTING DATA, FINDINGS, RECOMMENDATIONS, AND LESSONS

The presentation of data, findings, recommendations, and lessons is critical for the credibility of evaluations and the likelihood that they will be utilized for program and policy formulation. Evidence- or data-based findings are the foundation of credible and useful evaluations. Findings may be quantitative or qualitative, ideally both, but they must be objective and representative. Their sources must be reliable and identifiable. Findings comprise the basis from which evaluation conclusions are drawn, which in turn provide the basis for setting out conclusions, recommendations and drawing lessons learned. Therefore, the study examined this topic in detail.

5.1 Presentation of Data and Findings

The large majority of evaluation reports reviewed contained a clear presentation of the data and sources of the data. However, almost a third of the evaluations were deficient in some aspect of data presentation. Sometimes the units in which the data were expressed were not given, whether in tables or in charts. In other instances, the sources of data were missing. Such information is important for drawing reliable conclusions and recommendations and for the credibility of the evaluation.

Evaluations did not fare well in three important areas. First, a substantial number of evaluations did not provide a clear explanation of limitations of the data, which is necessary for the credibility of evaluations. As illustrated in Chart 5.1, only 25 percent of evaluations provided a clear and complete explanation of data limitations, while over 32 percent provided only a partial explanation and over 43 percent contained no such explanation at all. Box 5.1 provides some specific examples. These examples also suggest that evaluations that were strong on methodology in general also candidly pointed out data limitations. This was not true in every case, but held for a majority of instances.

Second, half of the evaluations failed to include the data collection instruments as annexes (e.g., survey questions, interview protocols, etc.). Only 23 percent of evaluation reports contained all data collection instruments, while 27 percent included only some instruments. This is a significant gap, as inclusion of data collection instruments gives the reader a reasonable idea about the strengths and limitations of data collection.

Chart 5.1
Data Limitations Explained
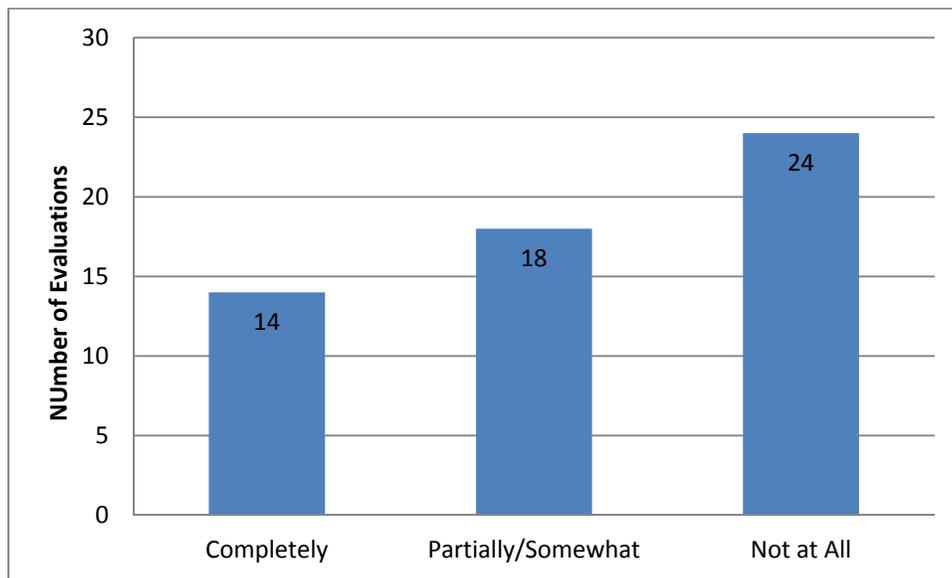
Box 5.1 Importance of Explaining Limitations in the Data: Some Examples

Afghanistan: The security and logistical contexts in Afghanistan tend to impose limitations on the availability and quality of evaluation data. In some cases, such limitations were explained, and in others, there was no explanation, giving the impression that the data were of satisfactory reliability and representativeness. This variation seemed to be the case regardless of the quality of the evaluation in other respects.

- In the case of the summative evaluation of the *"Local Governance and Community Development Program,"* implementation of the evaluation faced many problems in gathering data owing to such factors such as poor security and undependable transportation, among others. The evaluation report included a complete and clear explanation of data limitations and related contextual issues.

