QUALITY REVIEW OF RECENT EVALUATION STATEMENTS OF WORK (SOWs)

MARCH 2010

This publication was produced for review by the United States Agency for International Development. It was prepared by Micah Frumkin and Emily Kearney with Molly Hageboeck (editor/advisor), Management Systems International.

Quality Review of Recent Evaluation Statements of Work (SOWs)

March 2010

Management Systems International
Corporate Offices
600 Water Street, SW
Washington, DC 20024

Contracted under RAN-M-00-04-00049-A-FY-05-86

CONTENTS

INTRODUCTION

This report examines the degree to which USAID evaluation Statements of Work (SOWs) conform to USAID ADS guidance concerning the preparation of SOWs. Results reported in this document were compiled as part of a background research activity that supported the development of a separate submission entitled: *Good Practice Examples for Preparing Evaluation Statements of Work (SOWs),* which is designed to help USAID staff draft evaluation SOWs. This Good Practice volume includes examples from SOWs taken from USAID evaluations submitted to the Development Experience Clearinghouse (DEC) between the latter part of 2008 and January 2010, and identified through the background research discussed in this report. It was initially expected that this research report would lead to the identification of several SOWs which were exemplary in all regards. As explained below, however, the research MSI undertook yielded exemplary elements or sections in quite a few SOWs, but no single "Golden SOW" emerged.

The report also describes a small pilot test that brought together the approach for rating SOWs described herein and an approach MSI used in 2008 to rate evaluation reports against USAID ADS guidelines.[1] Using a variation of the rating instrument employed in 2008, MSI rated evaluation reports which were associated with SOWs that this study had identified as being above average, average or below average. The purpose of this small pilot study was to determine whether SOWs receiving an above-average rating (with respect to their compliance with ADS guidelines) lead to better-than-average evaluation reports. Findings from that small pilot study are also reported in this document.

SECTION 1 – RATING EVALUATION STATEMENTS OF WORK (SOWS)

USAID ADS 203. ADS 203.3.6.3 clearly states what is to be included in a SOW. These are:

- Identify the activity, project, or approach to be evaluated;
- Provide a brief background on the development hypothesis and its implementation;
- Identify existing performance information sources, with special attention to monitoring data;
- State the purpose of, audience for, and use of the evaluation;
- Clarify the evaluation question(s);
- Identify the evaluation method(s);
- Specify the deliverable(s) and the timeline;
- Discuss evaluation team composition (one team member should be an evaluation specialist) and participation of customers and partners;
- Cover procedures such as scheduling and logistics;
- Clarify requirements for reporting and dissemination; and
- Include a budget.

To determine how well recently-developed evaluation SOWs comply with these requirements - a number of which are more fully explained in other sections of the ADS - MSI used a 33-item variation of a SOW score-

---

[1] The results of the 2009 exercise were included as an annex to MSI's report for USAID entitled: *Trends in International Development Evaluation Theory, Policy and Practices.*

sheet originally created for USAID/Uganda under a long-term, Mission-funded monitoring and evaluation contract. The rating sheet took these eleven elements into account, expanding each into several questions that probe deeper into the details of the precise elements that should be included in each section of the SOW. For example, the ADS instruction to identify the evaluation methods was broken out into several questions addressing: 1) whether the SOW is clear about methods that must be used, 2) whether the evaluation team determines the methodology, 3) if data must be disaggregated, and 4) if any samples must be taken. The score sheet also included an additional question asking the scorer to assess the overall reasonableness of the evaluation based on the number of evaluation questions, the time available for the evaluation, and the budget for the evaluation. A copy of this rating instrument is included as Annex A. For most questions on the SOW rating sheet, the marking range for each question includes a four-point "above average," "acceptable," "partial," or "applicable but missing" scale. To ensure inter-rater reliability in scoring, a handbook was created that provided consistent directions as to how scoring should be conducted. This handbook can be found in Annex B.

MSI applied this instrument to 44 USAID evaluation SOWs for evaluations posted to the DEC between the latter part of 2008 and January 2010. These 44 SOWs represent 42% of the 105 late 2008 and all 2009 evaluations from the DEC that MSI examined for this report. In other words, 58% of these evaluations did not have attached SOWs.

In the analysis phase the team found that, for reporting purposes, a two-point acceptable/unacceptable scale was more useful than the four-point system raters had initially used when scoring the SOWs. All "partial" scores were thus re-analyzed and re-categorized into one of the two sides of this scale. "Above average" scores were combined with those scored "acceptable." The remaining low-scores were then treated as being unacceptable in terms of their compliance with the ADS SOW guidance. Once this process was completed for all 44 SOWs, they were ranked by their scores, and the three most highly-scored SOWs were set aside as the strongest representations of a quality SOW. These three are identified with an asterisk in the list of the 44 evaluations included in this study, which is provided in Annex E.

A key finding of this analysis was that none of the 44 SOWs achieved a rating of acceptable across all eleven SOW element categories. It had initially been expected that the application of this rating system would lead to the discovery of a number of SOWs that could be provided to USAID in response to its request for exemplary SOWs that could be shared with USAID staff.

Figure 1 below shows the results of MSI's tabulation of ratings for each of the 33 rating items across all 44 of the SOWS it scored. This figure illustrates visually, as well as mathematically, where the greatest strengths and weaknesses of recent SOWs lie.

