



Evaluation of Recent USAID Evaluation Experience

WORKING PAPER No. 320

by

Cynthia Clapp-Wincek
Richard Blue

June 2001

Center for Development Information and Evaluation



U.S. Agency for International Development
Washington

Evaluation of Recent USAID Evaluation Experience

13 February 2001

**By Cynthia Clapp–Wincek
Richard Blue**

**Submitted to U.S. Agency for International Development
PPC/CDIE**

Table of Contents

SUMMARY	iii
I. INTRODUCTION.....	1
II. USAID’S RECENT EXPERIENCE	5
A. How Many?	5
1. The numbers.....	5
2. Are there enough evaluations?	10
B. How Good are USAID’s Evaluations?.....	13
1. Quality.....	13
2. Methods	14
3. The teams	15
4. Scopes and guidance to teams	16
5. Time and money.....	17
6. Contracting complications.....	17
7. To criticize or not to criticize?	18
C. Are Evaluations Making a Difference in USAID Programs?.....	18
1. Are evaluations used and how?.....	18
2. Are we evaluating the right things?	25
D. Conclusions.....	27
III. USAID AS A LEARNING ORGANIZATION.....	29
A. The Learning Organization.....	29
B. Recommendations for a Learning Organization.....	32
APPENDICES	
1. Statement of Work.....	A1-1
2. Mechanics of USAID Evaluation.....	A2-1
3. Questionnaires	A3-1
4. Fieldwork Interview Schedules.....	A4-1
5. Bibliography.....	A5-1

ACRONYMNS

ACVFA	Advisory Committee on Voluntary Foreign Aid
AFR	Africa Bureau of USAID
ADS	Automated Directive System of USAID
ANE	Asia and Near East Bureau of USAID
BHR	Bureau for Humanitarian Response
CDIE	Center for Development Information and Evaluation
DEC	Development Experience Clearinghouse under CDIE
E&E	Europe and Eurasia Bureau of USAID
FSN	Foreign Service National employee
GAO	U.S. General Accounting Office
IG	Inspector General
IQC	Indefinite Quantity Contract mechanism
LAC	Latin America and Caribbean Bureau of USAID
NGO	Nongovernmental Organization
PD&S	Project Development and Support funds
PSC	Personal Services Contractor
PVO	Private Voluntary Organization
R4	Results Review and Resource Request
REDSO	Regional Economic Development Services Organization
SO	Strategic Objective
SOW	Scope of Work
USDH	U.S. Direct Hire employee

SUMMARY

Evaluation can be a powerful learning tool to identify lessons learned and to improve the quality of the work of all partners involved in the process. Evaluation is only one part of the Agency's current learning process. It is distinguished from the performance monitoring system because it looks back at the structure of the design of Agency programs to understand *why* things have changed. It is distinguished from the newly popular assessments because of its focus specifically on Agency programs and its greater expectation of good social science and objectivity.

As this evaluation was being written, USAID revised the Agency evaluation policy to add a new dimension to evaluations:

Applying the lessons of successes and failure systematically and providing leadership in tackling complex problems.

What has been learned about the Agency's evaluation experience is applied to help inform decisions and actions needed for the Agency to fully become this kind of learning organization. USAID's success in making a difference in people's lives around the world requires learning about what works and what doesn't, and applying those lessons.

Conclusions about Evaluation

This study was requested by USAID's Center for Development Information and Evaluation (CDIE) to evaluate the Agency's recent experience with evaluation. The study period begins in 1995 when Agency policy changed from requiring every project be evaluated to recommending that evaluations only be done in response to management need. The key questions of this evaluation are what have been the quantity, quality, and results of evaluations in USAID since that time?

1. Considerable evaluation work is being used in decision making and program planning as managers struggle to continue to carry out quality programs. Sixty percent of those that responded to the study's questionnaire reported that evaluation had been critical to their understanding of activity performance. Fieldwork presented many further examples of evaluation being used in mission decisions, designs, and debates.
2. However, USAID is doing markedly fewer evaluations and only a small percentage of completed evaluations is being submitted to Washington compromising broader learning. Of most concern is the very limited number of in- depth, program evaluations.
3. The monitoring/evaluation balance has swung too much toward performance monitoring. Recognizing a) the advances of improved performance monitoring (although the job is not done), b) the considerable efforts of those USAID staff that have continued to carry out evaluations, and c) that evaluation has a different role to play, the evaluation system has been ignored too long.
4. Partners are doing much of the evaluation work that is being done. The partner organizations are learning from the experience; USAID is not. USAID managers,

- particularly the busy team leaders, skim reports and react to the most pressing actions needed. But only the most committed managers find time to do the evaluations at all. Most of this work does not circulate beyond the partner and the SO team.
5. The field-level evaluation work tends to focus on *activity* performance, especially by implementing partners. R4 monitoring systems assess progress at the strategic level. There is a disconnect between activity evaluation and SO monitoring that cannot be closed by more activity monitoring. SO leaders, in the main, are not carrying out strategic evaluations at the SO level.
 6. The methods used for USAID evaluations are dominated by the “fly-in” approach having a serious effect on the quality of the evaluations. As things stand now:
Scopes ask a team to come for 4–6 weeks and interview the mission, the activity staff, and ‘representatives’ of the local people. There isn’t enough time to get any kind of representative sample. The team frequently tells the USAID manager pretty much what he already knows.
This isn’t worth the time and money we are spending on it.
 7. Insufficient knowledge and understanding of evaluation planning and methods have led to this. Fairly minimal advice and guidance from evaluation experts soon after the decision could lead to significant improvement. Refining evaluations questions and identifying appropriate and timely methods are the key. By the time a team arrives, it is too late to affect the methods in any significant way.
 8. Professional commitment to doing a good job is the primary incentive for good management, including evaluation. In general, professional staff do not see any Agency-wide incentive to advance learning through evaluations.
 9. There were too many examples of evaluations being linked with success stories. Continuing this practice will call into question the objectivity and independence of evaluation findings and conclusions.

The biggest structural change in USAID’s work is the increasing reliance on non-career staff, few of whom have any long term stake in Agency learning. Implementation is “wholesaled” to a variety of development partners. Privatizing our monitoring and evaluation undercuts the flow of information as organizations (both profit and nonprofit) recognize information as an element of competitive advantage. Moreover, contracting and granting rules create serious barriers to learning from experience and applying lessons to improve our assistance.

The clearest picture of this evaluation is from the comment by the USAID manager who said he spent 90 percent of his time on things that are required and that evaluation had a hard time making it to the top of the 10 percent list. There are good, committed professionals in the field who need support to accomplish the type of programs they know they should.

The linchpin of several pages of recommendations at the end of the report is that senior Agency managers make *learning* about what works and what doesn’t of primary importance. Only with visibility and high-level support will evaluation, along with performance monitoring, be a means by which “truth speaks to power.” More importantly, USAID needs to find a way to “use truth to empower.”

I. INTRODUCTION

The purpose of this study was to “evaluate USAID’s recent experience with evaluation”¹ in order to understand the decline in reports submitted to USAID’s library. The scope identified 1994 as the beginning of the decline and noted that this was close in time to the 1995 change in Agency evaluation policy. The key question of this evaluation is What has been the quantity, quality, and results of evaluations in USAID since the policy was changed in 1995?

Underlying this straight-forward question are a number of loaded questions:

Wouldn’t fewer evaluations be better, when they are so expensive and time consuming?

Hasn’t the R4 largely replaced evaluation?

What is the balance between evaluation and monitoring? What is really the difference anyway?

Do we need formal evaluations or do the more flexible “assessments” better serve USAID?

This study addresses these questions in two contexts: first in the context set by the policy and guidance in the 1995 ADS and secondly against the expectations in the September 2000 revision of the ADS.

Overview of the Report

In section II, USAID’s experience with evaluations is evaluated against the expectations and standards from the 1995 evaluation policy. The policy stated:

A decision to carry out an evaluative activity shall be driven primarily by management need. Evaluations are not required as a matter of formality.

That policy emphasized evaluations’ usefulness to managers and that had many implications for how things changed. So first, evaluation in USAID is analyzed in the context of the 1995 policy. Did the evaluations and managers’ use of evaluation change as expected? Section II describes how the system did shift in the direction of managers deciding when it was useful to evaluate, and analyzes why things changed the way they did.

As this evaluation is being written, USAID has just revised the Agency evaluation policy again.² ADS 200 series dated September 2000 has added a new dimension to evaluations. The “Reform Vision” in ADS 200 states these expectations:

Applying the lessons of successes and failure systematically and providing leadership in tackling complex problems that demand multi-agency or multi-donor responses.

¹Scope of work in appendix 1.

²The original intent was for study findings to inform and help shape the ADS guidance.

In section III, what has been learned about the Agency’s evaluation experience is applied to how that informs decisions and actions that could help the Agency fully become this kind of learning organization that meets the expectations stated above. Missions and SO teams are still expected to conduct evaluations in response to management need, but the vision has been broadened. So secondly, evaluation in USAID is analyzed against the *new* standards of a learning organization operating in a much more complex environment.

It needs to be acknowledged that Section III applies a set of standards and expectations to the Agency’s evaluation experience that were not in place during the study time period. That the Agency did not achieve that standard is not a shortcoming in any way. Evaluating USAID as a “learning organization” is done in an attempt to use the findings and lessons from past Agency experience to make recommendations on how best to move in the new direction.

The scope of work limits this evaluation to Agency and partner evaluations. Much of the current thinking in the evaluation community addresses the importance of building evaluation capacity in host countries. Another key trend is donors coordinating their evaluation requirements and efforts. Both are significant for a full understanding of evaluation efforts in the development community, but this study stays in the narrower confines of the Agency and partners.

Definitions

Several key definitions are necessary from the beginning. *Evaluation*, for this study, was defined as: any empirically-based analysis of problems, progress, achievement of objectives or goals, and/or unintended consequences for missions’.³

- country strategy,
- sectoral programs, and/or
- activities

whether the evaluation was conducted by

- mission staff,
- partners, or
- outside consultants.

Evaluation refers to systematic analysis of the performance of USAID-supported programs and not just the more general types of studies such as sector assessments or analyses of the political, economic and social environments that help to define mission’s strategic focus and context.

Monitoring is different from evaluation in two ways:

- data are collected repeatedly, preferably on a schedule, and

³This evaluation uses missions rather than “operating unit.” The emphasis in our work was on field evaluation but 30-40 interviews were conducted in Washington supporting our understanding of evaluation for the Agency as a whole. More analysis of Washington operating units is included in Appendix 2 “Mechanics of USAID Evaluation.”

- monitoring asks: is the activity/strategy progressing as planned? The monitoring function does not analyze causal relationships although a monitoring system may.

Managers review the monitoring data with two questions in mind: how are things going? Is the expected amount of change occurring? It is an assumption that the change measured is because of the intervention as planned. This data added to the manager's knowledge of the environmental context and other competing explanations can lead to a pretty good informal "evaluation" of the results. Together these might be parts of a monitoring system, but monitoring alone means only the repeated collection and reporting of data.

Methods

Capturing change in the complex system of USAID's evaluation program required several approaches:

- The pattern analysis. An annotated list of all 1998–1999 evaluations catalogued in the USAID library was analyzed by Bureau, sector and type of evaluation. Regional and sectoral patterns yielded little of interest. Others are discussed in section II.
- Review of 1998–1999 evaluations. 128 of 1998–1999 evaluation reports from USAID's Development Experience System (DEXS) were reviewed for 1) use of methods, 2) authors, 3) scopes of work and 4) sponsorship. The only reports read for content were those collected from the fieldwork countries.
- Questionnaires. A brief questionnaire was sent by email to 283 mission staff. 94 responded (representing 44 missions). A brief questionnaire was developed for partners as well. It was advertised in Interaction's "Monday Developments," sent to members of the Advisory Committee on Voluntary Foreign Aid (ACVFA) and made available at an ACVFA public meeting. Attempts to systematically reach contractor and university audiences directly were unsuccessful. The 21 responses received were too few to tabulate. Comments from the partner questionnaires have been included where appropriate. The questionnaires and tabulations for the mission questionnaires are in appendix 3.
- Field visits. One evaluator briefly visited Russia, Armenia and Macedonia in July, 2000. The second evaluator visited Mali, REDSO and Nairobi in August, 2000. The evaluators interviewed mission managers, mission directors, program officers, team members, and evaluation officers where they existed. They also visited partners that had either done evaluations or been evaluated. They reviewed mission strategic plans, R4s, and, in detail, the reports of the evaluations that had been conducted. Interview schedules and lists of reports found during fieldwork are in appendix 4.

The evaluators

Cynthia Clapp–Wincek worked as an evaluator in USAID for 14 years. Her time was split between CDIE and the Africa Bureau where she was the evaluation officer. The last decade has been spent consulting on evaluation, managing for results, and strategic planning.

Richard Blue is an evaluation consultant and trainer specializing in the area of foreign assistance programs, especially democracy and civil society development. From 1990 to 1996 he was the Representative for The Asia Foundation in Bangkok, covering the Asia Foundation's programs in Thailand, Cambodia, Laos and Vietnam. He served for 15 years with USAID holding various positions, including Associate Assistant Administrator for Evaluation and head of the Impact Evaluation Program. With Molly Hageboeck, he helped create the Center for Development Information and Evaluation. He retired as a Senior Foreign Service officer in 1994. As a consultant, he has participated in or led a number of evaluations and evaluation-related studies for USAID.

II. USAID’S RECENT EXPERIENCE

To assess USAID’s recent experience with evaluation, we begin with the basic questions:

- How many evaluations are being done?
- How good are they?
- Are they making a difference in USAID programs?

A. How Many?

1. The numbers

It was the report on the numbers of evaluations in the DEXS⁴ that prompted this study to be undertaken. Those numbers were:

Year	Number of Evaluations
1993	425
1994	497
1995	342
1996	240
1997	139

Source: Katrina Greene, “Narrative Summary of FY97 Evaluations,” USAID/R&RS, 7 January 1999.

We counted up more recent evaluations and found:

Year	Number of Evaluations
1998	104
1999	68
1999 projected	138

Source: DEXS

This looks like a drastic drop but several immediate clarifications are required. First, we need to point out that experience shows that it takes several years for evaluations to be sent in, catalogued, and show up in the system. The 68 entries for 1999 will eventually double if past patterns hold. But even in the mid-one hundreds, this would still be a significant decline from the 497 in the peak year of 1994. Second, the old and the new numbers turned out to be apples and oranges, because of discrepancies in counting methods.

Although we did not have access to how the earlier numbers were developed, an updated version of the same raw data (Bibcodes 22 and 24 from the DEXS) was reviewed. For purposes of this study, 1995 was used as the base year and 1998 and 1999 were used as the

⁴USAID staff tend to say they are sending reports in to ‘CDIE’ or ‘the USAID Library.’ Although these are related, they are not in actuality the same. See appendix 2.

study years. Two study years were used because 1999 data was too new to be complete and 1998 was too stale. Duplications and evaluations that did not meet our definition were discarded.⁵

Total Numbers of Evaluations		
Base Year	1995	253
Study Years	1998	104
	1999	68
	1999 projected	136

Source: DEXS

The initial concern was diminished because the numbers of evaluations conducted by the Agency hadn't dropped as precipitously as we thought. But even if the number of evaluations for 1999 reaches the projected level, this is still a *major drop*.

The most straight-forward explanation for this drop was that evaluation reports were not being sent in. During fieldwork, one evaluator brought back 22 reports from Africa that had not been submitted and there were others he couldn't carry. The evaluator that went to E&E found 11 reports that had not been submitted. Recognizing that we visited 6 USAID posts out of 74, we conclude that reports not being sent in to Washington is an important explanation of the apparent "decline" in the numbers of reports.

The requirement to send evaluation reports to the DEC is not widely known either by USAID staff or partners. Some USAID staff that do know, fall back on the letter of the law (i.e. 1995 ADS) that says partners are required to submit final reports. Partners have become unwilling to "finalize" evaluation reports and send them in—they see that as the job of their clients, in most cases the missions. Several mission staff reported in interviews that they didn't want partners sending in evaluation reports. One instance was reported in which a partner submitted a final report according to the 1995 guidance only to have the mission ask DEC to take it out of the system because they were concerned about accuracy. The report was not an evaluation in this instance but it demonstrates the problem. The partner had sent the report to DEC in good faith but ended up being embarrassed.

⁵*How we counted:* We discarded duplications, e.g. sometimes a summary would be listed separately from the full report. Then we looked at the entries in terms of our definition of evaluation. This issue rose particularly in deciding whether to include project assistance completion reports. Some PACRs are quite similar to evaluation reports, depending on how they are done but generally no effort is expended to address the question of "why" the changes occurred. Holding to our primary definition of evaluation, PACRs as a group do not fit. We also do not include CDIE evaluations in our analysis because they are not directly reflective of the same kinds of incentives and disincentives that affect the rest of the Agency. Because of the definitional issues and coding issues, not to mention the issues of what got lost or was never sent in, these numbers should not be considered sacrosanct—but they do show trends and that's what we need to understand.