- In the process evaluation of the *"Building Education Support Systems for Teachers"* project, the evaluation team was judged as having done a "remarkable job in marshalling all kinds of evidence and presenting them in a coherent way to draw conclusions and recommendations." However, the report lacked any indication of limitations of the data it utilized.

Rwanda: In a quite different setting from Afghanistan, the evaluations of two USAID projects in the human resources area reveal similar variations in addressing data evaluation issues.

- In the summative evaluation of the ***"Community-Provider Partnerships for Quality Improvement: Rwanda Decentralization and Health Program,"*** treatment of methodological issues was in general systematic, comprehensive, and clear, including a conceptual framework, appropriate data collection methods, and explanation of data limitations.

- In the process evaluation of the ***"Good Governance and Health: Assessing Progress in Rwanda"*** program, there was no explanation of data limitations, and other aspects of methodology were weak.

Third, an overwhelming majority (88 percent) of evaluation reports did not offer alternative explanation of their findings. Only seven reports (12 percent) provided some alternative explanation but these were skimpy or unclear. . Four examples are shown in Box 5.2. The latter two, for Nicaragua and Rwanda, are more explicit.

Box 5.2. Alternative Explanations of Findings

"SHOUHARDO: A Title II Program of USAID." This summative evaluation for a Bangladesh program discusses a number of alternative concepts, but none of these constitute alternative explanations for findings. For example, an alternative beneficiary strategy is explained as follows:
> "So long as group solidarity is strong, and each member is clear about the amount of moneys saved and participates in group decisions about the funds, group businesses are likely higher value alternatives than lending funds, which entails the risk of non-payment." (p. 60)

The evaluation also discusses at some length, ***"Alternative Livelihood Options,"*** but again these do not constitute alternative explanations for findings.

"Haiti Emergency Relief Efforts." This process evaluation compared populations without the program with program beneficiaries and found that "there was not another best alternative for their (the beneficiaries') time. Without the CHF cash-for-work activities, it would have been almost impossible for them to find another job. … the testimonies suggest that CHF has created a major impact on the short term job market in Gonaives after the hurricanes." (pp. 2,19)

"Extending Social Insurance to Informal Sector Workers in Nicaragua via Microfinance Institutions: Results from a Randomized Evaluation" This process evaluation explicitly considered alternative explanations as part of its methodology. One of its conclusions was that:
> "Potentially more costly alternatives included paying private doctors or avoiding care altogether, thereby incurring the risk of requiring more extensive and costly health care services down the line." (p. 2)

"Community-Provider Partnerships for Quality Improvement: Rwanda Decentralization and Health Program" This summative evaluation recognizes that other explanations are possible.
> "In addition, the assessment design does not allow us to present conclusive evidence on the extent of the PAQ (Partnership for Quality Improvement) committees' influence on health center performance indicators. Because use of prenatal care, assisted delivery and vaccination services increased both prior to and after the committees' initiation, it is unlikely that the PAQ

---

committees were the sole cause of the documented increases in service use over the project period. However, respondents' anecdotal observations suggest that PAQ outreach activities in the three content areas were considerable and probably deserve some credit for the rise in numbers. Finally, as extensively documented elsewhere, Rwanda's MOH has successfully introduced a number of progressive innovations over the past several years, including national scaling up of community-based health insurance (*mutuelles*), performance-based financing (PBF) and a variety of training and quality improvement approaches. In this rapidly changing climate, it is difficult to attribute improvements in specific indicators to one intervention without control groups and multivariate analysis." (p. 32)

5.2 Adequacy of Recommendations

Recommendations drawn from evidence-based findings and conclusions are one of the hallmarks of evaluations that make them useful to stakeholders. They distinguish evaluations from research, which may or may not include recommendations. In fact, of the 56 evaluations reviewed, 54 included recommendations, and of the two that did not, one employed solely a quasi-experimental design, more akin to a research study.

The clarity of recommendations, their logical link to conclusions and findings, and the extent to which they are actionable are more important than their quantity. In fact, too many recommendations are to be avoided if an evaluation is to have impact. There were a number of cases in which evaluators included far too many recommendations. For example, the summative evaluations for the "Private Sector Program in Ethiopia" (for public-private health-sector collaboration) contained 38 recommendations in no apparent order of priority. Similarly, "Land O'Lakes Zambia Title II Development Assistance Program" in Zambia contained 18 "key" recommendations.

This study particularly focused on two issues: recommendations should follow from the findings, and they should be actionable. On both these criteria, evaluations did well. The recommendations in 75 percent of the reports (42) were judged to have followed logically from their findings, and almost 25 percent followed somewhat from their findings. Over 50 percent of recommendations were actionable, while another 40 percent were somewhat actionable.

Box 5.3 provides brief assessments from the study's "Overall Comments" on the quality of recommendations in an illustrative selection of USAID evaluations. The examples tend to emphasize the weaknesses in recommendations and areas for

improvement, although the last two examples from Malawi and Egypt report stronger recommendations. They also reveal a general finding that virtually all the evaluations had room for improvement. Even those evaluations with clear, logical, and actionable recommendations had room for improvement in other areas.

Box 5.3 Assessments of Quality of Recommendations

"Evaluation of the Comprise-A (Communications for Behavioral Change Expanding Access to Private Sector Health Products and Services for Afghanistan) Social Marketing Program." Some recommendations are trivial or non-actionable.

"South Serbia and Sandzak Economic Crisis Impact Assessment" is a quasi-research/evaluation study, not a programmatic evaluation. Interviews were conducted with an extremely small sample of business owners in different parts of the country. However, there is a candid discussion of data limitations. But the basis for conclusions and recommendations is slim. Most recommendations are not very actionable. The language is not always clear.

"PEPFAR Public Health Evaluation Care and Support" (Uganda) is a thorough, mostly technical evaluation, containing 56 tables reporting on quantitative and qualitative findings. All three data collection instruments are included as annexes. Conclusions tend to be merged with recommendations, and the latter tend not to be actionable.

"Final Evaluation: The Private Sector Program in Ethiopia" (Public-Private Health Sector Collaboration). There are some actionable recommendations in this final evaluation but there are a plethora of recommendations that lack a strategic sense of priorities.

"Assessment of USAID/BASICS' Community Essential Nutrition Actions Program in Malawi" is a comprehensive, clear, and detailed evaluation report, which indicates a careful selection of program and control groups. Actionable recommendations are grouped by priority. There is also a useful comparison with five other countries.

"Takamol (Integrated Reproductive Health Services) Mid-Term Evaluation" (Egypt) is a first-rate evaluation report: analytical, well organized, complete, and clearly written. The recommendations are actionable, and the parties responsible for implementing them are identified.

5.3 Presentation of Lessons

Lessons are similar to recommendations in that they imply actions based on the findings and conclusions of evaluations, but they are intended to be broader in their applicability beyond the specific project being evaluated. They are typically not as immediate in their call for action as recommendations, nor are they necessarily addressed to specific stakeholder groups or action parties. Lessons are of the order of answers to the question: "what do we learn from this experience that might be applicable to situations with similar circumstances elsewhere?" The data presented in Chart 5.2 indicate that only half the evaluations reviewed contained lessons.

Chart 5.2
Lessons Mentioned
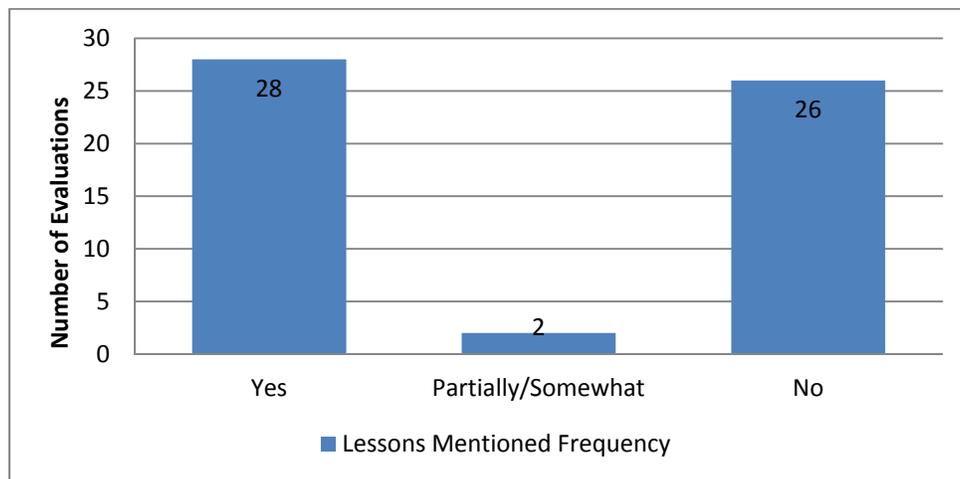
As shown in Chart 5.3, analysis of the data shows that of the 35 evaluation reports that referred to lessons, less than half (16) contained fully adequate explanations of lessons. Six cases summarized in Box 5.3 show features of what were judged by the investigators to be good examples, including varying but useful ways of organizing lessons. For example, good explanations were brought together in one place (in the main text and/or executive summary), identified in the table of contents, were based on the evaluation's findings, future-oriented, and programmatic as well as management-oriented. While they may not be as specific or actionable as recommendations, lessons should at least indicate lines of action (see the Nigeria case in Box 5.4 for a useful way of organizing lessons).

Chart 5.3

Lessons Adequately Explained
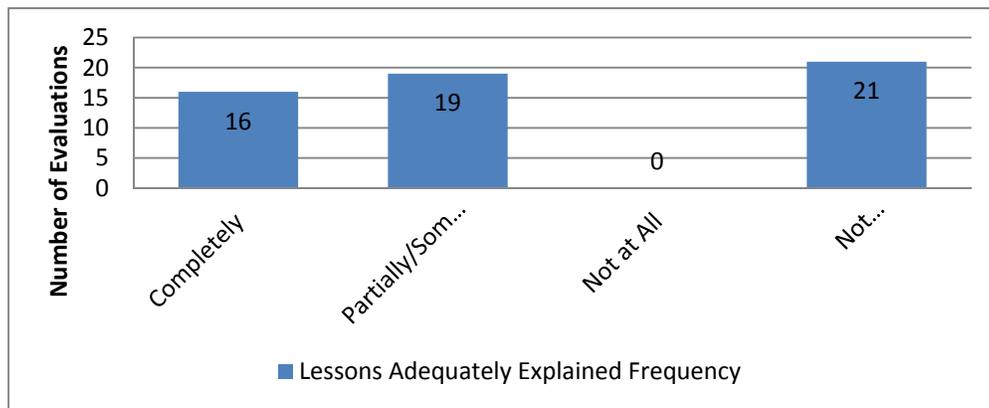
Box 5.4 Explanations of Lessons in Selected USAID Evaluation Reports

"Takamol Mid-Term Evaluation" (Egypt) contains a separate chapter on best practices and lessons learned for the purpose of "showing real potential for replication and sustainability beyond the life of the Takamol Project."

"Final Evaluation: The Private Sector Program in Ethiopia (Public-Private Health Sector Collaboration)" contains a three-page "lessons learned" section in the Executive Summary, which includes 23 lessons learned grouped into four categories.

"Community Stabilization Program: an Examination of the Youth Engagement Program" (Iraq). Examples of programmatic lessons learned identified in the report are that programs should:

- Be conflict-sensitive. Programs working in conflict situations (and especially working on the conflict) should be based on a conflict analysis to understand conflict drivers and triggers, and how program interventions may impact the conflict dynamic, in order to avoid exacerbating tension and to achieve program objectives.
- Try to have explicit "theories of change" that are relevant to youth and their conflict context, realistic in what they attempt to achieve, and where possible, based on evidence or growing evidence.
- Be designed in a way that ensures that their outcomes are relevant to their objectives.
- Be meaningful for youth.
- Be holistic. Holistic programming that is strategic, integrated, and cuts across sectors can better address the multiple challenges affecting youth.
- Pay attention to ensure that the selection of beneficiaries matches the program rationale.

"Child Survival - 20: Scaling-Up Community-Based Services In the Sikasso Region of Mali." Lessons are described in the Executive Summary and in a two-page section in a subsequent chapter of the main text,

preceding a recommendations section in the same chapter.

"Midterm Evaluation of the Conflict Abatement through Local Mitigation (CALM) Project" (Nigeria). The Executive Summary contains "Key Lessons and Future Considerations," and the report includes one-and- a-half pages on "Additional Lessons about Conflict Abatement in Nigeria." These are grouped under the following headings:

- Assumptions about causes of violent conflict
- Education in conflict prevention and peace building
- Role of the press
- Linkage and integration with related USAID programs
- Inter-agency and cross-donor coordination
- Program management and financing

"Mid-Term Evaluation Support for Accelerated Growth and Increased Competitiveness IQC Final Report Senegal." Chapter 5 contains "Conclusions, Lessons Learned, and Best Practices" under three headings: "Business Development Services," "Public-Private Partnership," and "Policy Reform." This is followed by a chapter on "Recommendations and Strategic Options."

5.4 Conclusions

A clear presentation of data, findings, recommendations, and lessons is critical for the credibility of evaluations and their use. Evidence- or data-based findings are the foundation of credible evaluations. To be useful, evaluations must formulate actionable recommendations and lessons that are based on findings and conclusions. There is room for improvement in USAID evaluations in these respects. Only a quarter provided clear explanations of the limitations of data, and only half included data collection instruments. Three-quarters of evaluations contained recommendations that were judged to have followed logically from their findings, but only about half were considered to contain actionable recommendations. Only half of the reviewed evaluations contained lessons for the future, and less than a third of evaluations adequately explained the lessons they sought to draw.

STUDY RECOMMENDATIONS

Making recommendations from this small study should be done with caution. As stated in the introduction, the study focused only on evaluations that were conducted in 2009, and were available in the DEC. Moreover, it is solely based on evaluation reports, as the investigators did not have time and resources to interview evaluators, evaluation managers, and other stakeholders. Even then, this report has identified many problem areas that deserve the attention of USAID and the Department of State. In formulating recommendations, the investigators focused on two areas: those that are often overlooked by experts and decision makers, and those that will improve the analytical and conceptual rigor of the evaluations and enhance their credibility.

The number of evaluations conducted by USAID is extremely small. It indicates that most of their interventions are not evaluated, although USAID mandates that each major intervention should be evaluated at least once in its lifetime. During the past year, USAID has reaffirmed its commitment to evaluations and has taken steps to increase both the quantity and quality of evaluations. These efforts should continue.

This report's recommendations are the following:

1. **Provision of sufficient time and person-days for evaluation teams.** As discussed earlier in this report, sufficient time and person days are often not allocated for evaluations to conduct necessary fieldwork. This is particularly the case with evaluations that are expected to assess the outcome and impacts of development interventions, and/or are conducted in war-torn societies such as Afghanistan and Iraq in which institutional, political, and security considerations pose barriers to data collection. It should also be noted that time and person-days are not perfect substitutes for each other. The implication is that increasing the size of the team is not a substitute for an adequate amount of time in country for the entire team.

   Recommendation: **USAID/DOS should require that Operating Units budget sufficient time and person-days, particularly for conducting summative evaluations, and evaluations in war-torn societies.**

2. **Participation of host-country evaluators.** The participation of host-country evaluators and researchers is necessary for many reasons. They provide an insider's perspective, and are usually able to better relate to local stakeholders and informants. Their participation also helps to build local capacities for evaluation and research, and can be instrumental in the diffusion of technical expertise. It can also reduce the overall costs of evaluations. The data presented indicate that close to 40 percent of evaluations did not include host-country evaluators or researchers (this figure excludes the participation of support staff).

   Recommendation: **Guidelines should be issued which require that all evaluations should wherever possible, have at least one local evaluator or researcher.**

3. **Involvement of host-county government officials and project staff in evaluations:** In many cases, government officials and project or program staff participate in evaluations. Their participation can be beneficial to the extent that they become aware of the ongoing intervention, the problems it faces, and the results it is likely to achieve. On the other hand, it can compromise the integrity of evaluations. Intended beneficiaries may be reluctant to share their opinions and particularly criticisms. Therefore, it is important that the role of government officials and project or program staff be clarified and explained in the evaluation.

   Recommendation: G**uidelines or a TIP should be issued to clarify the possible roles for host-country government officials and project or program staff in evaluations, and require that evaluations clearly spell out the nature and extent of their participation.**

4. **Documentation of qualitative and quantitative data collection instruments and data:** While most evaluations did provide questionnaires or the summary of questionnaires used in surveys, they usually did not include interview protocols or checklists for key informant interviews, group interviews, or even focus-group discussions. As such, the reader has no idea about what issues were covered in them. This undermines the credibility of evaluations. Therefore, if research instruments and the summary of important items covered are included in evaluation reports, the credibility of findings and recommendations would improve.

Recommendation: **Guidelines should be issued on the presentation of qualitative and quantitative data in evaluation reports.**

5. **Discussion of conceptual framework or model**: Almost 75 percent of evaluations do not refer to the underlying conceptual framework or model for the project or program. Even when they mention it, the discussion is brief and superficial. However, the importance of the underlying conceptual framework or model in examining the outputs, performance, outcomes, or impacts cannot be over-emphasized.[8]

Recommendation: **Evaluators should be encouraged to examine and report on the underlying conceptual framework or model while evaluating the performance and impacts of an intervention.**

6. **Improving the quality of evaluation reports:** The study has identified several problems with the reviewed USAID evaluation reports. An overwhelming majority of evaluation reports do not include SOWs. Many do not include research instruments used for data collection. More importantly, they do not mention the limitations of the data or offer alternative explanations wherever possible. Recommendations are usually not prioritized. Many times, evaluations did not distinguish between lessons and recommendations.

Recommendation: **USAID should consider revising its guidelines and TIP for preparing evaluation reports taking into consideration the above-mentioned shortcomings.**

---

[8] As part of, or along with the framework, the project document should include baseline data and information.

List of Evaluations Reviewed

Report Title	SPS Objective	Country	Region
Provincial Strengthening in Northern Afghanistan: Capacity Building and Innovation to Support Basic Package of Health Services	Obj. 3- Investing in People	Afghanistan	Asia
Local Governance and Community Development Program	Obj. 2- Governing Justly and Democratically	Afghanistan	Asia
Evaluation of the Comprise-A (Communications for Behavioral Change Expanding Access to Private Sector Health Products and Services for Afghanistan) Social Marketing Program, December 2008	Obj. 3- Investing in People	Afghanistan	Asia
Building Education Support Systems for Teachers	Obj. 3- Investing in People	Afghanistan	Asia
Mid-Term Evaluation of Afghan e-Quality Alliance	Obj. 3- Investing in People	Afghanistan	Asia
Tech Serve Program	Obj. 3- Investing in People	Afghanistan	Asia
"Life and Livelihoods: A Title II Program of USAID"	Obj. 3- Investing in People	Bangladesh	Asia

"SHOUHARDO: a Title II program of USAID"	Obj. 3- Investing in People	Bangladesh	Asia
Local Administration and Reform Project	Obj. 2- Governing Justly and Democratically	Cambodia	Asia
Evaluation of USAID Human Rights Program in Colombia	Obj. 2- Governing Justly and Democratically	Colombia	Latin American and the Caribbean
Victim of Torture Fund Evaluation	Obj. 3- Investing in People	Congo	Africa
USAID/Ethiopia External Mid-Term Evaluation of the HIV/AIDS care and Support Program	Obj. 3- Investing in People	Ethiopia	Africa
International Human Rights Law Outreach Program	Obj. 2- Governing Justly and Democratically	Egypt	Middle East
Takamol (Integrated Reproductive Health Services) Mid-Term Evaluation	Obj. 3- Investing in People	Egypt	Middle East
"Final Evaluation: The Private Sector Program in Ethiopia" (Public-Private Health Sector Collaboration)	Obj. 3- Investing in People	Ethiopia	Africa
Agricultural Recovery in the Commune of Gros-Morne Artibonite Department, Haiti	Obj. 4- Promote Economic Growth and Prosperity	Haiti	Latin America and the Caribbean
"Haitian Emergency Relief Efforts (HERE)"	Obj. 5- Humanitarian Assistance	Haiti	Latin America and the Caribbean

"PUBLIC LAW (PL) 480 TITLE II PROGRAM IN HONDURAS FY2005-2009"	Obj. 3- Investing in People	Honduras	Latin America and the Caribbean
Mid-Term Review of Vistaar Project	Obj. 3- Investing in People	India	Asia
"Sustaining the lives and dignity of IDPs in Purnea district – Bihar"	Obj. 5- Humanitarian Assistance	India	Asia
"COMMUNITY STABILIZATION PROGRAM (CSP):AN EXAMINATION OF THE YOUTH ENGAGEMENT PROGRAM"	Obj. 1- Peace and Security	Iraq	Middle East
Stabilization Program (CSP) in Iraq	Obj. 1- Peace and Security	Iraq	Middle East
The Impact of second National Kenya civic Education Programme on Democratic Attitudes, Values and Behaviour	Obj. 2- Governing Justly and Democratically	Kenya	Africa
"Plan International KIDCARE Child Survival Project"	Obj. 3- Investing in People	Kenya	Africa
"Summary: PEPFAR Public Health Evaluation – Care and Support– Phase I: Kenya"	Obj. 3- Investing in People	Kenya	Africa
"Decentralizing Kenya's Health Management System: An Evaluation"	Obj. 3- Investing in People	Kenya	Africa
"Evaluation of the Economic Management for Stability and Growth Program"	Obj. 4- Promote Economic Growth and Prosperity	Kosovo	Europe and Eurasia

Mid-Term Evaluluation of Liberia's Teaching Training Program	Obj. 3-Investing in People	Liberia	Africa
"IMPACT ASSESSMENT REPORT: LIBERIA ENERGY ASSISTANCE PROGRAM (LEAP)"	Obj. 4-Promote Economic Growth and Prosperity	Liberia	Africa
ASSESSMENT OF USAID/BASICS' COMMUNITY ESSENTIAL NUTRITION ACTIONS PROGRAM IN MALAWI	Obj. 3-Investing in People	Malawi	Africa
"MID-TERM EVALUATION OF THE BEGINNING LITERACY PROGRAM OF MALAWI"	Obj. 3-Investing in People	Malawi	Africa
"Tube Poka Child Survival Project FINAL EVALUATION REPORT"	Obj. 3-Investing in People	Malawi	Africa
"Child Survival - 20: Scaling-Up Community-Based Services In the Sikasso Region of Mali"	Obj. 3-Investing in People	Mali	Africa
"Synergy and Action for Nutrition+ (SAN+) - Child Survival Project – Koulikoro Region, MALI"	Obj. 3-Investing in People	Mali	Africa
"EVALUATION OF INJECTION SAFETY AND HEALTH CARE WASTE MANAGEMENT IN MMIS EXPANSION AREAS IN MOZAMBIQUE"	Obj. 3-Investing in People	Mozambique	Africa

"OTI Nepal Program Evaluation: 2006-2009"	Obj. 1- Peace and Security	Nepal	Asia
"EXTENDING SOCIAL INSURANCE TO INFORMAL SECTOR WORKERS IN NICARAGUA VIA MICROFINANCE INSTITUTIONS: Results from a Randomized Evaluation"	Obj. 4- Promote Economic Growth and Prosperity	Nicaragua	Latin America and the Caribbean
"MID-TERM EVALUATION OF THE CONFLICT ABATEMENT THROUGH LOCAL MITIGATION (CALM) PROJECT"	Obj. 1- Peace and Security	Nigeria	Africa
"EVALUATION OF INJECTION SAFETY AND HEALTH CARE WASTE MANAGEMENT IN NIGERIA"	Obj. 3- Investing in People	Nigeria	Africa
"Assessment of USAID's Child Welfare Programs in Russia"	Obj. 3- Investing in People	Russia	Europe and Eurasia
"Evaluation of USAID Support to Tuberculosis Control in the Russian Federation"	Obj. 3- Investing in People	Russia	Europe and Eurasia
"Community-Provider Partnerships for Quality Improvement: Rwanda Decentralization and Health Program"	Obj. 3- Investing in People	Rwanda	Africa
"EVALUATION OF THE HOME BASED MANAGEMENT OF MALARIA STRATEGY	Obj. 3- Investing in People	Rwanda	Africa

IN RWANDA 2008"			
"Evaluation of the Capacity Project's Human Resources Information Systems (HRIS) Strengthening Process in Swaziland, Uganda and Rwanda"	Obj. 3- Investing in People	Rwanda	Africa
"Good Governance and Health: Assessing Progress in Rwanda"	Obj. 3- Investing in People	Rwanda	Africa
"Mid-Term Evaluation of Task Order No. 1Support for Accelerated Growth and Increased Competitiveness IQC Final Report"	Obj. 4- Promote Economic Growth and Prosperity	Senegal	Africa
"SOUTH SERBIA AND SANDZAK ECONOMIC CRISIS IMPACT ASSESSMENT"	Obj. 4- Promote Economic Growth and Prosperity	Serbia	Europe and Eurasia
Fostering Justice in Timore-Leste:Rule of Law Program Evaluation	Obj. 2- Governing Justly and Democratically	Timor-Leste	Africa
EVALUATION OF INJECTION SAFETY AND HEALTH CARE WASTE MANAGEMENT IN UGANDA"	Obj. 3- Investing in People	Uganda	Africa
"PEPFAR Public Health Evaluation Care and Support"	Obj. 3- Investing in People	Uganda	Africa
"YEMEN BASIC HEALTH SERVICES (BHS) PROJECT Mid-	Obj. 3- Investing in People	Yemen	Middle East

Term Evaluation"			
"INTALEQ Student Assessment Report"	Obj. 3- Investing in People	Yemen	Middle East
"USAID/Zambia Education Program Evaluation"	Obj. 3- Investing in People	Zambia	Africa
"An Evaluation of Interactive Radio Instruction in GRZ Schools in 2008"	Obj. 3- Investing in People	Zambia	Africa
"Land O'Lakes Zambia Title II Development Assistance Program"	Obj. 3- Investing in People	Zambia	Africa
"TEACHING IN THE WINDOW OF HOPE"	Obj. 3- Investing in People	Zambia	Africa

Checklist for Reviewing Foreign Assistance Evaluations

Background Information

      Title

      Type of Evaluation

      Objective areas

      Operating Unit

      Region

      Reviewer

      Date of review

Scope of Work

(a) Was SOW attached?
      Yes_____ No_____

(b) Was the purpose or rationale of the evaluation stated, including specific objectives?
      Rating_____ Yes_____ No_____

(c) Were evaluation questions clearly stated?
      Yes_____ No_____

(d) Did SOW describe evaluation design, including relevant comparisons to be made?
      Yes_____ No_____

(e) Did SOW describe data collection methods?
      Yes_____ No_____

(f) Could evaluation questions be answered within given time and resource?
      Rating_____

(g) Was the audience for the evaluation identified?

      Yes_____ No_____

(k) What was the size of the evaluation team?

Number_____

(l) What was the number of evaluators/researchers from the host country?

Number_____

(k) How long was the evaluation exercise?

Number of days_____

(l) How long was the team in-country?

Number of days_____

Evaluation Methodology

a) Did the report present to a theory of change, development hypothesis, logic model or a conceptual framework relating interventions to results and outcomes?
Yes_____ No_____

b) What was the nature of evaluation design for impact evaluations?
Experimental_____ Quasi-experimental_____ *Interrupted Time-Series and Cross-Sectional :)*_____ Single case study---multi-case study_____

c) Were data collection methods spelled out in the report or annex?
Yes_____ No_____

d) What data collection methods were used?
   a. Review of project/program documents?
      Yes_____ No_____

   b. Secondary analysis of data
      Yes_____ No_____

   c. Focus group discussion
      Yes_____ No_____

d. Group interviews

Yes_____ No_____

e. Key informant interviews

Yes_____ No_____

f. Surveys

Yes_____ No_____

e) Were research instruments included in the evaluation report?

Yes_____ No_____

f) Were the methods appropriate to answer evaluation questions?

Rating_____

General
Comments_____

_____

g) Did impact evaluations distinguish between impacts and outcomes?

Yes_____ No_____

h) Were the data and sources of data clearly presented in the main report or an annex?

Rating_____

i) Did the report describe in a transparent manner the limitations of the data collected?

Yes_____ No_____

General
Comments_____

Findings

   a) Were all evaluation questions answered in the report?
       Yes_____ No_____

   b) Were the findings clearly stated and explained?
       Rating_____

   c) Did the findings logically follow from the data and information?
       Rating_____

   d) Were alternative explanations of findings explicitly considered and explored?
       Yes_____ No_____

   General Comments_____

Recommendations

   a) Did the recommendations follow from the evaluation's findings?
       Yes_____ No_____

   b) Were the findings clearly stated and explained?
       Yes_____ No_____

   c) Did the findings logically follow from the data and information?
       Rating_____

   d) Were alternative explanations of findings explicitly considered and explored?
       Yes_____ No_____

   General Comments_____

Lessons:

   a) Did the report mention lessons which may be used in designing new projects
       and programs?
       Yes_____ No_____

   b) Were they adequately explained?
       Yes_____ No_____

Contents and Style

a) Was the language clear?
   Rating_____

b) Were tables and figures properly presented?
   Yes_____ No_____

c) Did the report have an executive summary?
   Yes_____ No_____

d) Did it include a list of abbreviations and acronyms and annexes?
   Yes_____ No_____

e) Were references properly cited?
   Yes_____ No_____

f) Were research instruments included?
   Yes_____ No_____