Figure 1: Table of SOW Strengths and Weaknesses

Percentage of SOWs deemed acceptable for each question within each section recommended by ADS 203.3.6.3 (n=44)

Project Description	1. Clear	73%
	2. Project Duration Provided	80%
Background	3. Problem Being Addressed	73%
	4. Development Hypothesis	64%
	5. Expected Inputs/Outcomes	64%
	6. Changes in Environment	27%
	7. Changes in Project	34%
Data Sources	8. Availability of Existing Data	57%
	9. Availability of Other Data	20%
Management Purpose	10. Management Purpose	89%
	11. Evaluation's Audience	70%
Evaluation Questions	12. Specific List of Questions	66%
	13. Appropriate number of Questions	11%
	14. Priority of Questions	11%
	15. Questions are Consistent w/ Purpose	82%
Methods	16. Responsibility for Methods Designated	77%
	17. Methods Recommended	11%
	18. Data Disaggregation	20%
	19. Samples/Analysis/Response Criteria	7%
Deliverables	20. Deliverables Specified	86%
	21. Deliverables Described	75%
	22. Evaluation Start/Completion Dates	66%
	23. Dates for All Deliverables Provided	61%
Team Composition	24. LOE Available or Team Size	68%
	25. Specific Skills Required	57%
	26. Evaluation Specialist Required	57%
	27. Whether and How Participatory	50%
Scheduling/ Logistics	28. Specific Dates Mentioned	9%
	29. Logistics Discussed or Not	39%
Report Requirements	30. Report Requirements	68%
	31. Dissemination Requirements	50%
Budget/LOE	32. Budget or LOE Provided	45%
Reasonableness	33. SOW Reasonable: Q/$/Time	7%

While none of the SOWs were deemed to be acceptable on all 33 elements, several came close, with the highest of them achieving an overall rating of 26 out of 33. As for which elements received the greatest proportion of acceptable ratings, the ADS expectation that an evaluation SOW would identify a management purpose for undertaking the evaluation stood out. Of the 44 SOWs scored, 89% received a rating of acceptable on this element. In all, 70% or more of the SOWs were rated as being acceptable on nine of the 33 rating items (27%).

On the other side of the scale, there was a tie for the question whose results were least commonly found acceptable. Only 7% of SOWs were rated as acceptable on 'Methods provided for Samples/Analysis/Response Criteria,' which addresses whether SOWs included criteria or standards the evaluation data were expected to meet (e.g., beneficiary survey responses must be statistically representative of the target population). The other was 'Reasonableness based on: Qs/$/Time,' which refers to the critical linkages between the number of evaluation questions asked and the evaluation schedule and budget. This element is one which called upon raters to integrate what they knew about each SOW they had scored and reach a judgment; as such, this element and its scoring is probably best treated somewhat separately from rating items that involved direct scoring against ADS elements.

When item scores are aggregated into categories corresponding to the SOW elements specified in the ADS, the lowest ratings were in the areas of Scheduling/Logistics and Methodology. What this illustrates is that, in practice, SOWs written over the last few years have tended to include quite a bit of flexibility. Key decisions, such as what methods are most appropriate and the specific schedule and relevant dates, are not always predetermined in SOWs. Presumably this is to allow USAID staff an opportunity for interaction with an evaluation team before such decisions are finalized.

With respect to the reasonableness item that focuses on the triangulation of questions, budget and schedule, the low scores in this area indicate that the number of questions included in SOWs are more than can plausibly be addressed with any serious rigor or level of confidence within the time and/or budget allotted for the evaluation. Linked to this, there was a research finding that only 11% of the 44 SOWs were found to be acceptable for the question addressing whether the SOW included an 'appropriate number of questions.' The criteria for 'acceptable' was ten questions or less, based on the directive in ADS 203.3.6.2 to *identify a small number of key questions* in an evaluation SOW. When applied, all question marks were counted. MSI found that in a reasonably large number of the SOWs it scored, USAID clusters several questions together in a group. There may have been only six or seven groups of questions, but the actual number of individual and distinct questions in this type of situation was sometimes 20 or more, and a few included closer to 60 separate evaluation questions, when every question mark in that section was considered.

On a related note, the research showed that most evaluation SOWs do not explicitly state whether their evaluation questions are presented in order of their priority. Prioritization is treated in the rating system as a dimension of clarity. While prioritizing five or six questions may not seem terribly important, when the number of questions comes closer to 30, prioritization becomes critical. Without it, evaluators, pressed for time and faced with too many questions to realistically address, may set their own priorities, and those priorities may not turn out to match USAID's. This lack of specificity as to what the evaluation team is meant to address directly relates to the weaknesses in "reasonableness" of the SOWs.

SECTION 2 – RELATIONSHIP BETWEEN SOW QUALITY AND EVALUATION QUALITY

USAID efforts to improve its staff's ability to write evaluation SOWs, as well as MSI's emphasis on SOW preparation in the Certificate Program in Evaluation it presents for USAID audiences, presume that the quality of SOWs has a direct impact on the quality of the evaluation products USAID receives. Capitalizing on both the analytic work on SOWs undertaken for this report, and MSI experience in scoring evaluation reports against USAID ADS guidance, MSI undertook a small pilot test to determine whether any patterns emerged when the scores for SOWs were compared to the scores for the evaluation products that resulted from them.

For this pilot, MSI modified an evaluation rating tool it has used previously (e.g., in the 2009 Evaluation Trends report annex on evaluation quality and for participant exercises in the Certificate Program in Evaluation); this modification focused on evaluation methods used in the evaluations, drawing in part on an approach used in a 2005 study that examined the quality of USAID Democracy/Governance evaluations and found them lacking.[2] A copy of this updated evaluation rating checklist is included as Annex C, together with MSI's guide for rating evaluations using this instrument as Annex D.

To pilot test this revised instrument and compare SOW ratings to the ratings for evaluations based on those SOWs, MSI defined a small criteria sample of cases to examine. The established criteria focused on SOWs from among the 44 MSI had coded that had received high overall ratings, SOWs that had received low overall ratings and SOWs that had received average ratings. