

“Meta-Evaluation of USAID Trade Capacity Building Evaluation SOWs and Reports”

This paper reprints sections that report on a meta evaluation of USAID trade capacity building evaluation Statements of Work and evaluation reports described in USAID’s 2010 evaluation *From Aid to Trade: Delivering Results: A Cross-Country Evaluation of USAID Trade Capacity Building* (http://pdf.usaid.gov/pdf_docs/PDACR202.pdf USAID), pp. 119-128 and 227-232.

3. TCB PROGRAM AND PROJECT EVALUATION

In addition to considering how best to monitor TCB investments and their outcomes, the joint OECD/WTO effort to oversee TCB assistance on a multi-donor basis has focused in some degree on the appropriate approach to evaluation of TCB activities. For a 2008 OECD Policy Dialogue on Aid for Trade, the German Marshall Fund produced a volume that suggested a number of types of evaluations that might be advisable to conduct, depending on needs of a specific TCB project or program. This background paper, entitled *Evaluating Aid for Trade: Why, How, and What*, also reinforced the idea that results chains, or Results Frameworks, are appropriate for TCB programs, since they make explicit the hypotheses that evaluations might be asked to examine.

The OECD subsequently undertook another study (not yet published on its website), to examine whether special evaluation methods needed to be developed to assess the impact of TCB programs and projects. The conclusion of this OECD report—that adequate evaluation methodologies, including rigorous impact evaluation methodologies, exist for evaluating TCB programs and projects—is consistent with the observations of this evaluation team. Unlike TCB performance indicators where issues exist, there is no indication that TCB as a field requires a unique set of evaluation methods. Against this backdrop, this evaluation examined USAID efforts through 2009 to evaluate its TCB investments.

USAID’s TCB evaluation experience is one component of USAID’s overall evaluation experience. Historically, USAID had a strong reputation for carrying out program and project evaluations, and until the early 1990s required midterm and final evaluations for all projects. At the same time, USAID introduced new performance monitoring requirements in the early 1990s; over the next decade the number of evaluations USAID carried out each year dropped precipitously, from 497 in 1994 to 73 in 2007.¹ USAID has recently (2009) reintroduced a mandatory evaluation requirement at the program level.

a. TCB Evaluations at the Program Level

Prior to this evaluation, USAID conducted one program level TCB evaluation. As explained in the previous section, this evaluation used performance monitoring information found in annual R4 reports from missions as its primary data source. In addition, as a precursor to this evaluation, USAID/EGAT invested in a paper that examined alternative approaches for conducting evaluations at the program level for USAID’s TCB program, entitled *Trade Capacity Building Evaluation Methodologies and Indicators*. That paper recommended the use of a cluster approach and served as a foundation for the cluster approach used in this evaluation.² It also suggested that for those clusters of TCB projects this evaluation examined, results chains of some type would be an appropriate way to identify intended results, the hypothesized links between them, and appropriate performance indicators on which to look for evidence of achievements.

As part of this evaluation, a review was undertaken of previous evaluations by USAID or other donors. Early in the study period, MSI reviewed the findings of the initial set of TCB project evaluations it located,

¹ Hageboeck, Molly, *Trends in International Development Evaluation Theory, Policy and Practices*, prepared for USAID. Washington D.C., Management Systems International, 2009.

² Nathan Associates Inc. “Trade Capacity Building Evaluation Methodologies and Indicators” (2007) at: http://pdf.usaid.gov/pdf_docs/PNADJ708.pdf

extracting key findings from each. The results of this early effort was somewhat disappointing as most of these studies yielding broad findings that apply equally well to most types of development assistance, e.g., programs are more effective when the enabling environment is conducive to the achievement of program objectives.³ Subsequent to this effort, the team located a volume prepared by OECD entitled, *Trade-Related Assistance, What Do Recent Evaluations Tell Us?* (2006), that reviewed many of the same evaluations as had the MSI team, reaching similar conclusions.

Aside from this early review, which proved valuable in developing the Results Framework used in this evaluation, MSI concentrated on finding project-level evaluations and using them along with other project reports as a source of evidence about project performance and the kinds of results yielded by USAID TCB investments.

b. TCB Project-Level Evaluations

While USAID encourages, but does not require, project-level evaluations be undertaken, it does provide detailed guidance on what should be included in evaluation statements of work (SOWs) and the expected coverage and structure of evaluation reports delivered to USAID. These ADS guidelines are used in this section as a basis for examining the quality completeness and quality of TCB evaluation reports.

The MSI team was able to locate 30 evaluations representing 38 (15 percent) of the 256 projects it examined. A list of these 30 project evaluations is provided in Table 27, which also indicates the evaluation’s scope. The evaluation team also found 11 additional TCB evaluations that were for projects it did not examine during the evaluation, bringing the total number of TCB project evaluations it located to 41. It is difficult to place this number in context, as USAID does not analyze the number of evaluation reports forwarded to the Development Experience Clearinghouse (DEC) each year on a sector or topical basis. Furthermore, it is the view of a number of USAID staff and consultants to the Agency who work with evaluations that not every evaluation USAID conducts is submitted to the DEC, despite the Agency’s mandatory requirement in ADS 203 in that regard.

USAID TRADE CAPACITY BUILDING EVALUATIONS LINKED TO 38 PROJECTS EXAMINED BY THE EVALUATION

Evaluation Title	Year	Evaluation Scope			
		Single Project; Single Country	Multiple Projects; Single Country	Single Project; Multiple Countries	Multiple Projects; Multiple Countries
SAGIC Mid-Term Evaluation	2009	●			
Madagascar Business and Market Expansion Project	2008	●			
Evaluation of the Cambodia Strengthening Micro, Small and Medium Enterprise Program	2007	●			
Cambodia MSME Project – Final Monitoring and Evaluation Report	2008				
Assessing the Impact of the Micro and Small Enterprise Trade-Led Growth Project	2008	●			
Montenegro Competitiveness Project	2008	●			

³ Notably, this is not unlike the findings of this evaluation with respect to the types of problems that most clearly impeded project success. The two answers that were statistically significant, start-up problems and problems with partners, are typical of development projects in all sectors.

Evaluation Title	Year	Evaluation Scope			
		Single Project; Single Country	Multiple Projects; Single Country	Single Project; Multiple Countries	Multiple Projects; Multiple Countries
Impact Evaluation: Rebuilding Agricultural Markets Program (RAMP)	2007	●			
Evaluation of Trade Hubs - Africa	2006				●
Evaluation Report for Macedonian Competitiveness Activity	2006	●			
Agriculture Natural Resource Management Program- Wula Nafaa	2006	●			
PROALCA II Evaluation	2006				●
GATE	2006				●
EXPOLINK – Growth Through Globalization Evaluation	2005	●			
An Evaluation of the Bangladesh Agro-based Industries and Technology Development Project (ATDP II)	2005	●			
USAID/Expro Project Evaluation	2005	●			
Evaluation of Agricultural Cooperatives in Ethiopia (ACE) Program	2005				●
ATR Mid-Term Assessment	2004	●			
Armenia Small and Medium Enterprise Market Development Project (ASME)	2004	●			
Evaluation of the Enterprise Development Project	2004			●	
The Future of the LEAPSS Project	2004	●			
Evaluation of Economic Growth Technical Assistance	2004		●		
Aid to Artisans Shape II Program Evaluation Report	2003	●			
Ghana Trade and Investment Reform Program (TIRP)	2003		●		
SAIBL – Evaluation & Impact Assessment	2003	●			
Southern Africa Global Competitiveness Hub	2003				●
Evaluation of Trade Facilitation and Investment Activity	2003			●	
Final Evaluation Report: U.S. Department of Commerce Commercial Law Development Program in Egypt (CLDP)	2003			●	
Mid-Term Evaluation: Private Farmers Assistance Program and Private Farmer Commercialization Program	2002		●		
Evaluation of the Firm Level Assistance Group (FLAG) Program in Bulgaria	2002	●			
Assessment of USAID's JOBS Program in	2002	●			

Evaluation Title	Year	Evaluation Scope			
		Single Project; Single Country	Multiple Projects; Single Country	Single Project; Multiple Countries	Multiple Projects; Multiple Countries
Bangladesh within the Context of the Market Development Approach					
Partnership for Economic Growth (PEG)	2002	●			

Perhaps the best way to place the number of USAID TCB evaluations in context is in comparison to the total number of evaluations USAID received. For this purpose, the evaluation team uses the number of evaluations for the period 2002-2008 that an MSI review of DEC holdings, undertaken for USAID's Office of Evaluation, determined were actually evaluations and not audits, final reports, or needs assessments. For 2002-2008 the number of DEC submissions that were judged to be true evaluations, using the same standard applied to counting USAID TCB evaluations, was 744: the 41 TCB evaluations the MSI team located represent 5.5 percent of that total.

USAID does not formally keep statistics on the number of evaluations carried out by sector each year or on the percentage of projects in a sector that have been evaluation. Accordingly, the evaluation team had no objective basis for concluding that the number of evaluations carried out for TCB projects over the past decade is higher or lower than the percentage of projects evaluated in other sectors over the same period.

As noted above, findings from the 30 evaluations covering 38 of the projects the evaluation team analyzed were used along with data found in project performance monitoring reports to prepare this evaluation report. They are part of the set of evaluations that the team examined in the review of existing evaluations described above; the MSI team also extracted a summary of the key characteristics of 15 of these evaluations from both a substantive and evaluation quality perspective in the fall of 2009. Findings from that first synopsis of evaluation quality highlighted the need for a more systematic review of USAID TCB evaluations from an evaluation quality perspective, since coverage and quality are the dimensions of USAID TCB reports the Agency will be in the best position to improve going forward.

To assess the quality of USAID evaluations and the evaluation SOWs to which evaluation reports respond, the MSI team used two scoring tools that it developed in connection with the Certificate Program in Evaluation course it regularly offers for USAID staff. These tools were also used by MSI to review SOW and evaluation quality for a sample of USAID evaluations for the Agency's Office of Evaluation. Subsections below present the findings of these two coverage and quality reviews. The ten USAID TCB evaluations included in this analysis were all evaluations for which both the SOW and the evaluation report had been located. Furthermore, they were all for single country, single project evaluations, so that the playing field was level from a scoring perspective.

TCB Evaluation SOWs Rated

The rating sheet MSI used to score USAID TCB evaluations derives from ADS 203.3.6.2 and 203.3.6.2 which, together, define what an evaluation SOW should cover. The rating system is a simple checklist. The SOWs the evaluation team scored using this instrument were those SOWs it found included as annexes in evaluations for TCB projects examined during Phase II. MSI only rated SOWs for evaluations for which it also rated the evaluation reports, as described below.

Table 28 tallies positive ratings on each checklist item across all of the evaluations rated. A longer line across evaluations on an item indicates that overall evaluation SOWs were consistent with ADS expectations on that item. A short line, in contrast, highlights where very few evaluation SOWs included some form of specific

instructions to evaluation teams that the ADS says are important. On this table, two factors that experience suggests have a significant impact on evaluation quality are highlighted in red, namely clarity about the management purpose of the evaluation and a specific list of questions to be addressed, which the ADS explicitly says should be a “small number of key questions and specific issues answerable with empirical evidence.”

While the evaluation team’s ratings for USAID TCB evaluation SOWs highlighted areas for improvement, over half of the SOWs reviewed scored well on many of the key evaluation elements, including the use of a list of evaluation questions as the centerpiece of an evaluation SOW and a clearly expressed management purpose to help keep the evaluation team focused on the report’s intended use.

SOW COMPLETENESS & QUALITY RATINGS FOR TCB EVALUATION SOWS

(n =10)

SOW Elements	Element Aspects Rated	Number of Projects									
		1	2	3	4	5	6	7	8	9	10
Project Description	1. Clear										
	2. Project Duration Provided										
Background	3. Problem Being Addressed										
	4. Development Hypothesis										
	5. Expected Inputs/Outcomes										
	6. Changes in Environment										
	7. Changes in Project										
Data Sources	8. Availability of Existing Data										
	9. Availability of Other Data										
Management Purpose	10. Management Purpose										
	11. Evaluation's Audience										
Evaluation Questions	12. Specific List of Questions										
	13. Appropriate number of Questions										
	14. Priority of Questions										
	15. Questions are Consistent w/ Purpose										
Methods	16. Responsibility for Methods Designated										
	17. Methods Recommended										
	18. Data Disaggregation										
	19. Samples/Analysis/Response Criteria										
Deliverables	20. Deliverables Specified										
	21. Deliverables Described										
	22. Evaluation Start/Completion Dates										
	23. Dates for All Deliverables Provided										
Team Composition	24. LOE Available or Team Size										
	25. Specific Skills Required										
	26. Evaluation Specialist Required										
	27. Whether and How Participatory										
Scheduling/ Logistics	28. Specific Dates Mentioned										
	29. Logistics Discussed or Not										
Report Requirements	30. Report Requirements										
	31. Dissemination Requirements										
Budget/LOE	32. Budget or LOE Provided										
Reasonableness	33. SOW Reasonable: Q/\$/Time										

TCB Evaluation Reports Rated

The rating system used to assess evaluation reports is conceptually parallel to the SOW rating system in that it involves a checklist keyed to ADS 203.3.6.6 which emphasizes the importance USAID places on

distinguishing between findings (the facts the evaluation team collected), conclusions (the team's interpretations or judgments as to what the findings mean), and recommendations. Reflecting the importance of this distinction, for example, the evaluation report rating system will detract points from evaluations that co-mingle these different elements in the evaluation structure USAID prefers.

The results of this analysis are shown in Table 29. For each item, the number of darkened squares equals the number of evaluation that included the evaluation element discussed. Most evaluations included some of these elements, and few evaluations included other elements. For the most part, a high number of projects that fail to include a given item can indicate that problems may exist, e.g., USAID staff and evaluation teams are unaware of the need to include the item or have difficulty doing so. Overall, the results of this scoring exercise are similar to results when USAID project evaluations from other sectors are scored using the same instrument or one that is very similar.

With respect to the types of evaluations being undertaken, the split found between formative and summative evaluations is fairly typical for all USAID sectors. As to the methods being used, particularly for summative evaluations, while they are similar to what is being done in other sectors, they tended to fall on the low end of the methods spectrum with respect to the evidence they produce. Most are carried out over a short time span and depend heavily on interviews. Very few used comparison groups to try to determine what changes or results might have occurred in the absence of USAID's project. Notable in this regard was one internal project evaluation carried out by Aid for Artisans:

- For USAID's AGEXPORT project in Guatemala, Aid to Artisans collected pre-and post project survey data for artisans that were and were not affiliated with the project. While the pre-project samples on which data were collected were of a reasonable size, this evaluation noted that it was difficult to obtain post-project survey data. The number of observations the evaluation obtained for beneficiary and non-beneficiary groups both before and after the project was thus small, making it difficult to draw strong conclusions.

As noted above, these types of features are difficult to construct for retrospective evaluations, but some tools for addressing questions about the counterfactual on a retrospective basis do exist. Of greater concern is the fact that some TCB summative evaluations did not systematically compare the project situation on a pre-and post project basis. This problem may well be linked to the lack of baseline data in a fairly large percentage of TCB projects, as discussed above. Compared to its recent summative evaluations, USAID formative TCB evaluations may serve as a reasonably good basis for making improvements in ongoing activities.

While this evaluation did not focus directly on the utilization of evaluations or performance monitoring data by USAID and implementing partner staff, project documents sometimes included references to their utilization to guide ongoing projects and design new ones. These ad hoc references do not, however, provide a basis for drawing conclusions about the extent to which USAID has utilized the TCB evaluations it has carried out.

What was perhaps somewhat more evident "reading between the lines" in program and project documents and listening to USAID and implementing partner staff in the evaluation's stakeholder consultations sessions was that USAID and its partners may not be learning as much from one others' experience as is desirable—including from evaluations or new empirical studies with implications for TCB programs. In discussions at the three stakeholder consultation sessions the evaluation held for USAID implementing partners, one of the most frequent comments made to MSI staff concerned how rarely USAID TCB implementing partners are brought together to talk about experience with specific types of TCB projects or to jointly discuss impediments to making progress on those objectives across countries.

Evaluation Elements		Number of Evaluations										Evaluation		Number of Evaluations											
	Identified	■	■	■	■	■	■	■	■	■	■		F&C	■	■	■	■	■	■	■	■	■	■		
	Methods linked to Q/Is	■	■	■	■	■	■	■	■	■	■			No new Fs or Cs	■	■	■	■	■	■	■	■	■	■	
	Data Limitations ID'D	■	■	■	■	■	■	■	■	■	■				Actor clear	■	■	■	■	■	■	■	■	■	■
Chart	Easy to Understand	■	■	■	■	■	■	■	■	■	■				Action clear	■	■	■	■	■	■	■	■	■	■
	T&G: N shown	■	■	■	■	■	■	■	■	■	■				Linked to Q/Is	■	■	■	■	■	■	■	■	■	■
Evaluation Type	Formative	■	■	■	■	■	■	■	■	■	■			Linked to Purpose	■	■	■	■	■	■	■	■	■	■	
	Summative	■	■	■	■	■	■	■	■	■	■		LL	Doesn't fit C&R	■	■	■	■	■	■	■	■	■	■	
	Joint, government	■	■	■	■	■	■	■	■	■	■			ES	Other places/times	■	■	■	■	■	■	■	■	■	■
	Joint, other donor	■	■	■	■	■	■	■	■	■	■		Evidence		ES is a snapshot	■	■	■	■	■	■	■	■	■	■
	Participatory, Beneficiaries	■	■	■	■	■	■	■	■	■	■				ES adds no new information	■	■	■	■	■	■	■	■	■	■
	External	■	■	■	■	■	■	■	■	■	■					Qualitative	■	■	■	■	■	■	■	■	■
	Internal	■	■	■	■	■	■	■	■	■	■					Quantitative	■	■	■	■	■	■	■	■	■
Data Collection	Document Review	■	■	■	■	■	■	■	■	■	■			Comparative		■	■	■	■	■	■	■	■	■	■
	>IP Data	■	■	■	■	■	■	■	■	■	■				Time Series Break	■	■	■	■	■	■	■	■	■	■
	>Other	■	■	■	■	■	■	■	■	■	■				Econometric	■	■	■	■	■	■	■	■	■	■
	Interviews	■	■	■	■	■	■	■	■	■	■					Forensic	■	■	■	■	■	■	■	■	■
	>Client	■	■	■	■	■	■	■	■	■	■					Before/After	■	■	■	■	■	■	■	■	■
	>IP Staff	■	■	■	■	■	■	■	■	■	■					Comparison Group	■	■	■	■	■	■	■	■	■
	>Officials/Experts	■	■	■	■	■	■	■	■	■	■					Control Group (RCT)	■	■	■	■	■	■	■	■	■
	>Beneficiaries	■	■	■	■	■	■	■	■	■	■														
	>Other	■	■	■	■	■	■	■	■	■	■														
	>In Instruments Annex	■	■	■	■	■	■	■	■	■	■														

Results of the evaluation report rating exercise summarized in Table 29 include several important findings:

- Evaluation Questions: The lists of evaluation questions from the SOW that USAID intends will be used to structure evaluation reports are not migrating into the front section of most USAID TCB evaluations.
- Types of Evaluation Questions: the majority of USAID TCB evaluations that included a set of questions in their SOWS focused on process questions and questions about actual results compared to planned results. Questions about causality were rare, even though USAID’s ADS emphasizes that monitoring is usually sufficient to tell us what happened in a project. Evaluations are the best tool for getting at “why?”
- Most of the evaluations relied on two sources of data: documents and interviews. The use of structured observation, small surveys (written questionnaires), measurement instruments, photographs, and other data collection techniques were less frequently reported.
- While both qualitative and quantitative data were used, most evaluations lacked comparisons, including basis pre-post comparisons and efforts to work back from result that are evident at the end of a project to check on whether the project or something else was the likely cause (forensic methods). None of these evaluations involved data from comparison or control groups, which OMB’s 2004 white paper, *What*

Constitutes Strong Evidence of Program Effectiveness?, and other critiques of USAID and development assistance evaluations generally cite as an important weakness: that is, they fail to examine the counterfactual.⁴ While none of the 30 evaluations of TCB projects MSI reviewed included the use of comparison groups, the evaluation team noted that one or two recent USAID TCB projects have shown an interest in these types of comparisons, including the collection of data on comparison groups as well as target groups into their regular baseline data collection and project monitoring activities, and with the intention of making post-intervention comparisons towards the end of the project funding period.

- With respect to the issue of separating findings, conclusions, and recommendations, the ratings indicate that the lines between these evaluation segments were not rigid, and in some instances readers would have encountered findings for the first time in conclusions or recommendations sections.

Looking beyond current TCB evaluations, adequate guidance exists within USAID and the in the evaluation community more broadly to support improvements in TCB evaluation quality. USAID's ADS guidance on developing evaluation SOWs and conducting evaluations, including its evaluation TIPS series, are available through the EvalWeb section of the Agency's website. USAID also offers courses that include performance monitoring and evaluation components. Further, USAID/EGAT and its TCBoost have training materials on monitoring and evaluation for TCB projects developed by the evaluation team and used in a pilot course during the study period. TCBoost has been authorized by USAID to use and modify them as appropriate to help improve TCB M&E. Beyond USAID, there are numerous evaluation resources available, often online, from other development assistance agencies, the OECD website section on aid for trade, and through evaluation associations, academic institutions, and the like.

As USAID experience shows, more than guidance is needed to improve the coverage, quality, and use of evaluations in USAID. USAID's Evaluation Office, with support from the Administrator, is currently engaged in an effort to enhance evaluation in USAID in both quality and coverage terms. This Agency initiative offers an opportunity for progress on TCB evaluations, but how TCB can best be linked to that opportunity should be carefully considered: initiatives that are perceived as increasing staff workload without delivering valuable benefits will be resisted.

Impact evaluations, which are relatively new to USAID, are more rigorous than current TCB evaluations. While it is not likely that they will be needed frequently, impact evaluation may from time to time be appropriate for determining the effects of types of field projects that USAID has undertaken frequently, but for which it lacks clear evidence about effectiveness. They are also appropriate in pilot projects where replication or scaling up is envisioned. Based on the findings of this study, there are a few types of TCB projects that might warrant consideration in this regard, e.g., USAID fairly consistently sends producer/exports to trade fairs, but systematic information on the impact of participation is generally lacking.

⁴ This OMB evaluation "white paper" is available at: http://www.whitehouse.gov/omb/part/2004_program_eval.pdf

ANNEX B: Evaluation Methods and Instruments

Table 2. Statement of Work Checklist Keyed to USAID ADS 203.3.6.3

SOW Elements and Sub-Elements	How Well is the SOW Element Addressed ⁵					Issues Noted by SOW Reviewer
	Complete		Partial	Incomplete		
	Above Average	Acceptable		Applicable but Missing	Not Applicable	
Identify the activity, project, or approach to be evaluated						
1. Is the SOW clear and specific about what is to be evaluated, e.g., activity/project/approach; funding mission/office; sector/topic; budget; target group/area? (looking at the big picture)						
2. Is the duration of the project or program stated in the SOW, i.e., start and end years?						
Provide a brief background on the development hypotheses and its implementation						
3. Is the SOW clear and specific about the problem or opportunity the activity/project/approach was expected to address?						
4. Does the SOW provide a clear description of the development hypotheses; intended results; critical assumptions, e.g., narrative, and/or Results Framework/Logical Framework?						
5. Does the SOW clearly describe the nature of the intervention, i.e., what USAID would deliver (training, TA, etc.) and what was expected to change (at the output and especially outcome levels)						
6. Does the SOW include information about changes in the project environment since the start of implementation, e.g., policy, economic, political, other donor program changes, or any natural disaster other changed assumptions.						
7. Does the SOW include information about changes in the activity/project design or implementation since the start of implementation, e.g., changes in budget; team; substantive modifications; relationships with other entities?						
Identify existing performance information source, with special attention to monitoring data.						
8. Is SOW clear and specific about existing						

⁵ Key: Missing = element was not covered in SOW; Partial = At least one key aspect was not covered; Acceptable = all aspects were covered; Above average = covered all aspects but went beyond basics in at least one way that is likely to help evaluators.

activity/project/approach (program) monitoring data/reports that are available, i.e., specific indicators tracked, baseline data, targets, progress towards targets; narrative quarterly/annual reports; and when/how evaluators can access these data?						
9. Does the SOW describe other documents or sources of information that would be useful to the evaluation team, e.g., government or international data USAID is using to monitor activity/project/approach outcomes, e.g., growth rate, poverty rate, etc.?						
State the purpose of, audience for and use of the evaluation						
10. Is the SOW clear and specific about why, in management terms, the evaluation is being conducted, i.e., what management decisions an evaluation at this time will inform? (ADS 203.3.6.1 identifies several management reasons why USAID might undertake an evaluation).						
11. Does the SOW indicate who makes up the audience for the evaluation, i.e., what types of managers in which organizations, e.g., USAID; Implementing Partner(s); the host government, other donors, etc., are expected to benefit from the evaluation and how?						
Clarify the evaluation question(s)						
12. Does the SOW include a list of the specific questions the evaluation team is expected to answer? [Please enter the number of question in the far right hand column.]						Number of Questions SOW asks the evaluation to address [count question marks]:
13. Is the SOW list of evaluation questions/directives consistent with USAID expectations about limiting the number asked? (ADS 203.3.6.2 says “a small number of key questions or specific issues answerable with empirical evidence.”) [Small is often considered to be less than ten; every question mark signals a question.]						
14. Does the SOW indicate the relative priority of the evaluation questions/directives, e.g., are they in priority order or are “top priorities” identified?						
15. As a group, do the evaluation questions/directives appear to be consistent and supportive of the evaluation’s purpose?						
Identify the evaluation methods (USAID may either specify methods or ask the evaluation team to suggest methods)						
16. Is it clear from the SOW whether USAID requires the use of specific data collection/analysis methods or is leaving such decisions up to the evaluators?						Describe:

17. Is the SOW clear and specific about any evaluation methods it recommends, e.g., does it state which methods are to be used to answer each question, OR which methods of analysis will be used with which type of data?						
18. Is the SOW clear and specific about any data disaggregation, e.g., by gender, or geographic region, etc., it requires?						
19. Is the SOW clear and specific about any samples (e.g., representative); analyses (comparison of means for two groups); or response criteria (significant at the .05 level) it mentions?						
Specify evaluation deliverable(s) and the timeline						
20. Are the deliverables for which the evaluation team is responsible clearly specified in the SOW?						
21. If deliverables in addition to a draft and final version of the report are required, e.g., detailed evaluation plan, summary of findings prior to drafting the report; oral briefings for stakeholder, are these deliverables clearly described?						
22. Does the SOW include information about expected start and completion dates for the evaluation?						
23. Are dates provided for all of the deliverables specified as evaluation requirements?						
Discuss evaluation team composition (one team member should be an evaluation specialist) and participation of customers and partners.						
24. Is the SOW clear about the LOE available or size of the team that is required for the evaluation?						
25. Are specific positions and/or skills the team is expected to include clearly defined, e.g., specific positions and associated qualifications including technical, geographic, language and other skill/experience requirements?						
26. Is the SOW explicit about requiring that one team member be an evaluation specialist?						
27. Is the SOW clear about whether and how USAID expects its staff; partners; customer/beneficiaries or other stakeholders to participate in the evaluation process (i.e. developing the SOW, collecting/analyzing data or providing recommendations)?						
Cover procedures such as scheduling and logistics						
28. Is the SOW clear and specific about any specific dates that need to be reflected in the evaluation team's plan, e.g., local holidays, specific dates for oral presentations already scheduled, etc.						

29. Is the SOW clear about whether space, a car or any other equipment will be made available to the team or that they must make their own arrangements?							
Clarify requirements for reporting							
30. Is the SOW clear about what it requires in the evaluation report, e.g., Executive Summary; SOW as an attachment; methodology description and instruments; list of places visited, language(s) in which the report is to be submitted, etc?							
31. Is the SOW clear about dissemination requirements, e.g., numbers of hard copies of final report needed; PowerPoint/handouts for oral briefings; submission to the DEC, etc.							
Include a budget							
32. Is the SOW clear about the total budget or at least the LOE available for the evaluation?							
Reviewer Sense of Reasonableness							
33. In the reviewer's judgment, is the relationship between the number of evaluation questions/directives, timeline and budget for this evaluation clear and reasonable?	Yes	No	Insufficient Information				

	Yes	No		Yes	No
Findings (F)			Lessons Learned		
• Separate section			• Doesn't fit w/ C&R		
• Linked to Q/Is			• For other places/times		
• All Q/Is addressed			Executive Summary		
• No Cs or Rs in section			• Is a snapshot of Report		
• Raw data analyzed			• Adds no new info		
• Data all methods used			Evidence of Change		
• N with %s in text			• Quantitative		
Conclusions (C)			• Qualitative		
• Separate from F&R			Evidence of Causality		
• Supported by F			• Comparative		
• Not just Fs restated			• Time Series Break		
• No new Fs			• Econometric		
• Not Rs			• Forensic (alt. causes)		
• Linked to Q/Is			Comparisons		
Recommendations (R)			• Before and After		
• Separate from F&C			• Comparison Group		
• Supported by F&C			• Control Group (RCT)		
• No new Fs or Cs					
• Intended Actor Clear					
• Action needed is clear					
• Linked to Q/Is					
• Linked to Purpose					
Other Notes and Comments on the Evaluation					