Another possible explanation for the decline in the numbers of evaluations sent to Washington was that the 1995 ADS also effected a move away from ‘projects’ to ‘strategic objectives.’ With few exceptions, what we found were activity evaluations very similar to the old project evaluations. Therefore there did not seem to be the shift we expected to SOs as the unit of analysis. We concluded that the fact that the base number of SOs might have been smaller was less the issue than the evaluation approach missions took to those SOs. This is discussed further in section C.2.

The Agency’s change in evaluation requirements effected by the 1995 ADS was intended to have an impact on the number of evaluations and did have a noticeable impact. One evaluation officer in the field referred to this as “evaluation by exception.”

I believe, per the earlier version of ADS 200, that evaluations have a place (e.g. when monitoring identifies a problem or when there are significant lessons to be learned); but they are expensive and should not be used to cover for poor activity monitoring.

The change to evaluating when needed helps keep the focus on producing results and emphasizes other mechanisms for obtaining the information needed to make course corrections. However, missions do need to have formal mechanisms (e.g. annual or semi-annual portfolio reviews and/or a management control review committee) to assist mission and SO management identify when an evaluation, or other intensive review of performance and assumptions, is needed.

On the whole, people are very pleased not to be doing evaluation mindlessly. In one questionnaire, the respondent said, “The process has eliminated the need for much evaluative work” and then he goes on to outline three key reasons they should be done saying:

but the practice of impact evaluations done a bit after the completion of a major project or program should be supported/continued. They are indeed useful for strategic planning. Evaluations should also be done when a decision is pending on whether or not to continue or extend an important project or program (or intermediate result) for a significant period of time. Missions can use evaluations creatively to assist in strategic thinking but they are costly and take effort so their use should not be overdone.

Other factors related to the drop in the number of evaluations can be gleaned from reviewing patterns of evaluations submitted. One intention of the changed 1995 ADS guidance was that fewer pro forma evaluations would be done. The pro forma approach prior to the guidance change was that each project have a midterm and a final evaluation. We looked at what were called ‘midterm’ and ‘final’ evaluations to get a comparative sense of the types, although titles can be deceiving and many reports do not specify.

Types of Evaluations			
	Midterms ⁶	Final	Other
1995	92	94	67
1998	12	41	51
1999	9	21	38

Source: DEXS as analyzed by authors

The decline in the number of midterm and final evaluations was expected, even welcomed as the right response to the 1995 change in the guidance. The expectation was that time and resources would be saved by avoiding so many pro forma, empty evaluations. Forty-two people reported in the questionnaires that dropping the universal requirement saved time and eased their burden.

In general, the trend toward fewer formal evaluations in USAID is a positive one, allowing USAID to avoid the rote, pro forma, expensive and on occasion fruitless evaluation exercises of the past.

This was expected to free up time and resources for fewer, more comprehensive evaluations that focus on the results of Agency-assisted efforts. Although there were examples of this, it is *not* predominantly what was found. It most certainly does not describe most of the mission evaluations submitted to DEC that we reviewed or the bulk of the reports we collected in our fieldwork. In the Africa fieldwork, these were nearly uniformly midterms and finals done by partners as planned in their contractual/grant agreements with the missions. There were also “special” evaluations—largely initiated by the partners. In the E&E fieldwork, the picture was more mixed. The two smaller missions had done very little formal evaluation work but what had been done did largely fit this model. Although Moscow had some fairly routine evaluations of this type, their program was clearly initiated predominantly by the mission, with a greater mission role in defining what issues the evaluation addressed. Armenia’s informal assessments showed the same active learning and application of lessons as Moscow but not through formal evaluations.

The interviews in Africa showed that the old midterm/final approach continued but was delegated, or wholesaled, to partners. This was true to a lesser extent in E&E but it was of concern that this pattern would be as prevalent in the other regional bureaus that, like Africa, have not had the radical differences in operating approach as that found in E&E. This concern is one that is revisited throughout this report.

On the whole, what we found in the database and the fieldwork continued to be largely the pro forma, although not rote, evaluations that looked much like those of the early 1990s. They

⁶Not all 1998 or 1999s came from the Bureau for Humanitarian Response (BHR) which kept the “midterm” requirement. See Appendix 2 on mechanics of evaluation for further discussion.

captured a lot of valuable information but probably only a couple were strong enough to meet the initial expectations.

Almost none of the strongest evaluations were midterm or final evaluations. Evaluations that were not midterms and finals would either have been special evaluations, in response to a need, or program evaluations, including several strategic objectives or activities for lessons and/or impact.

Special and Program Evaluations as Per Cent of Total⁷	
	Special/Program Evaluations
1995	67
1998	51
1999	38

Source: DEXS as analyzed by authors

Assuming the time lag for reports to be submitted and processed, it can be predicted that the final total for 1999 would be closer to 76. Most of these were initiated and carried out by Washington bureaus, although there are notable exceptions. Two exceptions include the World Education report from Mali and the small and medium enterprise evaluation in Russia.

The intent of this evaluation was to focus on mission evaluation, though it is clear that they are actually done by the missions *and* partners. In fact, much of the evaluation work involved the Washington offices of the regional bureaus, as well as Global and BHR.

Sponsors of Evaluations			
	Field Only	Field Plus Washington Offices	% Total
1995	92	161	36
1998	32	72	31
1999	23	45	34

Source: DEXS as analyzed by authors

This is in accord with the 1995 ADS policy that encouraged “collaborative” evaluation, involving multiple actors. The guidance emphasized other donors, host country governments and NGOs, but included all parts of the Agency. Indeed, the policy turned offices in Washington into operating units, clarifying that they were responsible for evaluation as well. Roughly two thirds of evaluations have consistently included the participation of these Washington offices showing no real change due to the policy and guidance.

Only two of the 1998/1999 evaluations we reviewed were “collaborative” beyond USAID and partners, as we have been using the term. In both cases, host governments collaborated on the evaluation. In one case reported to us, a regional organization partially funded by USAID produced evaluations that were sent to DEC. Because USAID support was not adequately

⁷These represent all evaluations in the DEXS database that were not called midterm or final.

clear, DEC declined to catalogue them. Some collaborative evaluation may be masked by the limitations of the Agency to capture everything in which USAID plays a role.

In conclusion, the field, mission and partners combined, generated considerably more evaluation work than was sent into DEC and captured by the system. Although there was more than we initially thought, there was *less formal evaluation* work being done than in the mid 1990s.

2. Are there enough evaluations?

Perhaps the most striking finding from the questionnaire of USAID staff was that 32 percent said that USAID should be doing more evaluation. Eighty percent (80 percent) also said that they carried out evaluations “as needed.” There are several reasons for this apparent contradiction:

- *Need* is a relative concept. When staff actually focused on the *need* for an evaluation, they most often managed to get through the process and get the evaluation done. But at the same time, as this questionnaire got them to focus on evaluation, they realized that their scarce time and resources had kept them from getting to all the evaluations that would have been useful.
- While the Agency has focused on building a results monitoring system as part of reengineering, evaluation coasted along on autopilot. Evaluations were no longer required allowing missions to take a longer view. Rather than evaluating at the end of a five-year activity, they shifted their sights to the new strategic objectives (SO). One could say this defensibly allowed them to postpone evaluations for a time. The time is running out.
- Perhaps the single most important explanation for the contradiction in questionnaire responses is between USAID staff that are comfortable with partners or Washington offices driving their evaluations and those that are not.

Many people pointed out that evaluation was just one form of retrospective information gathering. Missions were doing more monitoring and more “assessing.” Opinions were strong, but mixed, on whether the existing combination of evaluation and these other forms was a step forward or a step backward.

It should be noted here that this is not a report on *monitoring*. It takes a systematic look at monitoring, but only as playing a role in explaining what was happening with evaluations. For this report, the focus on monitoring is used to help explain USAID’s recent experience with evaluation. The questionnaires were eloquent on this topic.

Monitoring has squeezed out evaluation.

We don’t do evaluations anymore. —From a field visit.

We have gone to the other extreme in not conducting evaluations. We seem to be depending on in-house, often subjective judgments of implementing teams staffed by PSCs and grantees/contractors as to the soundness of the programs/activities. An external assessment or evaluation mid-way through the SO and Strategy is essential...It has become easy to hide under ADS guidance, avoid the hard work that involves evaluations and tough decisions that may have to be taken on the basis of evaluations.

More evaluations are required because annual three-page R4 descriptions to describe what is happening are insufficient for mission management decisions regarding where best to allocate its limited funding to achieve the most significant impact.

I firmly believe that the proper use of results monitoring largely obviates the need for evaluations—unless communication and relations break down and there is a need to use them as a tool for engaging counterparts and partners.

Any good organization assesses its performance and uses that information to plan future activities. Unfortunately, I don't think USAID does a good job of analytically assessing the results of its programs on an ongoing basis. Without such ongoing analysis the more formal, independent evaluations are much less useful. I think both types are needed, and that we need to strengthen both.

The variation in opinion was shared by the variation in practice. Appendix 2 further addresses how the pieces fit together.

Twenty percent of 98–99 “evaluations” reviewed did not have evaluation in the title. Most often they were called assessments. Assessments are often broader than an evaluation, such as a sector assessment that looks at an entire sector and not just what has changed due to USAID’s programs (as an evaluation should). But, the name “assessment” is used widely and vaguely. Assessments are valued to the extent they tell USAID managers about the trends and dynamics in a sector’s political, social, and economic environment. Sometimes these assessments are “evaluations” in disguise, avoiding the negative stigma of an evaluation, and imposing no requirement that the findings and conclusions be widely shared.

I have avoided doing ‘evaluations’ preferring ‘assessments’ instead. Reason? An evaluation carried a formal set of baggage that includes the Project Evaluation Summary facesheets (PES), among other requirements. The PES was structured to force follow up (in 6 months or 12 months) on actions EVEN when it was an end-of-project evaluation, when the mission was closing out, etc. After completing the PES, some clown in Washington, maybe a year later, would bug you about the follow-ups of the closed project! There were not ‘assessment’ officers or Assessment Divisions within the mission. Assessments accomplished more for the project officer and with less risk (evaluations, at least in mission ____), were threatening and often precursors of project reductions. I was delighted when the evaluation idea was left to the project management structure.

I have perceived the need to do more quick assessments, rather than large evaluations. Time and budgets are always constraints.

Armenia mission was an excellent example of a mission that was strong on learning. But, they weren't as strong on documenting. There are many examples of “assessments.” The style in the

mission seemed to be for managers to participate in the assessment of their own programs. They were quite systematic. For example, the democracy assessment began with a group of people in the mission sitting down and deciding on the organizations to look at and the questions to ask. Then 3 mission staff carried out the assessment, and decisions were made based on the assessment. However, the only written product was a memorandum to the mission director. The program officer said “there is probably paper we should send into Washington but this is more useful and more efficient.” When pressed by many demands and priorities, ‘useful and efficient’ is the right choice. The loss is the documentation and sharing of lessons.

On the other hand, we saw many examples of “assessments” that were as expensive and time consuming as evaluations. In Kenya for example, three democracy assessments were done over five years. These assessments could not have been any cheaper than evaluations, and were much more frequent. But they were not called evaluations and did not have the same expectations for external distribution or need to explain progress. No one other than the purchaser of the assessment has to use it or take responsibility for its use. Whereas an evaluation theoretically has a broader audience, may expose problems, and requires some kind of response.

Many questionnaire comments addressed the relative nature of the various forms of retrospective information gathering.

I think the appropriate level of evaluation is determined by context....there are evaluation zealots at large within the Agency who would want to see unreasonable amounts of resources, especially staff resources, devoted to evaluating. This creates the danger of an imbalance between actual implementation and reflection/assessment. I think we need to be careful that the evaluations we finance are cost effective in terms of how their results are used.

The previous requirement of midterm and final evaluations for USAID projects was NOT useful. They generally served the purpose of keeping consultants employed and continuing existing projects.

This mission has done few formal evaluations (as opposed to partner-managed assessments)—I strongly support more systematic evaluations, not of every activity, but to objectively determine performance and future directions.

There are situations where M&E are potentially high leverage investments and others where it would be a waste of time and money with all combinations in between. I did a major evaluation design for USAID which recommended not doing the evaluation because no one was going to use it. —Partner

It can be assumed that balance is the answer, but like all elegant answers, it’s hard to achieve. This evaluation concludes that the balance has tipped too far on the monitoring side and needs to shift back toward evaluation. The decrease in midterm and final evaluations did take place as expected, but other changes in the field did not mean that time and resources were freed up for

the broader deeper evaluations that were intended. Our scope of work asked us to look at the quality of evaluations for a fuller understanding of this issue.

B. How Good are USAID's Evaluations?

Assessing the quality of an evaluation report really requires an understanding of whether the evaluators 'got it right' i.e. accurately and completely captured and explained the changes they saw. It is impossible for us to ascertain "quality" in that sense. In the few countries we visited we could at least ask those involved in the evaluation process.

Most information on "quality" of USAID evaluations is about the reports. It is expected that USAID's evaluation reports are part of a learning process that extends beyond the immediate users. Since the report is all that the learner has, clear and concise reports are still worthwhile. Clearly distinguishing findings (i.e. the facts) from the conclusions of the evaluators is the first step in an empirical presentation. Recommendations should derive from the findings and conclusions, and not be simply unsupported assertions that rely on the faith in the evaluator.

The best reports we reviewed, and there were some excellent ones, reflected this careful empiricism. Although there was considerable range in the quality of the reports themselves, there was a tremendous amount of information captured in those reports.

Not being able to recreate the evaluations adequately to reassess their inherent quality, the field researchers:

- identified the evaluations in the missions visited, read the reports, talked to mission staff, talked to the organizations evaluated and in some cases the evaluators themselves;
- reviewed the evaluation reports submitted to DEC in 1998–1999; and
- surveyed mission staff i.e. sent them the questionnaire asking their opinion.

This provided sufficient information to draw conclusions on the following issues:

1. Quality

The mission questionnaire did not directly ask about evaluation quality but several mission staff commented. One said:

Quality is about the same as in the past; utility is higher because unnecessary evaluations have been dropped.

They are disappointing when they whitewash politicized projects.

We have had difficulties getting objective, comprehensive evaluations with actionable recommendations.

On the whole, fieldwork interviews with mission and partner staffs indicated that the quality of evaluations was sufficient for decision making. In one instance, USAID staff said that subsequent feedback from sub-grantees caused them to question the validity of an evaluation. However, they said that the evaluation was still useful to them, but they didn't feel they could fully trust it.

Another USAID staff person expressed concern that an evaluation didn't adequately address issues of cost effectiveness and that the data set had not been given to the mission. But again it was sufficient that his walls were covered with charts that were part of the process of using evaluation findings and conclusions to 'set a new direction' for the sub-sector.

In fieldwork interviews, most partners who were evaluated responded that methods and evaluators were good. This finding should be bracketed with the awareness that this was a situation where there was a general unwillingness to criticize in the presence of anyone related to their source of funds. One partner questionnaire reported:

We have had difficulties getting objective, comprehensive evaluations with actionable recommendations. . . . Nevertheless, evaluations have been used to resolve bitter catfights with the mission.

The World Education evaluation done in Mali (English and French) was exceptionally well done. Getting the questions right in the SOW, followed by hiring someone who combines evaluation expertise, a good sense of how politics work, and sufficient technical understanding seems to be the best recipe for success.

The quality issue is a difficult one because people have different expectations about what an evaluation is supposed to do. If a sponsor wanted an evaluation to "confirm" his opinion about a project or partner, his opinion of the quality will be a function of whether the evaluator agreed with his view. Quality is always a function of the methods, the team, the scopes of work and the amount of time and money available. These will be discussed in the next sections.

2. Methods

There is one issue that cries out for attention based on a reading of the evaluation reports: "methods" has become nearly synonymous with sending a team for 4-6 weeks to ask questions of the relevant mission staff, activity staff and representatives of local people. One mission staff person reflected this in their questionnaire response:

evaluations in the sense of empirically-based analyses. The expatriate "fly-in" evaluations, where a team comes for 2-3 weeks and bases its 'findings' on a series of qualitative interviews with key people, are not very useful.

Many of the evaluation reports we looked at that had been submitted to DEC did not include more than a paragraph or two on methods in the report. Those that did nearly all conformed to the “fly-in” approach described in the previous paragraph.

The USAID evaluations employed very weak evaluation methods. The “Potemkin Village” bias is almost inevitable when the subject organization selects the sites, the people to talk to and makes all the arrangements. USAID officers sanction these designs and help make the arrangements.

The 1989 *Review of the Quality of A.I.D. Evaluations* also raised issues about choice of methods so this is clearly a longstanding concern.⁸

3. The teams

What we really need is a list of who the good evaluators are so we know who to hire.
—Fieldwork interview.

We’ve begun to examine the proposed team members very carefully, rejecting those who were unsuitable. The contractors always seemed able to come up with someone better.

Each of these quotations highlights the importance of the team members that carry out the evaluation, and the difficulties that managers have had getting the evaluators they think can do the job well. Creating a list of good evaluators has been discussed before but ruled out by the lawyers as laying the Agency open to trouble.

External evaluators are expected to provide objectivity due to their independence. One mission manager reported an “evaluator” who arrived in country and asked what the manager wanted the report to say. His livelihood depended on subsequent work and he wanted to insure it. External evaluators are frequently professionals who do empirically assess and objective conclude, but external is not synonymous with objective.

Most of the evaluation team members listed in the 1998-1999 reports were not experts in evaluation methods, whatever their substantive expertise might be. Twenty seven percent had identifiable evaluation expertise, but too often even this was based on past experience with the “fly-in” approach to methods. However, not every team member has to be an “evaluation expert.” Evaluation expertise during the design is perhaps even more important than having a team full of evaluation “experts.” More evaluation expertise in planning and on teams is needed in USAID’s evaluations.

Hiring outside “experts” does not, however, assure objectivity. A good methodology with checks and balances to avoid bias is required for any team to do a good evaluation. On balance,

⁸Paul J. Hopstock, Allan C. Kellum and Malcolm B. Young, *Review of the Quality of A.I.D. Evaluations FY 1987 and FY 1988*, A.I.D. Evaluation Occasional Paper No. 19, USAID, May 1989.

in the context of careful methods, the participation of one or two people from within the organizations can strengthen the evaluation.

4. Scopes and guidance to teams

One outside evaluator who answered the partner questionnaire said “Good scopes of work are the first step in a successful evaluation: what does the contracting entity want to learn.” The scopes of work included in the 1998-1999 evaluation reports were checked (58 percent had scopes), and surprisingly little was said on methods.

This concern is clearly shared by evaluators that responded to the partner questionnaire:

At no time was the existence of USAID evaluation guidance mentioned; nor were any previous pieces of work cited as exemplary. Either or both would have been very helpful, in every case.

Variable quality of guidance. Some are excessively detailed, some very vague. The biggest weakness in guidance is a clear understanding of the purpose of the evaluation and how it will be used.

One mission questionnaire response said:

‘We do not have enough time to put the scope of work together’ conveying a concern about the implications this has for the quality of the effort. They went on to say ‘coordinating with CDIE or other Washington offices is seen as a burden. . . . and just takes more time.’

In general, I don’t think we have learned anything new from evaluations than we learn from good monitoring. . . . evaluations are not as useful as ongoing monitoring because they are too expensive, too short, too time consuming, and too difficult to implement when you consider that you have to bring an external team up to speed very quickly in a complex environment.

This is not surprising since the evaluation methodology written into most scopes of work was very much the same as the techniques the managers used to oversee their programs: field visits, interviews, documentation reviews. If evaluation is to become a credible and useful part of the Agency’s new learning strategy, then more attention must be paid to developing sound questions, good research designs and methods that will go beyond the kind of knowledge already at hand.

USAID staff we talked to did not fully realize that in writing scopes of work they were making the key methodological choices. Even when managers do attempt to do an honest and objective job, they seldom assign sufficient time or resources to the task. Evaluation experts are routinely hired *after* all the important methodological choices have been made such as:

- Who is on the team? A team can only be interdisciplinary if that's who they hired. Or a senior evaluation specialist supported by a junior research assistant offers greater opportunity to ferret out and plough through masses of existing data.
- Time is the key to determining methods. Short time frames can, for example, preclude visits to all regions. A limited statistical survey can be carried out, perhaps by a local social scientist, as a preparatory step for a team, but not when the team is already hired and in country.

When scopes are collaboratively developed and research methodologies are arranged in advance with the partners, it leaves little room for the hired outside evaluation expert to influence the questions to be asked, the design of the evaluation methodology, or the selection of the team with whom he or she carries out the work. In most cases, the first contact the manager has with the professional evaluator is when they arrive in country or as they are getting visas and clearances. At this stage of the evaluation, it is too late for them to share their expertise in a way that can affect most types of methods.

5. Time and money

Time (to prepare the scope of work, go through the competition process, etc.) and budget (IQC's are expensive, and sometimes do not have the local expertise) are always constraints.

I've noticed that while activity/project budgets have remained fairly constant during my 12 years with USAID, the cost of evaluations has increased appreciably, in part because of increased personnel costs (the cost of a consultant at the FS-01 level). Thus, in my experience, good evaluations require a larger portion of project finances than in the past.

Evaluations, particularly those with rich evaluation methods, are expensive and can take a long time. Each factor affects the quality of the reports and the overall usefulness of the evaluation.

The Moscow mission, specifically its nearly full-time monitoring and evaluation officer, met this challenge by contracting for multiple evaluations on its own. The mission went to a firm that had produced a good evaluation in previous work and the mission went back to the firm with a contract for a series of evaluations. There was a firm manager for the series of evaluations who played a quality-control role, in addition to the mission monitoring and evaluation officer. Smaller missions would have seen this as a luxury.

Still, there are ways to be more efficient and increase quality that are not beyond reach. More needs to be done to identify and disseminate best practices in evaluation methods.

6. Contracting complications

Completing the contracting for an evaluation takes significant time and effort. When staff talk about not having time to do evaluation, they most often mean, they don't have time to do the contracting before the need for the evaluation becomes irrelevant. It takes many months to develop a scope of work, carry out

the contracting, and field the team. Even contracting mechanisms designed to be quick, such as the Indefinite Quantity Contract (IQC), are taking 5–6 months. Not infrequently, USAID staff reported that it took about a year to get an evaluation done.

Evaluation IQCs were designed to try to provide assistance both with the identification of good evaluators and how long full, competitive procurement takes. Unfortunately, in 1993, IQCs were awarded with a ceiling amount that was reached before the projected end date: they “ran out” before their planned term and there was a period during which missions had no access to the mechanism.

A new set of IQCs began as this report was completed. Now the constraints again become how many and how quickly the contracts office can execute.

7. To criticize or not to criticize?

Human nature assures that very few final evaluation reports contain the most pointed criticisms. They were probably in the first draft but toned down as actions were quickly taken to address problems. Certainly before reports are circulated, explanations and modifications are added to dissipate the heat.

As students of human nature, we know that as soon as one critical thing is said, most people will begin their mental defense and not hear anything else. Constructive criticism requires looking at all sides of the story and does not read nearly as straightforwardly as audit reports. But it does take into account the complex environment in which we work and the extent to which the highly complex signals, incentives and disincentives can lead to what at first glance can seem to be stupid decisions. Besides, the evaluated are more likely to hear the evaluators.

C. Are Evaluations Making a Difference in USAID Programs?

1. Are evaluations used and how?

In the part of the questionnaire asking about mission’s use of evaluations for management decision making, 60 people agreed with the statement: “Evaluations have been critical to our understanding of activity performance.” That’s nearly two thirds.⁹ This demonstrates that people have had experiences with evaluations being used to make a difference.

The mission questionnaire asked staff to list the evaluations from the last year and answer “What person/event/issue caused your office to undertake the evaluation?”

⁹One qualification is that we suspect that people were generalizing from their experience as a whole rather than the last-year time frame we gave them (i.e. people who said there had been no evaluations in the last year checked this question).

- There were 64 instances of evaluations tied to design: program, strategy or activity. This included informing the design of follow-ons, redesign of activities needed when strategies changed, etc.
- In 39 cases, specific problems, issues or questions were identified as the reasons for an evaluation. Examples include: requests for more funds, responding to budget cuts, and external factors such as the introduction of government subsidies, economic crises and elections.
- There were 35 instances of evaluations being carried out because they were planned (“midterm,” “final”), as well as 24 that were designed to capture the impact and/or lessons learned of a completed activity (not tied to design issues.)
- There were only 7 cases cited where problems identified in the Results Review process prompted evaluations and 5 examples were included where evaluations were done to feed into R4 reviews.
- A handful of cases each were reported to assess partner performance and respond to host government requests.

Some of the comments on use of evaluations from the questionnaires include:

In [this LAC mission], the number of evaluations we conducted in 1999 is misleading in terms of how important evaluations are to the mission. For example, one evaluation led to a mission decision to terminate a contract and to revise the strategic objective.

Our SO has used evaluations to help us modify and refine activities.

Use of evaluations in this mission is mixed, depending on how committed is the team leader for managing for results and how the concept of information for decision making is understood among activity managers.

Fieldwork was an even richer source for identifying uses:

- The Moscow mission is using its evaluation of four small and medium enterprise activities to redefine the strategy of the mission’s program in business development to be more regional, more multifunctional, integrated and consolidated, and with more Russian ownership. There will be a new procurement—probably one larger mechanism. The office director is currently drafting a scope of work for a consultant to come and “set the direction” in a concept paper.
- The democracy assessment in Armenia was done because the programs were ending. As a result of the assessment, they decided to phase out direct capacity building for political parties and established a new direction of civic education. This laid the foundation for the annual program statement and for future competitive actions.
- One evaluation was done on regional trade/food security in East Africa that was helpful in developing strategy in lieu of an existing REDSO strategy. A second comprehensive

evaluation is planned to best refine the program in accord with the new REDSO strategy.

The perspective of evaluating SOs and strategies was clearly in the thinking of field staff, even if they hadn't always gotten to the point of evaluating. Several questionnaires reported on the "periodic, continual evaluation process of our SOs and overall portfolio." One said:

USAID/Ghana has institutionalized comprehensive biannual performance reviews involving all major mission offices, in which SO teams make presentations and answer questions from Senior Management and the Office heads about all aspects of their programs. These reviews are sufficiently thorough to obviate the need for most evaluations. . . . except in response to a specific need.

Clearly, we found many examples where evaluations are making a difference in decisions, designs, and debates. And this certainly explains why 79 percent of the staff answering the questionnaire agreed with "we carry out evaluations as needed and they provide useful input to our thinking." *Need* is, of course, a relative thing since 32 percent of the same group also said "USAID should be doing more evaluations." This meta-evaluation has identified many good uses of evaluation, but also tries to address the red flag raised by the 32 percent. The next section addresses constraints to doing evaluations that can make a difference.

The section on how many evaluations are being done (II.A.) discusses the important signal sent by the change in guidance that universal coverage was no longer necessary. The change in guidance is an important piece but not the only explanation found for what evaluations were being done and when they were used to make a difference.

Wholesaling

Not since the 1970s has USAID staff actually implemented the bulk of the U.S. foreign assistance programs. Downsizing and reengineering accelerated the "wholesaling" of development through the practice of creating big contracts/grants intended to accomplish an entire strategic objective because:

- SO teams were made responsible for everything under the 1995 ADS
- There were sharp declines in operating expense funds and direct hire staff
- Contracting constraints created the incentive to do fewer contracts.

Evaluation was wholesaled along with everything else. Evaluations were being done, but not just by USAID staff. In one mission we visited, USAID initiated one evaluation and partners initiated over 10. Partners do evaluation, in part because they are required to do so, but also, in some instances, because they seem to have internalized the need to do them as part of the organizational culture and formal policy. There were several PVOs during fieldwork that said they would evaluate whether required or not.

The major source of evaluation information (for Mali's North SO) is the Partner community working in the North with USAID support. The Partners such as Africare regularly do evaluations of their programs, with USAID approval and support.

This rosy picture is certainly not true across the board for every PVO, but we did see a prevalent pattern of partners carrying out evaluations and frequently even initiating them. PVO partners stand out in this regard; contractors seem more concerned with getting the data needed to meet their contractual obligations.

There are problems though. USAID “wholesales” a good amount of its implementation work to American PVO partners, usually in coalition with local NGOs. These organizations are true “retailers” of the development product, and are the first line of accountability for results. If they don't perform, USAID will not achieve its projected results.

This partnership is imbedded in the formal agreement between USAID and the PVO, usually in a form of grant called a Cooperative Agreement. This sets out the terms, conditions and parameters of the partnership and the financial grant which makes it work. If something is not put into the agreement, USAID has no right to come up with a new requirement later on.

USAID's “evaluation by exception” approach was in direct contradiction to the cooperative agreement structure. According to the 1995 ADS guidance, evaluations were to be initiated as a response or reaction to a problem that emerges in an otherwise calm sea of unimpeded progress towards results as measured by the monitoring system. But monitoring and evaluation officers reported a number of instances where the guidance on cooperative agreements was in conflict with the guidance on evaluations and the evaluations were not done.

This would have caused greater difficulty but for the fact that all cooperative agreements funded by BHR were subject to BHR guidance requiring evaluation to be written into each one. Because many partners that work with USAID tend to have agreements with BHR as well, some were not aware that the universal requirement was dropped for non-BHR programs and planned evaluations thinking they were required.

This inflexibility issue existed for SOs with big contracts with for-profits as well. With performance-based contracting, the contractor was driven by the need to meet whatever benchmarks were initially set. They did not have the flexibility. As one experienced USAID evaluator said, “Managers who might have done small, issue-driven studies were stopped in their tracks.”

This was reported in fieldwork as an issue of finding funding in the budget that isn't already allocated. Willingness was expressed by several contractors, but the inflexibility resulted from the funding being 100 percent allocated. Their own time was already allocated to meeting benchmarks. They reported being amenable to responding to unforeseen needs but it required a contract modification to do so. Modifications to contracts are very difficult to accomplish.

Front line managers have a natural desire to understand what is changing with their programs and why. They make choices daily and want information to help them do so. USAID staff are not usually the front line managers anymore—they manage partners. In many cases, the Agency is managing managers who manage managers (i.e. USAID staff manage PVOs or contractors who manage sub-grantees or sub-contractors). Nonpresence countries are another case where Agency staff are divorced from the direct hands-on role in development. In that instance, the USAID staff do not even reside in the countries where they direct work.

Physical or managerial distance from activities changes the managers' priorities. It is not that USAID staff do not want to know the results of programs, but the burning desire is more balanced by the risks of drawing attention to shortcomings. These risks may include: losing control, your reputation suffering, having a budget cut (like in the late 1990s), etc. When the requirement for every activity to be evaluated was dropped, a rather insidious reworking of priorities took over—not ill intentioned, not lazy, just responding to the incentives and pressures they faced.

This is why in so many instances, the momentum to initiate, carry out and use evaluations shifted to the partners. In actuality, distance actually increased USAID staffs' need for evaluative information to make decisions. There was evidence that this has been recognized at REDSO where they manage several nonpresence countries. This recognition needs to spread more widely. But for evaluations to make a difference, there are other practical constraints that need to be addressed as well.

Privatization of Evaluation

In a privatized evaluation system, knowledge becomes a commodity. In one mission, an “impact assessment” loaded with good data was marked private information not to be reproduced or disseminated. The profit-making partner that carried out the assessment argued successfully that there was a new procurement coming up and use of this information by others could diminish their competitive advantage. The mission said the piece will be distributed when the procurement is completed. This could cause a delay of 6–9 and is clearly in contradiction to Agency procurement policies. In this swiftly moving environment, what will be the losses from information being one more commodity in the competition for government contracts?

However, there is also evidence of results from an evaluation in one country affecting what happened in another. FSVC Macedonia changed their monitoring because the organization had learned from an evaluation carried out in Russia. The Macedonia country director shared a copy of his organization's annual report with the results of another Russia evaluation. Although this cross-border sharing is excellent, in both examples, the learning is taking place within one organization and not between partners.

Another aspect of the “privatization” of evaluation is the need to contract for teams. The contracting process affects evaluations. One manager, when asked who made the final decision

on whether an evaluation would be done, said “the contract officer.” That is literally true: when an evaluation isn’t already written into an activity’s mechanisms, and if the contracting officer doesn’t approve or can’t get the contracting done, there will be no evaluation, regardless of what expectations are in the guidance. Facing serious resource constraints like the rest of the Agency, scarcity of contract officers’ time has become an increasing constraint. Even IQCs, designed to be the Agency’s quick contracting mechanism, were reported to take 5–6 months to sign.

A number of people reported the difficulties with the contracting rules to avoid conflict of interest. In other words, the organization that carries out an evaluation of an activity or SO is not allowed to bid on any follow-on work. Both USAID staff and partners feel their hands are tied:

The restrictions on contractors’ eligibility for subsequent work virtually guarantee that firms who know a subject will not bid on the evaluation. It’s just too much trouble to get an outside evaluator and not worth doing if you don’t. —A mission director.

Conflict of interest rules are draconian. —A partner.

There are political implications of USAID wholesaling its implementation and evaluation to its partners, both nonprofit and for profit. One is that the decisions of what USAID programs are evaluated are functionally no longer USAID’s. USAID’s partners are often strong, well-developed institutions with roots in important constituencies in the U.S. They have access to Congress, are well organized as a group, and provide one of the few sources of political support for the foreign aid program in the U.S.

This was worsened in those situations where the partner had significant political clout. Most reported cases of a powerful partner avoiding evaluation were for democracy and governance PVOs, but there were others as well. Instances of U.S. Senators’ names being invoked, the Administrator’s reported intervention, and general concerns of more senior managers “afraid to stick their necks out” were some of the problems reported. In one case, a mission proposed a “participatory” approach to evaluation to draw a politically powerful partner into the evaluation process. The partner “couldn’t find the time” to schedule the evaluation over a period lasting many months. There were few of these instances but they were very serious.

Decreased Priority of Evaluation

Because evaluation was no longer a requirement, the practical implication was that it was no longer a priority. As one USAID staff person reported:

Ninety percent of my time is taken up with requirements. I have a hard time getting evaluation to the top of the 10 percent.

By contrast, the Government Performance and Results Act of 1993 created great demand for building stronger “results” monitoring systems due to the need to feed an annual report on Agency results to Congress. Taken together with the Agency’s reengineering, monitoring was

the priority. Monitoring was not only required, it was discussed at mission Directors' conferences, TA was offered and monitoring was discussed in a blizzard of cable traffic. What has been referred to as "monitoring" is frequently the "R4," the nickname of the key reporting document in the system. The emphasis on monitoring distracted missions from carrying out and using evaluations. In questionnaires and in the fieldwork, many staff reported that they felt forced, for practical reasons, to trade one off against the other.

Increased Risks

Another well-intentioned reengineering innovation was to link mission budgets to positive results. Based on monitoring reports, staff in Washington allocated mission budgets according to reported progress. It became risky to report anything other than success. This success-oriented environment played a role in evaluation decisions as well—if you don't ask, you don't fail, your budget isn't cut.

Use of evaluations in this mission is mixed, depending on how committed the team leader is to managing for results.

It has become easy to hide under ADS guidance, avoid the hard work that involves evaluations and tough decisions that may have to be taken on the basis of evaluations.

With growing recognition of the unwelcome distortions this caused, guidance was changed again in the late 1990s. This linking and delinking of budget has bred cynicism.

[The questionnaire for mission staff] should have given respondents more chance to identify the advantages of in-depth evaluations over *the hopeless, endless, process and gamesmanship of R4s* [emphasis added], performance monitoring, and promotion of a self-feeding "indicator industry" to suck up our development dollars.

Evaluation Skills and Training

The 1995 guidance decentralized authority, making evaluation a function of each SO teams' decisions. No central authority was responsible for what got evaluated. To carry out their evaluation responsibility, each team leader needed substantial knowledge and skill:

- a good understanding of how to decide when to evaluate
- sufficient understanding of evaluation methods to write a good scope of work
- enough savvy to know how to work the system to make an evaluation happen (identifying a *good* team, contracting, etc.)

Our review of the evaluation reports submitted to DEC demonstrated that the biggest weakness was that SO team members were not, as a group, adequately well versed in evaluation methods (as indicated in section II. B.).

With this widened need for specialized skills, evaluation training and technical assistance should have been increased. The reality was that the Agency began a five-year moratorium on training

and TA was addressed almost exclusively to monitoring, data collection and strategic planning issues.

CDIE's Development Experience Clearinghouse

CDIE holds the formal responsibility for expanding the learning circle both through their own evaluations and “capturing missions” work as well. For the latter, they created the Development Evaluation Clearinghouse to collect, computerize and publicize evaluation and other reports. This evaluation that you are reading was prompted by the noticed decline in the number of reports being submitted to DEC.

One program officer guiltily pulled the guidance for sending reports to CDIE from a deep stack on his desk. He knew he needed to do it but hadn't been able to find the time. Recognizing what else was in that stack, I couldn't blame him.

As indicated in section II.A., fieldwork demonstrated that sending evaluation reports to DEC was so low on mission's priority lists that most didn't get to it and some staff (particularly PSCs) didn't even know it was a requirement. (Guidance for submitting documents is in ADS section 203.3.9).

Field staff recognized this as a lost opportunity:

I can see the problem of not sending reports in. I'm designing a new social transitions program for the mission and sent a request to CDIE. So few evaluations have been done in the five years that USAID has been doing this type of program that CDIE had nothing. I had to contact individual missions that had these programs and it was very time consuming.

2. Are we evaluating the right things?

Evaluations can only make a difference in aid programs if the right things are being evaluated. There is a lot of evaluation going on but is it the right type of evaluation?

Need for sectoral/impact evaluations

We do know from comments made on the questionnaire that some of the people that said “do more” evaluation particularly wanted to see more “sectoral evaluations” or “impact evaluations.”

I think you need to focus attention on reviving Agency sectoral and impact evaluations which have been abandoned since the 90's. R4s probably offer some good comparisons of what's working and where we have a need to learn and replicate lessons.

Why not . . . go back in time and evaluate projects that have closed for 8–10 years, some in countries where we no longer have missions? I managed a first-ever granting of land use titles of hill-tribe villagers in national forests in Thailand. . . . It would be interesting to see what became of that titling effort. And the progress of the Lam Nam Oon irrigation project?

I think we need to do more retrospective evaluations. We have two SOs that have a more than 10 years history. We tried to interest CDIE in both these exercises but unfortunately did not get much joy. This kind of review is critical to agency learning.

The mission supports the use of evaluations to supplement the results monitoring process. The process has eliminated the need for much evaluative work but the practice of impact evaluations done a bit after the completion of a major project or program should be supported/continued. They are indeed useful for strategic planning.

We did find examples of “sectoral” or *program* evaluations such as the small and medium enterprise program in Russia. This evaluation pulled together findings and lessons learned on four similar but unrelated activities. Other examples included: the program evaluation of the economic rehabilitation activities in the Sudan, Hurricane Mitch response, media transitions supported by the Office of Transition Initiatives, environmental programs in Poland, and environmental policy dialogue worldwide. This list is probably not exhaustive but close.

Most other “program” evaluations were of individual contracts/grants/cooperative agreements managed by one entity with many sub-activities. A number of “program” evaluations were done of the centrally managed contracts/grants/cooperative agreements in E&E such as the ABA/CEELI evaluation.

Having excluded CDIE evaluations, nearly all of the evaluations found were of one contract/grant/cooperative agreement. Many of these have gotten larger and are approaching a sectoral scope. Therefore, some offer the potential for similar lessons and insights to the types of evaluations the questionnaire respondents mentioned. Unfortunately, the predominance of the fly-in methods has limited their usefulness.

An evaluation of nonformal education in Ecuador looked back after 25 years but that was the only *ex post* evaluation we found. In spite of our understanding that development is a long-term process, the pattern found was that evaluation work responds to the more immediate pressures of the day.

Success stories

Comparing the low apparent priority of evaluation to the quite striking awareness of the requirement for “success stories” and how these two have become intertwined is another cause for concern. On the whole, knowledge and awareness of the need to produce success stories was more widespread than of the need for systematic unbiased evaluation work.

- In one mission that prides itself on following guidance, most of the evaluation scopes ask for success stories as an appendix. One of its evaluations, when constrained by time and money, chose to concentrate on the more positive of the two cases.

Every year, a cable goes out from USAID/W to all the missions asking for success stories. One mission evaluation officer reported that that was the only time he ever heard from the bureau’s evaluation officer.

- Many staff in another mission expressed the view that the R4 document is a public relations tool, pointing to the constant call for success stories.

There were a surprising number of examples both in the questionnaires and in the fieldwork of evaluations being carried out to document success. This makes sense from the point of view of the SO teams and activity managers, because they can learn the most from the instances where the greatest was accomplished. But it is important not to discount what can be learned about the constraints that cause less progress. From the view of the overall Agency system, the tendency to evaluate positive cases is problematic. Not only can evaluation start to look like PR, but the Agency needs a mechanism to know *on balance* how their activities are doing. Reviews of all evaluations for the Annual Performance Report are probably not representative, taken as a group.

It was disturbing that the predominant focus on success was creeping into evaluation work. Missions were being responsive to the signals of priority, but “success” was far too prevalent for an evaluator to feel comfortable.

D. Conclusions

1. In response to the changed guidance in the 1995 ADS, USAID is doing fewer evaluations. Only a small percentage of completed evaluations is being submitted to CDIE’s DEC undercutting the sharing of lessons.
2. Most field evaluation work is focusing on activity performance, especially by implementing partners. R4 monitoring systems do assess progress at the strategic level. There is a disconnect between activity evaluation and SO monitoring that cannot be closed by more activity monitoring. SO leaders, in the main have not been planning for strategic evaluations at the SO level.
3. Much of the evaluation work that is being done is being done by partners. The partner organizations are learning from the experience; USAID is not. USAID managers, particularly the busy team leaders, skim reports and react to the most pressing actions needed. But only the most committed managers find time to do the evaluations at all. Most of this work does not circulate beyond the Partner and the SO team.
4. The methods used for USAID evaluations are dominated by the fly-in approach having a serious effect on the quality of the evaluations. As things stand now:
 - Scopes ask a team to come for 4–6 weeks and interview the mission, the activity staff, and ‘representatives’ of the local people. There isn’t enough time to get any kind of representative sample. The team frequently tells the USAID manager pretty much what he already knows.

This isn’t worth the time and money we are spending on it. With fairly minimal advice and guidance from evaluation experts, this could be significantly improved. This advice needs to quickly follow the decision to evaluate. Refining evaluations questions and identifying appropriate and timely methods are the key. By the time a team arrives, it is too late to affect the methods in any significant way.

5. The monitoring/evaluation balance has swung too much toward performance monitoring. Recognizing a) the advances of improved performance monitoring (although the job is not done), b) the considerable efforts of those USAID staff that have continued to carry out evaluations, and c) that evaluation has a different role to play, the evaluation system has been ignored too long.
6. Professional commitment to doing a good job is the primary incentive for good management including evaluation. There is little perceived linkage between either good SO performance or good learning and typical rewards such as increased budget, more staff, or personal advancement. In general, professional staff do not see any Agency-wide incentive to advance learning through evaluations.
7. Notwithstanding our concerns, considerable evaluation work is being used in decision making and program planning as managers struggle to continue to carry out quality programs. Sixty percent of those that responded to the questionnaire reported that evaluation had been critical to their understanding of activity performance. Fieldwork presented many further examples of evaluation being used in mission decisions, designs and debates.

The biggest structural change in USAID's work is the increasing reliance on non-career staff, few of whom have any long term stake in Agency learning. Implementation is "wholesaled" to a variety of development partners. Partners are interested in achieving results from the activity they manage. Policy and training still focus on direct-hire staff without adequate recognition that they do not do the implementation or the evaluation of USAID assistance.

Privatizing our monitoring and evaluation undercuts the flow of information as organizations (both profit and nonprofit) recognize information as an element of competitive advantage. Moreover, contracting and granting rules create serious barriers to learning from experience and applying lessons to improve our assistance.

The evaluation of USAID's recent evaluation experience appears to confirm that USAID can do less with less. The clearest picture of this evaluation is from the comment by the USAID manager who said he spent 90 percent of his time on things that are required and that evaluation had a hard time making it to the top of the 10 percent list. There are good, committed professionals in the field who need support to accomplish the type of programs they know they should.

III. USAID AS A LEARNING ORGANIZATION

The 1995 guidance encouraged evaluation efforts to respond to management need. Section II shows that this led to a lot of good, but fragmented evaluation work. The 2000 revision of the ADS guidance has taken a giant first step by stating that USAID is a Learning Organization as well as a Planning and Achieving one. This section goes beyond the retrospective analysis of recent evaluation experience to look at what we've learned through the lens of a learning organization.

A. The Learning Organization

The basic elements of a good learning organization are well known. They are:

1. Leadership commitment and involvement in driving and using the learning process.
2. A systematic process for setting priorities and the learning agenda.
3. An integrated system for sharing knowledge among various levels and parts of the organization.
4. Sufficient specialized staff and budget to harvest the knowledge and process it for broader consumption.
5. Incentives and sanctions that link performance and rewards to the learning process.
6. A process that insures accuracy, objectivity and relevance.
7. A recognition that good learning requires multiple approaches. There are no "silver bullets."
8. Recognition that knowledge that is not owned is knowledge that will not be used.

Based on what we have learned from our systematic analysis of the Agency's evaluations, how does USAID measure up as a learning organization?

Leadership and agenda setting

USAID participation in evaluation can range from active initiation, planning, budgeting and monitoring the work of an outside team, to active participation by one or more USAID officers on the evaluation team. Too much of the evaluation function has been shifted to partners for USAID to be a learning organization.

There is no process by which senior leadership sets the learning agenda and knowledge-harvesting priorities. CDIE establishes its agenda without much guidance from senior leadership or much response from other Bureaus.

Sharing Knowledge

There is little indication that knowledge is being processed and fed back into decision making in a manner that can be readily absorbed and used by senior and mid-level management. There must be a kind of “knowledge SWAT team” in CDIE or Global that can quickly cull, assess, and prepare “state of the knowledge” papers on key issues of direct interest to decision makers. Field evaluations and other formal assessments should be treated as part of the raw material of “development intelligence.” These must be supplemented with more broad, strategic evaluation. This raw material must be assessed, synthesized, and made usable in the same way the NSC and the President rely on the “strategic intelligence” to guide security and diplomatic decision making.

Expertise

Evaluations are a special part of the Agency’s current learning process that has grown to include strengthened (though far from perfect) monitoring and assessments. These are useful tools that need continued development and support. Insufficient training and technical assistance has been provided. The missions do not, and in the foreseeable future will not, have the skill, resources or incentives to capture the overall sense of the impact and effectiveness of Agency resources.

USAID officers’ participation on evaluation teams has become extremely difficult as staff resources have shrunk. Real learning takes place not through reading reports but by going out and asking the questions directly.

The constant call for more success stories with a deafening silence on evaluation puts USAID at risk of not doing “organizational learning” but “organizational PR.”

Incentives and Sanctions

Self-evaluation can be a powerful learning tool. Objectivity and good social science remain the key ingredients, but the purpose is to find lessons learned and to improve the quality of the work of all partners involved in the process. This approach does require outside expertise to both facilitate the learning as well as to insure that sound methods were used to generate the data, analysis and findings. Stakeholder participation is also a key ingredient in this approach. Bad news is not rejected, but when properly done, it is turned into constructive criticism and advice to all managers. Such an evaluation is not “done to” people, it is done “by people” who are presumed to be seriously committed to being successful, accountable and transparent. It is a tool for learning, and a method for teaching at the same time. Whether and how an organization or a coalition uses evaluation as part of the organizational learning strategy is itself a measure of the institutional capacity and maturity of that organization.

There is a place for evaluation as a punitive tool for investigating poor performance and improving accountability. Bad news with political consequences does get everyone’s attention

when other forms of analysis and inquiry do not. USAID, particularly in Washington, spends far more time reacting to IG Audits and GAO evaluations than responding to its own evaluation studies. Whether this is time well spent is another issue.

The 1995 guidance emphasized *management need*, i.e. that there be a management question that triggered an evaluation, but with such scarce staff and resources this slid into *problem-driven mode* i.e. the only questions pressing enough to trigger an evaluation were when a real problem was evident. This played right into the view of evaluation as a *punitive exercise*.

Policy and Relevance

Policy and guidance must define explicitly the difference between evaluations as punitive exercises, and evaluations that contribute to the learning process. USAID must do more to provide a framework of resources and mechanisms by which evaluations can be done, not because there is a “problem” but because there is now a “reason to build knowledge and to learn from our experience.”

It is true that there are many paths to organizational learning. Field trips, quarterly implementation reports, discussions with partners, other donors, local government officials all yield valuable information leading to an increased level of management comfort. Unless the manager is extraordinarily disciplined, this type of knowledge is rarely turned into “organizational learning.” A successful organization is one which understands the importance of “harvesting knowledge” from within, and from without. Self-evaluation can play this role. When combined with stronger evaluation methods that can provide new insights, evaluation can best add value to the learning process.

Learning

To be successful, the organization must have a system of procedures, rewards, and sanctions in place for supporting the knowledge harvesting process. It must be an integral part of the cycle of planning, achieving and learning (as well as training). The new policies in the revised 2000 ADS begin this process.

The ADS guidance itself says:

The process of assessing and learning requires that we go beyond meeting specific requirements that can be spelled out in a few pages of guidance. We must demonstrate a broader commitment to some basic principles and practices and encourage our partners to do the same.

Following are recommendations to get some momentum behind the changes in the revised 2000 ADS.

B. Recommendations for a Learning Organization

1. Agency leadership should convene a high level task force with one question: “how can we become the best learning organization possible?” Issues of organization, responsibilities and common understandings need to follow. Some examples: both CDIE and Global do best practices type studies; BHR guidance differs radically from Agency guidance; OP needs to participate in order to work towards addressing serious constraints posed by the contracting process; success stories need to be delinked from evaluations and an alternative approach identified (LPA).
2. Recognizing that missions will not be able to capture the overall effectiveness and impact of Agency programs, a strategy needs to be developed on how the Agency will more adequately meet these needs.
3. More training is clearly required. This system has been running on the professionalism, good faith and project evaluation training of Agency staff and partners. With the expected staff turnovers in the next five years, this will be increasingly insufficient.
4. An Evaluation Handbook directed to both USAID *and* their partners needs to be developed.
5. Evaluation policies need to be changed so that no bureau has blanket policies that directly contradict to Agency policies.
6. It should be someone’s job on an ongoing basis to market evaluation policies and the guidance captured in the handbook (i.e. “get the word out”) and to update the handbook on an ongoing basis.
7. The Strategic Objective has become the Agency’s primary unit of results expectation and management oversight. This is the level to which the R4 system is geared. Paradoxically, much evaluation work is going on at an activity level, much less at the SO level. The current ADS strongly recommends evaluations be conducted at the SO level at least once in the life of the SO. The Agency should consider upgrading this to a requirement after adequate attention from senior management, training and support. The requirement should be that all programs are covered at some point by an evaluation, not that each SO should be evaluated.
8. Every strategy submitted to Washington for approval should be required to have a section reviewing experience on what works, what doesn’t and why. These should explicitly include the mission’s previous strategic experience.
9. More attention needs to be given to timely harvesting and processing of the knowledge that is being produced, especially by Agency partners.

10. Participating in a strategic evaluation in someone else's mission ought to be an integral part of the terms of service of any aspiring USAID officer. As USAID does less and less of their own implementation, the Agency should be doing more and more of their own evaluation. People learn completely differently when they are out in the world asking the questions than they do sitting in an office quickly scanning a report.

Modern evaluation theory and experience demonstrate that evaluation can be both objective and constructive if the organization is willing to invest leadership and staff resources in the ownership of the evaluation process. Since staff resources are highly constrained and fewer evaluations are being done, the evaluations should be of the highest quality.

11. The quality of evaluation design and methodology must be improved. If evaluation research is to become a credible and useful part of the agency's learning strategy, then more attention has to be paid to developing sound questions, good research designs and methods that will go beyond the kind of knowledge already at hand. Good evaluations will challenge the conventional wisdom, not just confirm it.

In practical terms, this means drawing on professional evaluators to assist in the development of the key decisions about evaluation questions, research design and methodologies. Hopefully, USAID will have in-house expertise from which to draw to actively facilitate the process of use-based evaluation design. If not, USAID needs to find a contractual device for bringing in outside evaluation expertise at the stage when scopes of work, methods, and resources for the evaluation are being determined. Poorly stated questions and bad evaluation designs lead directly to poor evaluation results.

12. Standards and guidance for assessments must be developed. The assessments, with their quick backward look and a scan of the environment to decide how to move forward, make good sense in some contexts.
13. It is important to capture success but it needs a mechanism other than evaluations. Evaluations will always capture success when it occurs but linking the two distorts the evaluation process. An approach with guidance needs to be developed.
14. The role of mission evaluation officers should be reconsidered. Many missions have FSN evaluation officers by name that report being limited in their ability to carry out their job. How best can they be used and what needs to be done to support M&E officers where they exist.
15. Put evaluation back into the list of the Agency's priorities. The most senior Agency staff possible must demonstrate an interest in evaluation. Senior staff participation on evaluation teams would send a strong signal. Regional bureau senior staff could systematically debrief evaluation teams.

In Conclusion

There is much to be done to reestablish evaluation as a means by which “truth speaks to power.” More important, USAID needs to find a way to “use truth to empower.” Only then will USAID be able to say that it is a true learning organization.

Appendix 1

Statement of Work

I. Purpose

To analyze USAID's recent experience with evaluation in order to propose revisions in the ADS and Agency policies and provide guidance in appropriate methodologies for using evaluation to assess impact at the SO and interim results level.

II. Background

A thorough search of the USAID Development Experience System (DEXS) database in July 1998 revealed that the number of evaluation reports in the database has been steadily declining since 1994. Some of the decline may be due to a lag in the time that a document is produced and the time it is sent to the Development Experience Clearinghouse for inclusion in the DEXS. Part of the decline may be attributed to a change in Agency guidance contained in the Agency Directive System (ADS) that no longer requires conducting periodic evaluations, the results of which would be sent to the DEXS. More broadly, the issue that needs to be addressed is what is happening to evaluations in the Agency. It appears that as a result of the new guidance fewer evaluations are being undertaken. While the reform agenda clearly intended to reduce the number of unnecessary evaluations, it was never intended to devalue the importance of evaluations as a tool of results management. There appears to be a misunderstanding within the Agency of the new guidelines, the type of evaluations to be carried out, and the continuing need for appropriate evaluations. These studies will attempt to review the Agency's current experience with evaluation to determine whether revised directives or guidance is needed.

Statement of Work

The consultant(s) should review a representative sampling of operating unit evaluations undertaken since new Agency guidance on evaluations (contained in the ADS) was issued in November 1995. In addition, the expert(s) should carefully review the summary assessment of USAID evaluations in FY 97/98, completed in December 1998 by Katrina Greene of the Research and Reference Services. The consultant(s) should also review any evaluations submitted to the DEXS subsequent to that review.

Based on these documents, the consultant will formulate a questionnaire to "ground truth" some of the conclusions reached by the earlier desk study. The questionnaire will include a representative sampling of the entire Agency, but will focus on field missions. The consultants will utilize the existing desk study, a review of current evaluations in the DEXS, a review of R4s, a field survey, and interviews, among other approaches, to attempt to answer the following questions:

1. What has been happening with evaluations in the field? Are evaluations still being conducted periodically? What types of evaluations are being conducted? Why have the numbers of evaluation documents coming in to the DEXS been declining? Among the possible reasons for the decline, the following will be investigated:
 - Operating units giving priority to establishing strategic plans, building performance monitoring systems, and preparing R4s, with little time left over for conducting evaluations.
 - Operating units interpreting the reengineering guidance through the ADS as indicating that evaluations were no longer deemed either necessary or important to USAID/W.
 - Operating units misunderstanding the focus/purpose of evaluations within a reengineered Agency.
 - Operating units not understanding how to carry out evaluations when there is a failure to make discernible progress toward an objective or to determine impact.
 - Operating units understanding evaluations only in terms of those previously undertaken on well-defined “projects.”
 - Operating units perhaps undertaking more informal evaluations, or perhaps undertaking evaluation activities that are not so labeled. (Perhaps such efforts would be called “performance information reports,” “research,” or “sectoral analyses” rather than evaluations.)
 - Whether or not evaluations are being used to inform management decisions.
 - Whether or not evaluations are being used to test the validity of hypotheses and assumptions. (The consultant needs to distinguish between “performance measurement and monitoring reports” and “evaluations.”)
 - Whether or not evaluations are being undertaken at the mission level but are not being forwarded to the DEXS.

2. How should changes in USAID evaluations be characterized?

How can one evaluate the quality as well as other characteristics of USAID evaluations in recent years? How can one codify the changes to evaluations that have occurred since December 1995 in response to reengineering guidance?

How can one characterize current USAID evaluations?

- Scope (e.g., a country, region, multi-program, multi-country, etc.)?
- Focus (e.g., activity, IR or SO level)?
- Purpose (e.g., program improvement/performance issues, achieving results or objectives)?
- Sector (e.g., Agency goal/ objective area)?

- Type of methodology used (e.g., informal, rapid appraisal, sample survey, participatory, etc.)?
 - Who sponsored it?
 - Who participated in the decision to carry out an evaluation, prepare the scope of work, for an evaluation, carry out the evaluation, and implement evaluation findings?
 - What was the cost of the evaluation?
 - Do evaluations relate to the operating unit's performance and monitoring plan? Have performance monitoring issues been the impetus to conduct an evaluation?
3. What methodology is employed in evaluations?
- Are more participatory approaches being used, as suggested by the guidance?
 - Are partners involved in the formation of evaluation scopes of work?
 - Are more joint evaluations (using involvement from partners such as host country officials, other donors, NGOs, etc.) being undertaken?
 - Are program beneficiaries and stakeholders being involved by the missions in the evaluation process?
 - Are evaluations more frequently using rapid appraisal techniques, as the guidance recommends?
 - Are evaluations project oriented or results focused?
 - Do evaluations serve a complementary function to performance measurement systems?
 - Does the evaluation contain clear recommendations for action? Does the mission concur with the evaluation findings? If so, have the actions recommended in the evaluation been taken?
4. Are evaluations meeting management needs?
- Despite a reduction in numbers and other characteristics, the most important question is whether evaluations are adequately meeting management needs, both at the operating unit level and Agency-wide.
 - Do evaluations meet operating unit management needs for explaining unexpected performance and for making informed program decisions?
 - Are operating units doing evaluations as a result of: 1) periodic performance reviews that identify key questions / problems that need to be answered; 2) questioning the contributions of USAID projects to results or strategic objectives; 3) a need to question results frameworks, hypotheses, and critical assumptions.¹⁰

¹⁰For example, suppose performance-monitoring data indicates that farmers are adopting a new agricultural technology as planned, and yields are increasing but not farmer incomes. Suppose an evaluation finds that the reason incomes are not increasing is due to a lack of marketing facilities and good roads. The evaluation might recommend increasing program funds for facilities and roads as a means to improve program performance. If, on the other hand, the evaluation finds the reason farmer incomes are not increasing is due to dropping international prices for the crops, about which managers can do little, the recommendation may be to terminate the program.

- Are operating units using evaluations to assess the contribution of USAID program activities to achievement of specific strategic objectives?
- To what extent have missions been conducting evaluations for the purpose of assessing the impacts of USAID programs on strategic objectives? Have they been able to convincingly demonstrate attribution?
- Are operating units using evaluations for reporting and explaining results in their R4s?
- To what extent are findings from an operating units evaluations drawn upon and reported in their R4s when discussing and explaining performance, as required by the ADS even when individual evaluation results are no longer reviewed by USAID/W? Are many unexpected results going unexplained in the R4s? Has the failure to meet targets and objectives in the R4 triggered an “evaluation”? If so, what type of “evaluation” or other analysis is done? Are program decisions being made based on performance monitoring data without the benefit of evaluation recommendations?
- Are operating units using evaluations to determine progress toward Strategic Objective or Interim Results? Are missions able to use evaluations to attribute project or activity level outputs to higher level results and objectives, particularly at the SO or interim results levels?
- At the Agency level, is the decline in the number of mission-sponsored evaluations (or other aspects of these reports, such as their purpose, focus, quality) affecting CDIE’s ability to make a case of “plausible association” in the Annual Performance Report—that USAID programs are having a demonstrable impact on achievement of Agency goals and objectives (performance goals)?

This year’s APR is supposed to be using recent evaluation findings to make a case, in specific theme areas, that USAID programs are indeed contributing to achievement of agency (country-level) performance goals. To the extent that USAID operating units have been conducting program impact evaluations of the type—as proposed by the GAO (see section above), this may be a feasible approach. How is this effort going? Are useful evaluations/case studies being uncovered for use in the APR? What was the experience with last year’s APR review of evaluation findings? Was the emphasis last year more on lessons learned than on accountability for results?

III. Deliverables

The activity should be undertaken in two phases.

Phase I will involve the creation and delivery of a survey instrument to be used within selected offices and missions throughout the Agency. This document should be based on a thorough study and analysis of the information outlined in the Katrina Greene report as well as on an analysis of additional documentation obtained through reading selected evaluations found in the DEXS.

Phase II envisages 75 days of effort by the Principal Researcher and 20 days of effort by the Senior Advisor.

It is expected that one or more trips to the field will be required to carry out additional study after analyzing the results of the survey. Selection of missions for more in-depth study will be made in consultation with PPC/CDIE. For planning purposes, trips to 3 missions in Asia and 3 missions in Africa are envisioned.

Phase II will include the preparation of a report addressing the questions identified above, including the tabulated findings of the survey, an analysis of the results, and a description of the methodology.

A preliminary report, not to exceed 40–50 single-spaced pages (not including annexes) will be produced. It should be presented to PPC/CDIE no later than September 30, 2000, and should contain the following:

1. Recommendations to the Agency relating to the current ADS Section 203.5.6, including suggestions for modification, additions, or revisions, if needed.
2. A statement outlining the role of evaluation in the performance monitoring system.
3. A recommendation to CDIE relating to the need for additional guidance/methodologies on conducting impact evaluations.
4. If additional guidance is deemed necessary, an outline for a TIPS on evaluations.
5. A recommendation relating to actions that the Agency should consider to ensure that missions undertake evaluations.
6. A selection of lessons learned from FY 99 evaluations that can be incorporated into the Agency Annual Performance Report, if time and resources are available.

Four copies of the report, with an electronic copy in Word, should be delivered to the Deputy Director, PPC/CDIE, within four weeks of termination of the field trips.

Appendix 2

Mechanics of USAID Evaluation

Mechanics of USAID Evaluation

The scope of work for this evaluation focused primarily on missions, and the main report is slanted that way. Much of what we found applies to “operating units” in Washington as well. Due to limited time and resources, we did not survey Washington operating units in as structured a manner. We did conduct dozens of interviews in Washington between March and July 2000 to help us round out our understanding of USAID evaluations as a whole. The findings and conclusions about the Washington side of the picture is less about carrying out evaluations than it is about managing and supporting. Believing in the importance of learning from our experience, we begin by looking back at the history of evaluation in USAID.

1. History of Evaluation in USAID

As a learning organization, USAID has had flashes of brilliance that create enough momentum to sustain the system for years afterwards. Although not widely loved, the Logical Framework, developed in the early 1970s, initiated the formal planning-implementing-evaluating link that provided the elegantly analytic underpinnings of the Agency’s approach to learning. Narrowly focused on projects and fatally prone to static fill-in-the-boxes use, the LogFrame has nonetheless been recognized by the international community as an exemplary tool.¹¹

As then Director of CDIE Janet Ballantyne concluded in 1989, ‘For many years A.I.D. was in the forefront of donor agencies in its support for, and use of, systematic evaluations. More than a dozen countries have come to adopt a version of A.I.D.’s Logical Framework, a technique originally developed by the Agency in the early 1970s to support project planning and evaluation.’¹²

The 1970s were also the era of the “Spring Reviews.” These were supposedly annual efforts to bring together the learning on key development issues. The Spring Review on credit is widely considered to have changed the Agency’s approach to all credit programs.

As early as 1979, senior Agency management realized that project evaluations were not providing them with adequate information on the “people level impact” of the Agency’s programs. Soon after Doug Bennett took over as Administrator in 1979, he asked, “What do we know about USAID programs that have been completed? Can we say anything about the long-term impact of our work? To respond, the Evaluation Office (CDIE’s predecessor) created the Impact Evaluation Series. The ideas behind the new series were simple. They were to look at completed projects and find out whether there was a lasting impact and why, or why not. In order to make them readable for busy managers, the reports were to be no more than 15 pages and in a style of a “good *New Yorker* magazine article.” They should focus on issues of poverty as objectively and systematically as possible. They were to look at intended

¹¹The World Bank has recently implemented a logframe-based approach and the Japanese are considering it.

¹²Paul J. Hopstock, Allan C. Kellum, Malcolm B. Young, *Review of the Quality of A.I.D. Evaluations FY 1987 and FY 1988*, A.I.D. Evaluation Occasional paper No. 19, May 1989.

consequences and beyond to unintended impact. USAID professionals were the evaluators, with care given to avoiding conflicts of interest. They were supplemented by outside experts, particularly anthropologists and other country specialists.

The Impact Evaluation series was a unique experiment. It had the highest Agency leadership behind it. It strove for objectivity and relevance while remaining sensitive to resource constraints. It had a strategy for synthesizing and disseminating knowledge, and by using Agency professionals primarily, it sought to get the Agency to “own” the knowledge that was being created. When the Administration changed, the link between the knowledge harvesting and processing chain and the policy interests of the new administration was severed.

The 1980s were characterized by the efforts to strengthen and systematize USAID’s project evaluation system. In 1987, the Agency issued an Evaluation Handbook. Although this has been superseded by further guidance, it is still in use in at least one mission today “because it includes all the basics.”

Another important effort was the series of evaluation training programs that were conducted primarily in the late 1980s. This week-long training was organized around the themes of:

- Evaluation’s utility to management,
- Collaborative evaluation, and
- Program level evaluation.

The program evaluation emphasis was added to the later training, probably due to Africa’s changed legislation. Training programs were held more than once in each of the Agency’s regions, which successfully encouraged greater participation. Hundreds of people were trained, both Agency staff and partners.

By 1989, in an introduction to a *Review of the Quality of A.I.D. Evaluations*, Janet Ballantyne characterized USAID as having “fallen behind in its efforts to build and apply new approaches that seem most appropriate and useful to foreign assistance.” Having emphasized the institutionalization of evaluations, the pendulum was swinging back towards a call for new approaches.

The *Review* was just one of several assessments of USAID’s evaluation system. Former USAID Administrator Peter McPherson reviewed CDIE’s services and organizational structure in 1987. In 1988, a management assessment of CDIE was led by mission Director Julius Coles. Having been quite thoroughly studied, in 1990, the Administrator approved an “Evaluation Initiative” based on these studies and CDIE’s “The A.I.D. Evaluation System: Past Performances and Future Directions.” The “Initiative” resulted in greatly increased staff and resources and a return to an expanded set of evaluations.

In retrospect, the Evaluation Initiative of the early 1990s was hijacked by the Reengineering and the Government Performance and Results Act. The two combined to focus USAID's energies on monitoring "results" with an emphasis on "strategic objectives" rather than projects. The 90s were the era of "Managing for Results" for country programs, regions and the Agency as a whole.

Reporting results became a higher priority and the R4 (Results Review and Resource Request) was introduced. Again, not widely loved, the Agency's approach to monitoring has been recognized by a U.S. Government-wide community as ground breaking, although perhaps flawed in implementation (see IG/GAO audits).

2. The Planning-Achieving-Learning Cycle

USAID's recent evaluation experience is part of what the new guidance calls the planning-achieving-learning cycle. Our scope of work asks us to evaluate USAID's recent evaluation experience. Most of this appendix will address evaluation's role in learning but first there are several observations on recent experience with planning and achieving and how they affect the learning third of the system.

Planning

As USAID direct engagement in implementation and evaluation has diminished, *planning* is one area where the organization should excel. We have some anecdotal evidence that raises concerns.

One team leader reported: 'The Logical Framework, whatever its weaknesses, is not used now. As a result, too many assumptions are left unstated, cause and effect is not well structured, hypotheses about links between development interventions and expected results are not clearly stated or analyzed. This has had an effect on evaluations, because the linkages between activities (inputs and outputs) and results as measured by indicators are much less explicit.'

This sense that good planning (good design of development activities) had suffered was repeated elsewhere during the field trips and in Washington interviews, usually by USAID and Partner officers who had grown up professionally under the old system. Several missions reported that they didn't use results frameworks but those were also missions where we would characterize their programs as a series of activities with a strategy wrapped around them. Good evaluation relies on clearly spelled out linkages between the activity outputs and the strategic objectives to know whether our investments made any difference or whether it was just serendipitous changes that caused the results.

In the Reengineered USAID, most planning has moved to the more *strategic*. In the questionnaire, we tried to get a sense of what evaluations were used for, so we asked questions about the role of evaluation in strategy planning. What we found is that many mission staff have

been waiting to get enough experience with a “strategic objective” under their belts to evaluate it. They think of it as a discrete unit to be evaluated much like the old project evaluation. They don’t think of it as part of an ongoing system to learn and understand. Therefore, we heard really too often, that strategies and strategic objectives were too new to evaluate.

Overall, the problem with the strategy as the unit of action is that it tends to be somewhat abstract, long term, and passive with regard to driving actions. The tendency seems to be to see it as something that was done in order to organize the work, but lacking in the kind of specificity that you have to pay attention to every morning when you come to work.

This reflects that the day to day work of staff tends to focus on activities. Strategies are the planning framework, but people still think of themselves as managing projects i.e. activities. This means that activities fall into a kind of planning void created by the shift of reporting and accountability strategies and strategic results. The connecting tissue is the indicators system, the numbers for which are created by the activities. Because the links to the strategy may not be well defined, it is now logically possible for a poorly designed activity to still generate the desired results.

Because of this weak design and because of lack of experience strategizing at the country level, strategies, within country programs and for specific SOs, seem to be constantly changing. Strategies do change for good reasons. Such as when the Macedonia mission shifted from humanitarian to developmental assistance only a few years ago. A pattern of always-postponed evaluation is a source of concern. It is important to think about what types of retrospective analysis do make sense in situations where the context requires constant modification of strategy.

Managing Achievement

While *achieving* is almost completely “wholesaled” to partners, USAID is still responsible. “Achieving” has become surprisingly synonymous with “contracting.” An excellent example was the E&E Bureau’s Partners Conference, a commendable attempt to develop the relationship with partners on Managing for Results. According to the proceedings, it was all about *managing* and very little about *results*. The ‘speakers’ were all from the contracts office. There were many questions from the partners about “What does Results mean?” They were asking for a better understanding of what they were being held accountable for in their contracts and grants—these were really planning and strategy questions. The design of the agenda was more to answer contracting questions. Is this relationship about contracting or achieving results? The main report discusses the difficulties in these relationships as does the section on partners below.

It is not just in our relationship with partners that this plays out. Agency staff are no longer managers of the development process, they are *cognizant technical officers*. They are delegated certain authorities from the contracts officers and require that contracts officer’s approval for most actions. This has implications for the ability of the organization to achieve,

especially in an environment requiring a certain degree of flexibility, speed, and risk taking. Contract officers are neither trained in development nor rewarded for taking risks in pursuit of development objectives. With the contracts-manager role that the Agency staff play, the whole mentality has shifted too far toward bureaucraties: it brings a real *accountability* sensibility to the fore. This is the environment that has led to reports of evaluations carrying a very negative connotation.

As one team leader reported: 'It is a "judgment" process, the end result of which is to "punish."

This mentality is not new to USAID but so much emphasis on delegations of authority, conflict of interest, accountability, etc. have caused it to flourish.

Every organization investing money, whether a private bank or USAID, has a tension between those charged with "using" the money and those charged with making sure "proper procedures and internal controls" are followed. Finding the balance between these two necessary functions is a critical issue for senior management. It may be that the pendulum has swung too far in favor of those whose job it is to protect the public purse and away from those whose job it is to *achieve* the public purpose.

Learning

The third function is *Learning*. For many years, USAID has struggled to develop some kind of institutional memory that would serve the Agency in doing a better job of planning and achieving in a rapidly changing world. In 1983, CDIE was created joining the evaluation and library functions in order to turn the Agency's data on its experience into a usable institutional memory. The supply side approach to solving the problem was to create a unit capable of responding to requests for the Agency's experience with a type of program by combining the library to create a response tailored to the specific request.

Doing this evaluation, we learned that it is simply not enough to create an institutional memory when there are no incentives operating in favor of using it. This brings us back to the person who reported that 90 percent of his time was spent on required activities and he had a hard time getting evaluation to the top of the remaining 10 percent!

There has to be a demand side. In the principles laid out in the ADS, strategic plan documents require the mission to "summarize prior experience in working with the country. . . ." as part of communicating the "Overall Assistance Environment." Buried with the quality of the development partnership, the likely impact of the strategy and a brief review of cross-cutting issues, it is one of seven requirements. If USAID more explicitly required every strategy to have a section reviewing experience on what works, what doesn't and why (including the mission's previous strategic experience), that would be the "demand" for people to use evaluation as a

learning tool, and start demanding better ones. Evaluation needs to again become a key part of the learning function.

3. Actors and Roles in USAID’s Evaluation System

Our scope of work directed us to look at USAID’s evaluation system. “USAID’s system” is closely tied to

- Partners: PVOs/NGOs, contractors, and universities;
- Other parts of the U.S. Government, primarily the State Department and the U.S. Congress.

Partners (all kinds taken together) implement over 95 percent of USAID’s programs. That puts them *in* the evaluation system and they will be discussed as part of it below.

Congress has long played an important role in foreign assistance—most noticeably by the “earmarking” (formal and informal) of funds. In 1993, Congress passed the Government Performance and Results Act which required that an annual report on agency results be submitted to Congress (all agencies of the U.S. government). Also in the 1990s, the Agency was part of a Government-wide Reengineering effort initiated by Vice President Gore. Together, these two initiatives focused the Agency’s attention on *results monitoring*.

The State Department’s role at the end of the Cold War has increased their interest in development as a foreign policy tool. Many senior Agency managers reported their diminishing control over decisions.

We don’t need evaluations because we don’t make the decisions anymore. They’re made at State.

I don’t make program choices the way I used to. Now, the Ambassador does that.
—A mission director

We are not aware if State Department has an “evaluation” function to help inform these program decisions. Their systems get them excellent information, but not specifically on the impact and lessons from past experience with activities supported with U.S. foreign assistance. USAID needs a strategy for using information to influence decision makers in State.

a. CDIE is the Center for Development Information and Evaluation. Their mission is to foster the use of development information in support of USAID’s assistance efforts.

CDIE has three functions and three corresponding divisions:

- Manage what is now called the learning system, including reporting to Congress;
- Carry out lessons-learned evaluations (a.k.a. best practices);
- Catalogue, store, and disseminate ‘development information’ produced by USAID.

To understand USAID's recent evaluation experience, we primarily needed to look most specifically at policy and guidance. As the experts, CDIE makes policy working with other PPC offices in the development of the ADS. Other supplemental guidance issued by CDIE included the TIPS (not an acronym): an ongoing series of best evaluation practices. TIPS number 11, titled the "Role of Evaluation in USAID," was intended to supplement the ADS policies although it was not issued until 1997.

As managers of the learning system, CDIE has the responsibility to provide technical assistance, train for, promote and insure high quality evaluation work throughout USAID. It is questionable whether they have been given the authority and resources to accomplish this.

The E in CDIE's name is for *evaluation*, but under reengineering they recognized the need for support in monitoring. In so doing, even they lost some of the emphasis on evaluation. Moreover, because of apparent methodological concerns, they shifted to calling their thematic studies "assessments" rather than evaluations. This contributed to the Agency's de-emphasis on evaluation.

CDIE's products and services will be discussed in a separate section below.

b. Regional Bureaus

Over the years, regional bureaus have played greater and lesser roles in the learning process. Some Bureaus have had very vibrant evaluation offices with staff and resources, but they seem to come and go. There has long been a debate as to whether this function should reside in regional bureaus and/or in PPC.

The Asia Bureau of the late 1970s and early 1980s was an early example of a particularly vibrant evaluation division, with three evaluation experts and a budget for doing evaluations. The Africa Bureau under the separate legislation of the Development Fund for Africa (DFA) built an evaluation function designed to meet the specific needs of the DFA. At times, bureaus have developed separate guidance, usually termed "supplemental" to avoid the appearance of contradiction.

E&E has two unique characteristics: a menu of objectives from which missions pick, and centrally managed projects. It was reported in questionnaires and interviews that, to missions, this meant that the evaluation responsibility was that of the regional bureau offices in Washington. In fieldwork, missions talked about a shift presently occurring of responsibility and authority to the field, evaluation included.

Currently, Africa and ANE have direct-hire evaluation officers, E&E has a PSC evaluation officer (due to their special authority to hire PSCs in Washington) and LAC has assigned evaluation as one role of the program officer. CDIE is calling this group together monthly to consult with them and keep them informed about changes in Agency policy and practices.

c. The Bureau for Humanitarian Response

BHR no longer has an evaluation officer. They gave up an evaluation position under reengineering with the understanding that that function was being centralized in PPC. In an interview, one senior staff person complained that CDIE should have provided them services in return but that had never happened. Their previous evaluation officer had drafted a supplemental guidance piece but it was never issued due to what they saw as the centralization of the function.

Two of BHR's offices have evaluation requirements written into their Request for Assistance guidelines. Both the Office of Private and Voluntary Cooperation (PVC) and the Office of Food for Peace (FFP) still require midterm and final evaluations for every activity. FFP explained that although they recognize that this is counter to Agency policy, they had just created the Food Aid Policy in 1995 when the Agency dropped the universal coverage requirement. FFP felt that they still needed universal coverage to keep tabs on how the shift towards new food aid policy objectives was working.

PVC seems to have retained the universal coverage requirement more because of a commitment to ongoing learning. In some instances, they provided grants to PVOs specifically to build up their evaluation functions. We found evidence of the positive effects of that strategy, but PVC now needs to think rather differently about the types of evaluations being done to best achieve learning.

d. Global Bureau

The Global Bureau is made up of the technical specialists for each of the Agency's objectives. One type of product is reports on best practices. We reviewed several of Global's best practices pieces and found that they usually do include evaluation of Agency-supported development assistance as a source of information. Functionally, they are sectoral or program evaluations. Although they might be strengthened by a more explicit consideration of best evaluation methods, they are generally excellent and widely used. They came up often in interviews in the field – much more often than CDIE studies. In addition to their written reports, the Macedonia democracy and governance (DG) staff said that the annual DG training would be an excellent source of up to date information on best practices—unfortunately, neither of them had been able to attend.

Global is where mission staff seem to first look for lessons and best practices. Their resources far exceed CDIE's. Ironically, in an example of one of CDIE's best Best Practices reports where there was extensive cooperation between CDIE and Global, Global subsequently hired the CDIE staff that had worked on it!

e. Missions

Missions vary widely in their attention to evaluation and learning. Many of the reasons are discussed generally in the main report and in our anecdotal observations on planning and learning above. There are specifics that play a role as well.

In several of the E&E missions visited during fieldwork, the issues of newness of the program and the reliance on PSCs added to or worsened the constraints:

- One new fast growing E&E mission was having difficulty covering basic administrative needs much less worrying about retrospective analysis.
- In a mission with only two years of development programs, it really is too soon to do evaluation.
- A PSC from a mission that was largely reliant on PSCs said how helpful it would be to have a list of who the good evaluators are. PSCs tend not to have been around and know the resources. Likewise, they were only vaguely aware of CDIE and the resources they could offer.

These patterns were seen in other Bureaus to lesser degrees as well.

The reduction in personnel has been accompanied by a change in the formal responsibilities of remaining direct hire staff. Most are now “strategic objective” managers, rather than “project managers.” In a context where consultants are routinely called in to write a mission’s own strategy, program officers and team leaders clearly are not in a position to worry about documenting lessons for the Agency as a whole.

This shortage of USDH staff time has also meant that the monitoring and evaluation function is the responsibility of local national or PSC staff. Some missions just include evaluation as one of the many functions of the program officer. Having FSN or PSC evaluation officers is not necessarily a negative, as many of these staff are very well trained and they are more likely to have the time to give it attention. However, if leadership does not actively support evaluative work, FSN staff reported that it was very difficult for them to become advocates within the mission. Most of their time and training in recent years, particularly in Africa, has been in implementing the indicator system. All felt the need to do more evaluation, and wanted more training and support. One mission director acknowledged that he had to be the mission monitoring and evaluation officer if there was going to be active learning.

The experience in Moscow highlights how much it helps to have someone who is a champion for evaluation, and the Moscow evaluation officer is an FSN. A professional evaluator who makes it his business to be aware of Agency evaluation policy, keep abreast of TIPS and other supplemental guidance, as well as what is going on in the evaluation field (i.e. by attending American Evaluation Association meetings) is reflected in the professionalism of Moscow’s

products and processes. Even in this case, the full support of the program office has been required to get this established and make it work.

The problem of running an agency with a patchwork quilt of staff resources is an entire subject for analysis and review, but the extraordinary reliance on non-career staff is the biggest structural change in the Agency. It has an impact on Planning, Achieving and Learning. It may be a contributing factor in the growth of Contract officer power, the weakness of the evaluation work, and the tendency toward micro-management at a time when reinventing government is trying to work the other way.

It is probably not a coincidence that Moscow was one of the posts that had a PD&S-type evaluation budget. PD&S funds were budgets set aside for design and evaluation work managed by the Program office. One of the policies in the 1995 ADS was that all funds would be allocated to the missions strategic objectives, functionally doing away with PD&S. The problems this causes are discussed in the main report. The 2000 revision to the ADS has created the Program Design and Learning Fund as an updated solution.

Another unintended consequence of the 1995 ADS guidance came from requiring partners to send in all “final reports,” including final evaluation reports. Evaluators generally consider missions to be their clients. They submit their reports to missions. There have been several instances of missions submitting reports to the DEC that were then called back. None of the instances reported to us were actually evaluation reports but uncomfortable nonetheless. Most partners have developed the practice of submitting “final drafts” to missions. That way, they didn’t have to deal with the requirement and risk the wrath of the mission for a mistake.

Missions need to have the responsibility to know when a report is final and appropriate to be shared widely. Any concerns about missions burying negative reports will not be ameliorated by expecting partners to send in reports. This is an untenable position for them.

USAID is striving to be a learning organization. USAID participation in evaluation can range from active initiation, planning, budgeting and monitoring the work of an outside team, to active participation by one or more USAID officers on the evaluation team. People learn more when they are actively engaged than when they passively read a report. USAID officers participating on teams has become extremely rare as staff resources have shrunk. Office directors are very reluctant to release personnel for the 3 to 5 weeks it takes to do the average USAID evaluation. Occasionally, they do field work for an evaluation in their own mission, but there has not been cross bureau participation in evaluations since the Impact Evaluations of the early 1980s.

Evaluations are a special part of the Agency’s current learning process. Agency managers are far more likely to rely on more convenient, less costly and more informal means for monitoring their programs. These means include field visits, talks with key informants, discussions with other donors, all filtered by their own experience and sense of judgment. The formal monitoring

system of results indicators does not seem to be that useful a tool for SO managers, as several reported to us that it is a tool for Washington not the missions.

Because of staff and budget constraints, as well as some reluctance to expose one's program to outside scrutiny, USAID managers are not initiating many evaluations. As mentioned above, they do require their implementing partners to do them, and they do participate in SOW development, team selection, and in review of findings. A more popular tool is the "assessment." Assessments are valued to the extent they tell USAID managers about the trends and dynamics in a sector's political, social and economic environment. Sometimes these assessments are "evaluations" in disguise, avoiding the negative stigma of an evaluation, and imposing no requirement that the findings and conclusions be widely shared.

The old guidance relies exclusively on the judgment of the operating unit to conduct an evaluation or similar study when it determines it has a problem or need. We need to recognize that missions have, to varying degrees, responded and taken up the re-engineered approach to overall learning. They are too tight on resources to fully document the findings and lessons useful to the development community as a whole.

f. Partners

Partners play two types of roles in USAID's evaluation system,

- As implementers of over 95 percent of our programs
- As external evaluators

Sometimes the same individual will even play both roles at the same time.

As implementers

Partners are the on the ground implementers of our programs. They do what gets evaluated. "Our" success depends on them. They do not have the same long-term incentives to build the kind of transparent, objective and widely-shared knowledge base necessary for USAID as a learning organization. I don't mean to say they don't have the incentives to build a knowledge base; the experience was clearly that they do. But their incentives and their base are not the same as USAID's. The example cited in the main report of the organization reporting evaluation lessons in their annual report is a case in point. They circulated the lessons according to their incentives within their organization. We need to rely on our partners but currently we are relying on them too much.

Worse yet, USAID's contracting and conflict of interest regulations put a firewall in the middle of what needs to be a free flowing of information, knowledge and learning throughout the system. Legally, this is probably unavoidable, but we need to get a lot smarter and more active about how we work around it. The issue of how to get the information to where it's needed in a timely way is much more serious than how many evaluations we are doing.

The regulators of the bureaucracy dictate that ADS guidance be directed to USDH staff. Again this reinforces the separation between Agency planners and partner implementers. An evaluation handbook could be written for the whole audience, not just half of it. Similarly, training to share common language and expectations is as necessary for partners as for USAID staff.

As ‘external evaluators’

Most evaluations reviewed did not involve any USAID staff, some involved Partner staff, but the majority were done by outside hired consultants. Whether this assured objectivity is difficult to ascertain. A surprising finding from the Africa fieldwork was that local firms using African experts was more the rule than the exception. This is not a negative finding, but it may reflect relatively minimal budgets for evaluation, especially when the evaluations are the usual midterm/final variety as practiced by several PVOs. In the long term, this will be a strength. This is part of the audience that needs support from the Agency on evaluation methods and practices.

4. CDIE’s products and services

a. Policy/Guidance

In response to the questions about the guidance in the questionnaire, only 19 percent reported that the evaluation policies in the ADS were clear, highly useful and well known. Another 20 percent reported that they were not very useful. Indeed several of the staff commented that they understood them themselves but did not think that other mission staff did. In this instance, we questioned the positive as well as the negative and only 40 percent of respondents chose to address the issue.

It should also be noted that 12 mission staff (13 percent) were willing to admit that they didn’t know the requirement to evaluate every activity had been dropped. This was pretty even across regions and sizes of mission. This is a marketing failure. It is not sufficient to just write the guidance; you need to make an effort to get people to know it and use it.

Many people began conversations saying that they weren’t sure that they knew the difference between monitoring and evaluation. At the activity level, they did. The one real confusion we found was about “monitoring” beyond the mission. Many mission staff said that the R4 is not a monitoring document; it is a report for Washington. Some characterized it as a PR report or a sales document. Several said it couldn’t be monitoring because it is only once a year—not understanding that there is no set time frame for monitoring. Most did not understand that the R4 is the monitoring that Washington does of the field. Similarly, they didn’t see the Annual Performance Report as the Congress’ monitoring of the Agency. Lack of understanding contributed to these efforts getting scant attention because there was no recognizable value to the field staff.

The emphasis in ADS 200/2000 on Performance Monitoring Plans (PMP) is on a document—one the mission staff will think they need to hire an outside expert to come in and do for them, rather than on what they need to learn, know and understand. The Project Logical framework was deep-sixed by the restrictiveness of the boxes. Why would we expect the PMPs to overcome the same hurdle? A PMP is a project logframe without the cause-effect hierarchy of the objective tree. We've kept the boxes and lost the analytical framework. Teaching by paper is the issue; no guidance is sufficient without people to people contact.

It is worth repeating here that the Agency policy on submitting documents to the central repository is found in ADS section 203.3.9. CDIE has a contract for the Development Evaluation Clearinghouse to collect, catalogue and disseminate all documents produced with the support of the Agency. The guidance for submitting documents electronically is in section 203.3.9 which can be found on the Agency website (www.usaid.gov for partners). The policy is in place, the DEC is in place, but the sharing can only take place if Agency staff can find the time to submit their documents.

b. Tips

Tips function as guidance in support of the policy in the ADS. Forty five percent of questionnaire respondents said they found the *Tips* useful in managing their evaluation work. Comparing this to the 19 percent that said they found the policy useful suggests that mission staff find it more useful to have advice than requirements. One partner-evaluator claimed that an “out of date *Tips* was the only guidance I was given” when doing an evaluation. Updating and pulling them together as part of a handbook would be very useful.

c. Technical assistance

The technical assistance offered by CDIE in the second half of the 1990s was very monitoring focused including data collection and strategic planning. This was a sign of the times and a classic example of 1990s wholesaling: do a big contract and bring in consultants. The lapse in services available to the field during the exceedingly awkward transition from one contractor to another demonstrated one problem with this approach.

There was also a failure of leadership. CDIE did not adequately keep their focus on the balance between monitoring and evaluation and allowed the balance to stay too long with building up the monitoring systems. It was reported to us that the contractor proposed evaluation efforts that CDIE did not take. It is similar to the field situation where the partners are the day-to-day managers and understand the needs for evaluation. The methodological problems with the evaluations cited in the main report will require leadership from CDIE. To successfully provide this leadership, CDIE needs more expertise in evaluation methods.

d. Evaluations/assessments/studies

CDIE makes a grand effort to codify and synthesize lessons learned from its own crosscutting studies, but the existence of these often excellent studies is not a measure of their actual use in the planning and management process.

Global's best practices have an edge in the sectors where Global has more resources. CDIE stopped doing "evaluations" and started doing assessments. Notes from a staff retreat in the late 1980s seem to indicate that they were concerned with providing timely lessons to missions. In trying to model an understanding of sound evaluation methods, they unintentionally sent the message that good evaluation practices were out of reach! Maybe they held the bar too high.

Decentralizing the system (i.e. through the 95 ADS guidance) also exacerbated the already difficult task of providing an overall view of what information the system as a whole needs. It is very difficult for CDIE with shrinking budget and resources to get the whole picture from so many little pieces. Indeed, the annual report is not much based on mission evaluations or R4 monitoring. We need to look at the investments in these knowing that they don't serve an overall agency purpose.

e. Evaluation Agenda

The "Evaluation Agenda" is in no way an Agency agenda. It is very much a CDIE agenda and several senior staff in Washington were unhappy that they couldn't draw on CDIE resources. Sending a paper for clearance after a quick meeting with Bureau evaluation officers is not sufficient.

When this team interviewed the senior staff of one regional bureau, they were so pleased they actually contacted CDIE and suggested that this type of contact take place more often. Use this report to help get evaluation's visibility and priority raised, and actively involve Bureaus.

f. Summer Seminars

Summer seminars filled the training need to some extent but their primary emphasis was not for training in evaluation planning and methods but as a vehicle to communicate the findings and lessons learned of CDIE's studies. The tactic to have the seminars during the summer when staff come through Washington wasn't bad, but the team did not meet anyone who had attended the summer seminars. The seminars need the resources to go on the road.

5. Conclusions about the Management of the USAID Evaluation

1. USAID, during the study period, has not had an organized evaluation system with clearly understood rules, expectations and responsibilities.
2. The ADS begins the process of clarifying the rules.

3. With the emphasis on building a performance monitoring system, expectations have been too low. Senior management has demonstrated little interest in evaluation.
4. CDIE has had overall responsibility but other than their own evaluations, even they found it hard to give the evaluation system of the Agency much priority.
5. Three regional bureaus have retained evaluation officer positions but one is part time and another is a PSC.
6. Global plays a large role and capturing lessons and best practices but with variability in how systematic these are and how much the focus on the results of USAID's own assistance.
7. BHR retained the requirement to evaluate every activity when that requirement was eliminated for the rest of the Agency, but then downgraded the bureau evaluation function. This created considerable confusion in the partner community, particularly the PVOs.
8. There were too many examples of evaluations being linked with success stories. Continuing this practice will call into question the objectivity and independence of evaluation findings and conclusions.

6. Recommendations to CDIE

1. The Agency's Evaluation Agenda needs to capture all the retrospective analyses planned, not just those done in CDIE. Global, regional bureaus and BHR all do "evaluations" and these should be considered as a whole, at least as CDIE decides how to allocate resources. This would require wider, and more frequent, consultation with others in the regional bureaus than just the evaluation officers.
2. CDIE could undertake analyses of the relationship between certain management structures, procedures and incentives and the way programs get designed and done, particularly contracting. The Management Bureau would be a stakeholder.
3. CDIE needs to identify partners as important clients since they carry out so much of the Agency's evaluation work. They need to have the resources to take on this new challenge.
4. CDIE must play a role in getting expert evaluators involved in strategy and program design. Would we be surprised if most of USAID's overstretched staff read section 201 during planning and didn't even read section 203 until the project was almost over?
5. CDIE needs to find a way to provide evaluation expertise in the development of mission evaluation design, identification of evaluations and development of scopes of work.
6. CDIE staff must direct resources toward training and communication.
7. CDIE should use RFNet for quick response tutorials on evaluation planning and methods.

8. CDIE needs to be sure they have adequate expertise on evaluation methods on their staff. Hiring from the outside to identify someone adequately versed in social science methods should be considered.

9. CDIE should consider a two-stage approach to developing a new Evaluation Handbook. Initially, it needs to support the Agency policy in the ADS with evaluation wisdom such as: what to write into a grant/contract/ cooperative agreement on evaluation requirements; how to go beyond the fly-in approach to methods; mandate missions to take the responsibility to send in evaluation reports. It should at least append any supplemental Bureau guidance on evaluation. The six TIPS on evaluation provide an excellent basis. Models of good evaluation reports and scopes of work should be included.

Subsequently it needs to address such issues as guidance on assessments, particularly sector assessments, and best practices in the balance between monitoring and evaluation. Guidance on close out reports should be included as well. Guidance on strategic evaluations and how they are different from activity evaluations should be developed. Case studies of evaluations, from the initial identification of an issue to how the completed evaluation affected decisions should be developed.

Appendix 3

Questionnaires

1. Mission Questionnaire
2. Partner Questionnaire
3. Tabulated Results of Mission Questionnaire

1. MISSION QUESTIONNAIRE
Study on Recent Evaluation Experience

May 2000

Sponsored by PPC/CDIE

MISSION QUESTIONNAIRE

For the purposes of this study, *evaluation* is characterized as any empirically-based analysis of problems, progress, achievement of objectives or goals, and/or unintended consequences for mission's

- country strategy,
- sectoral programs and/or
- activities

whether the evaluation is conducted by

- Mission staff,
- partners, or
- outside consultants.

Evaluation refers to analysis of USAID-supported programs and not just the more general types of studies such as sector assessments. We also do not include contractor/grantee performance assessments that do not also include an analysis of the mission's programs/activities.

For multiple choice questions, click on as many lettered boxes as needed to describe your circumstances. Multiple answers are O.K. There is space for comments at the bottom of the questionnaire. Comments are very welcome.

A note on confidentiality: replies will go directly to consultant Cynthia Clapp-Wincek and you will not be identified by your email address. Individual responses will not be shared with anyone other than the study's senior advisor, Richard Blue. If you have concerns about what is in writing, Question 14 offers the opportunity for us to contact you directly.

Question 1:

Name: _____.(optional)

Responsibility in mission: _____.

Years in this mission: _____.

Regional bureau of your mission: AFR LAC ANE E&E

Size of mission (# of USDH): _____.

Question 2:

I have personally:

- a. Managed an evaluation or similar analytic study for this mission.
- b. Conducted an evaluation or similar analytic work while in this mission.
- c. Asked for or directed an analytic study to be done.
- d. Had no direct experience with managing, doing, or using mission evaluation reports during my time here.
- e. Participated in an evaluation conference (examples: E&E Energy and Infrastructure Division's annual conferences or G/PHN's Virtual Workshop on Collaboration).

Question 3: Please list the evaluations you are aware have been produced, sponsored, or requested for your mission during CY 1999. Complete titles preferred if readily available; otherwise just characterize the reports. Also please list evaluation conferences if you checked 2e.

- a. _____.
- b. _____.
- etc. _____.

Question 4: For each product, please identify the person/event/issue which caused your office to undertake the evaluation.

- a. _____.
- b. _____.

Question 5: For each product, who has been the primary user of the product?

- a. _____.
- b. _____.

Question 6:

Please check the budget sources for evaluations listed in question 3:

- a. mission strategic objective,
- b. mission program development and support funding,
- c. central bureau funding, such as Global Bureau,
- d. regional bureau funding, such as Africa Bureau Sustainable Development,
- e. other (please specify) _____.

Question 7:

Which of the following statements best describes your mission's situation reevaluation (multiple answers are always O.K.)?

- a. The results monitoring system has largely replaced the need for evaluations.
- b. The results monitoring system is an inadequate system for determining what programs and activities are working.
- c. The results monitoring system is an inadequate system for determining why programs or activities are working or why not.
- d. We carry out evaluations as needed and they provide useful input to our thinking.
- e. We would like to do more evaluation work but are constrained by lack of budget, time on the part of team leaders, and/or the difficulties of contracting for them.
- f. Mission leadership strongly supports doing empirically based analytical investigations of the results and impact of our development programs.

Question 8:

Which best describes the mission's use of evaluation for management or decision making?

- a. We have carried out evaluations but not found them useful in management or other decision making.
- b. Evaluations have been critical to our understanding of activity performance.
- c. Evaluations are useful to the activity evaluated but have no usefulness to any other mission work.
- d. I am not aware of any evaluations the mission has done and therefore cannot reply to this question.

Question 9:

Which statement best describes your mission's use of evaluations for preparing the mission's R4?

- a. Mission evaluations help inform our strategic thinking, but are not formally cited in the R4.
- b. Mission evaluation findings are regularly cited in our R4 narrative.
- c. Mission evaluations have been used to support R4 annexes, but have played little role in the preparation of the narrative.
- d. Mission evaluations play little role in our R4 reporting or other reporting.
- e. There have been no evaluations to inform our R4 reporting since I have been at this mission.

Question 10:

In your experience with this mission, how would you describe the process by which the mission develops its strategy and implementing activities?

- a. The mission systematically uses the results monitoring process to develop its strategy and program activities.
- b. The mission regularly conducts sectoral assessments and other strategic analyses which are used to formulate strategy and make decisions about program activities.

- c. This mission has conducted an evaluation of a strategic objective as part of the decision making of extending, changing or ending that strategic objective.
- d. Our strategy and activities are more a function of Washington mandates and “directed” budget allocations from Congress than they are of systematic analysis.
- e. Strategy and program activity decisions tend to be opportunistic and personality driven within the larger framework of budget constraints.
- f. Our strategic plan is a rationalization of our existing portfolio of programs and projects.

Question 11:

What best describes your situation since the requirement to evaluate every project was dropped under the new Automated Directive System?

- a. I was not aware that the evaluation requirement was dropped.
- b. It has saved valuable time and eased our burden.
- c. It made it more difficult to do evaluations.
- d. It has been more difficult to justify budgets to do evaluations.
- e. There has been no change in our practices.

Question 12:

Which best describes your experience with Agency evaluation guidance?

- a. USAID evaluation policies (ADS 200 series) are clearly stated, highly useful and well known to me and this mission.
- b. USAID evaluation policies are unclear and not very useful.
- c. I have found the performance Monitoring and Evaluation TIPS from CDIE useful in managing our evaluation work.

Question 13:

Based on your experience, should USAID be:

- a. Doing more evaluations
- b. Fewer evaluations
- c. About the right level now

14. I have more to say, please contact me:

Yes No

If yes, please provide details on how you want to be reached (e.g. phone number, email address, etc.)

Comments: _____.

2. PARTNER QUESTIONNAIRE
Study on Recent USAID Evaluation Experience

May 2000

Sponsored by USAID/PPC/CDIE

Please check the choices and fill in the blanks to share your experience. For multiple choice questions, multiple answers are OK. If you don't find the right possibility, add it to the comments.

What do we mean by 'evaluation'?

For the purposes of this study, evaluation is characterized as any empirically-based analysis of problems, progress, achievement of objectives or goals, and/or unintended consequences for mission's:

- country strategy
- sectoral programs and/or
- activities

whether the evaluation is conducted by

- mission staff
- partners, or
- outside consultants.

Evaluation refers to analysis of USAID-supported programs and not just the more general types of studies such as sector assessments or analyses of the political, economic and social environments which help to define a mission's strategic focus and context. We also do not include contractor/grantee performance assessments that do not also include an analysis of the mission's programs/activities.

Question 1:

Name: _____.(optional)

Organization: _____.(optional)

If organization not filled in:

- Private voluntary organization
- Contractor
- University

Responsibility in organization:

- Management
- Operations
- Evaluation/research
- Other: _____.

Years in this organization: _____.

Region of your work:

- Global
- Africa
- Asia & Near East
- Europe and Eurasia
- Latin America and the Caribbean

Question 2:

I have personally:

- a. Participated in an evaluation of activities or programs supported by USAID.
- b. The USAID-supported activities that I work on have been evaluated.
- c. Managed an evaluation or similar analytic study for my organization.
- d. Asked for or directed an evaluation or similar analytic study to be done.
- e. Other: _____.

If you checked A or B in Question 2, please answer questions 3-9. Otherwise, go on to Question 11.

Question 3: How many evaluations have you done with/for USAID?

Question 4:

What part(s) of USAID sponsored the evaluation?

- a. Mission
- b. Global
- c. Bureau for Humanitarian Response
- d. Other: _____.

Question 5:

What part of USAID funded your participation?

- a. The activity in the mission/our grant or contract.
- b. Other mission funds
- c. Global
- d. Bureau for Humanitarian Response
- e. Other: _____.

Question 6: For evaluations in which you have been personally involved, please indicate the person, event or issue that prompted the evaluation.

Question 7:

What kind of guidance did your scope(s) of work or contract/grant give you about the evaluation methods to use?

- Written guidance
- Verbal guidance
- Example from other evaluations
- Other: _____.
- Comment: _____.

Question 8:

On what background and experience did you rely to determine the methods you used to carry out the evaluation(s):

- a. I rely on my academic training.
- b. I use training from my current organization on how to do evaluations.
- c. I use training from another organization where I carried out evaluations.
- d. I use training provided by USAID.
- e. I rely on my experience from participating in previous evaluations.
- f. Other/comment: _____.

Question 9:

Which best describes your experience with USAID evaluation guidance?

- a. I have no direct experience with USAID evaluation guidance.
- b. USAID evaluation policies (Automated Directive System 200 series) are useful in conducting evaluations.
- c. I have found the Performance Monitoring and Evaluation TIPS from CDIE useful in managing evaluation work.
- d. To carry out evaluations, I rely most heavily on _____ (e.g. our organization's own evaluation guidance).
- Comment: _____.

Question 10: Describe the relationship with USAID during the evaluation(s); e.g., did you have adequate independence? Were they supportive? Were they defensive?

If you checked B in Question 2, please answer questions 11, 12 & 13. Otherwise, go on.

Question 11:

What best describes your use of evaluation for management or decision making?

- a. We have carried out evaluations as required but not found them useful in our management or other decision making.
- b. Evaluations have been critical to our understanding of activity performance.
- c. Evaluations are useful specifically to the activity evaluated but have no usefulness to any other work we do.

Comment: _____.

Question 12:

What best describes your situation since USAID’s requirement to evaluate every activity was dropped under the Automated Directive System initiated in 1995?

(NB: performance monitoring continues to be required of every activity.)

- a. I was not aware that the evaluation requirement was dropped.
- b. Every contract/grant/cooperative agreement we do with USAID does require evaluation.
- c. It has saved valuable time and eased our burden.
- d. It made it more difficult to do evaluations.

Comment: _____.

Questions for anyone who has worked with USAID:

Question 13:

Which of the following statements best describes your view of USAID’s approach to evaluation?

- a. The results monitoring system has largely replaced the need for evaluations.
- b. Evaluations are carried out as needed and they provide useful input to USAID’s thinking.
- c. USAID staff are too constrained by lack of budget, time on the part of team leaders, and/or the contracting difficulties to carry out needed evaluations.
- d. Mission leadership strongly supports doing empirically based analytical investigations of the results and impact of our development programs.
- e. Other/comment: _____.

Question 14:

In your experience working with USAID, how would you describe the process by which missions develop their strategy and implementing activities?

- a. I have no idea.
- b. The missions I’ve worked with systematically use the results monitoring process to develop their strategies and program activities.
- c. Missions I’ve worked with regularly conduct sectoral assessments, strategic analyses and/or strategic objective evaluations which are used to formulate strategy and make decisions about program activities.
- d. Mission strategies and activities are more a function of Washington mandates and “directed” budget allocations from Congress than they are of systematic analysis.
- e. Mission “strategies” are a rationalization of existing portfolios of programs and projects.
- f. Other/comment (e.g., partners’ role in strategy development): _____.

Question 15:

Based on your experience, should USAID be:

- a. Doing more evaluations
- b. Fewer evaluations
- c. About the right level now
- d. Including partners more
- e. Putting less burden on partners?

Question 16:

I have more to say; please contact me:

Yes No

If yes, please provide details on how you want to be reached _____ (e.g. phone number, email address, etc.).

Should you prefer to contact us directly, our email addresses are ccwincek@aol.com and rbluemont@aol.com.

Comments: _____.

MANY THANKS

3. Tabulated Results of Mission Questionnaires

August 2000

N = 93

Question 1: Name – optional

Responsibility in Missions

Director/Deputy	15
Program/Project Officers	32
Evaluation Officers	7
Team Members	36

Regional Bureau of Your Mission

Africa	28
Asia/Near East	18
Latin America/Caribbean	27
Europe/Eurasia	16

Size of Mission (# U.S. Direct Hires)

9-15	13
9-16	20
9-17	26
> 15	23
Egypt	7

Question 3: Please list the evaluations you are aware have been produced, sponsored or requested for your mission during CY 1999.

Question 4: For each product, please identify the person/event/issue which caused your office to undertake the evaluation.

Because it was planned (Includes “midterm” “EOPS”)	35
Impact	7
Lessons Learned (Not tied directly to new designs)	10
Impact and Lessons Learned	7
New Program Design/ Realign new strategy	64
Specific Questions/Problems (7 of which identified in Results Reviews)	39

Assess Grantee/Contractor Performance	6
Feed into annual performance reviews (e.g. “needed for R4”)	5
Host Government Request	4
Take stock	8
M/OP Request	1

Question 5: For each product, who has been the primary user of the product.

Not tabulated because the usual response was: “the mission.”

Question 6: Please check the budget sources for evaluations listed in question 3.

Mission Strategic Objective	67
Mission PD&S	18
Central Funding	15
Regional bureau (e.g. AFR/SD)	4

Other:

Office of Transition Initiatives; mission-funded OE; Title II; Project funds; CDIE

Question 7: Which of the following statements best describes your mission’s situation regarding evaluation (multiple answers are always OK)?

The results monitoring system has largely replaced the need for evaluations.

26

The results monitoring system is an inadequate system for determining what programs and activities are working.

18

The results monitoring system is an inadequate system for determining why programs or activities are working or why not.

44

We carry out evaluations as needed and they provide useful input to our thinking.

74

We would like to do more evaluation but are constrained by lack of budget, time on the part of team leaders, and/or the difficulties of contracting for them.

25

Mission leadership strongly supports doing empirically based analytical investigations of the results and impact of our development programs.

41

Question 8: Which best describes the mission’s use of evaluation for management or decision making?

- a. We have carried out evaluations but not found them useful in management or other decision making. 3
- b. Evaluations have been critical to our understanding of activity performance. 60
- c. Evaluations are useful to the activity evaluated but have no usefulness to any other mission work. 15
- d. I am not aware of any evaluations the mission has done and therefore cannot reply to this question. 6

Question 9: Which statement best describes your mission's use of evaluations for preparing the mission's R4?

- a. Mission evaluations help inform our strategic thinking, but are not formally cited in the R4. 33
- b. Mission evaluation findings are regularly cited in our R4 narrative. 29
- c. Mission evaluations have been used to support R4 annexes, but have played little role in the preparation of the narrative. 12
- d. Mission evaluations play little role in our R4 reporting or other reporting. 14
- e. There have been no evaluations to inform our R4 reporting since I have been at this mission. 10

Question 10: In your experience in this mission, how would you describe the process by which the mission develops its strategy and implementing activities?

- a. The mission systematically uses the results monitoring process to develop its strategy and program activities. 36
- b. The mission regularly conducts sectoral assessments and other strategic analyses which are used to formulate strategy and make decisions about program activities. 50
- c. This mission has conducted an evaluation of a strategic objective as part of the decision making of extending, changing, or ending that strategic objective. 33
- d. Our strategy and activities are more a function of Washington mandates and "directed budget allocations from Congress than they are of systematic analysis. 18

- e. Strategy and program activity decisions tend to be opportunistic and personality driven within the larger framework of budget constraints. 16
- f. Our strategic plan is a rationalization of our existing portfolio of programs and projects. 19

Also grouped:

Systematic processes	119
(a + b + c)	
Responding to external pressures	53
(d + e + f)	

Question 11: What best describes your situation since the requirement to evaluate every project was dropped under the new Automated Directive System?

- a. I was not aware the requirement was dropped. 12
- b. It has saved valuable time and eased our burden. 42
- c. It made it more difficult to do evaluation. 14
- d. It has been more difficult to justify budgets to do evaluations. 11
- e. There has been no change in our practices. 28

Question 12: Which best describes your experience with Agency evaluation guidance?

- a. USAID evaluation policies (ADS 200 series) are clearly stated, highly useful and well known to me and this missions. 18
- b. USAID evaluation policies are unclear and not very useful. 19
- c. I have found the Performance Monitoring and Evaluation TIPS from CDIE useful in managing our evaluation work. 42

Question 13: Based on your experience, should USAID be:

- a. Doing more evaluations. 30
- b. Doing fewer evaluations. 5
- c. About the right level now. 47

Appendix 4

Fieldwork Interview Schedules

- Mission
- Partners

EVALUATION OF USAID'S RECENT EVALUATION EXPERIENCE
FIELD INTERVIEWS
MISSION STAFF

COUNTRY

Name/Office/Job title

What evaluations have you personally participated in? When? Your role?

What was the person/event/issue that caused your office to undertake the evaluation?

Who wrote the scope of work? Did it tell the evaluator what methods to use?

Do you use a model for scopes of work? Where did you learn to write scopes of work?

Did you feel that the methods used were sufficient to get you the quality of data you needed?
Why or why not?

How did you find the evaluation team? What types of expertise did it include?

Was the evaluation useful? Who was the primary user (or intended user) and what was the use?

What other evaluations are you aware of? Who should I contact to find them?

What assessments have you personally participated in?

Why was it called an assessment? How did it differ from an evaluation?

How was it used and by whom?

What did the evaluation/assessments that you were personally involved in cost?

How much does the mission spend each year on evaluation and assessments? [percent of program]

Are there other instances where you thought about doing an evaluation and decided not to?
Why?

Are evaluations used in your R4 process? [cited? In annexes?]

Do you/your mission/office routinely review your R4 data?

Are you aware of any instances where R4 monitoring has triggered an evaluation?

Are evaluations used in your office's/your mission's development of strategy and implementing activities?

What types of systematic analysis are used in strategy development and decisions about program activities? Examples

Or do Washington mandates and earmarks drive the program?

Were you aware that the requirement to evaluate every activity was dropped in 1995?

Yes _____ No _____

What has been the result? Saved time; eased burden; made it more difficult to do evaluations? Difficult to justify budgets? No difference.

Do you know where to find USAID's evaluation policy?

Yes _____ Where? _____ No _____

What do you think of ADS 200?

Have you every used CDIE's Monitoring and Evaluation TIPS?

Did you know that final reports are required to be sent to CDIE/R&RS?

Yes ____ No ____

Do you send them? Yes ____ No ____

If not, what is the constraint?

Are they never finalized?

Are they sensitive?

Do you think USAID should be doing more evaluations? Why?

One third of the questionnaire respondents said we should be doing more evaluations. Why, do you think?

EVALUATION OF USAID'S RECENT EVALUATION EXPERIENCE
FIELD INTERVIEWS
PARTNERS

COUNTRY

Name/Office/Job title

What does USAID require for monitoring and evaluation of your activity?

How does it differ from what you would do?

Do you do monitoring/evaluation other than that required by USAID?

How much did you spend in 1999 on monitoring? Was it a normal year?

How much did you spend in 1999 on evaluation? Was it a normal year?

What evaluations have you personally participated in? When? Your role?

Why was the evaluation undertaken?

How was the scope of work developed (including methods)? Did you have a role?

How was the evaluation used? What decisions were made resulting from the evaluation?

How was the evaluation team chosen? Did you have a role?

What role did USAID staff play?

How independent was the evaluation team?

Do you know of USAID's evaluation policies? Have you ever heard of or seen ADS 200?

Have you ever heard of or seen CDIE's Monitoring and Evaluation TIPS?

Appendix 5

Bibliography

Bibliography

- Adams-Matson, Michelle, Douglas Daniell, Anne Inserra, Janet Tuthill. 2, July 1999. "Accuracy, Sufficiency and Innovation of Reporting: An Analysis of 26 Africa Bureau Operating Units' R4s." Africa Bureau FY 2001 R4 Analysis – Task. Washington: USAID.
- Advisory Committee on Voluntary Foreign Aid. August, 1998. "ACVFA Quarterly Report." (including discussion from June 10, 1998 meeting on "Reporting Needs of External Audiences" and "Missions' Role in USAID's System to Report Results to External Audiences.")
- Danida. "Evaluation Guidelines" by the Ministry of Foreign Affairs, Hedeusene, Denmark. [<http://www.um.dk/udenrigspolitik/udviklingspolitik/evaluering>] February 1999.
- Evaluation Office, IDB. June 19-20, 1997. "Evaluation Cooperation Group of the Multilateral financial Institutions; Framework and Practices for the Assessment of Development Effectiveness." ECG Presentation before special Meeting of Officials of Multilateral Financial Institutions. Washington.
- Greene, Katrina. January 7, 1999. "Narrative Summary, Evaluation of FY 1997 Evaluations." R&RS. Washington: USAID.
- Hopstock, Paul J., Allan C. Kellum and Malcolm B. Young. May 1989. "Review of the Quality of A.I.D. Evaluations Fy 1987 and FY 1988." A.I.D. Evaluation Occasional Paper No. 19.
- Kerley, Janet. March 23, 2000. "A Report on the Use of Evaluations in the Europe and Eurasia Bureau/USAID; An Inventory of Evaluations." Office of Program Coordination and Strategy, Europe and Eurasia Bureau. Washington: USAID.
- Mansfield, David. "The United States Case Study; NGO Evaluation Policies and Practices," Appendix 8. In *Searching for Impact and Methods: NGO Evaluation Synthesis Study*. A Report Prepared for the OECD/DAC Expert Group on Evaluation. [<http://www.valt.helsinki.fi/ids/ngo/appcav.htm>]. May 1997.
- Office of Program Coordination. December 3, 1999. "Reporting on Results: Review of the Revised R4 Process." Bureau for Program and Policy Coordination. Washington: USAID.
- U.S. Congress. January 5, 1993. "Government Performance and Results Act of 1993," Chapter 3. In *title 5, United States Code*. One Hundred Third Congress of the United States of America. Washington.
- USAID. August 3, 1998. "Reaching-4-Results Workshop." Course Number 65J, H/HR/LS Syllabus. Washington.

USAID. September 2000. "USAID Programming Policies; ADS 200 Series Information Package." [<http://www.usaid.gov/pubs/ads/200>]. Washington.

USAID. N.d. "The History of CDIE." Memorandum.

Vondal, Patricia. May 1999. "Indicator Quality Assessment; An Analysis of Seven Cases." Prepared for AFR/DP. Washington: USAID.

Working Party on Aid Evaluation. December 17, 1999. "Results Based Management in the Development Co-operation Agencies; A Review of Experience." Organization for Economic Cooperation and Development DCD/DAC/EF (2000)3.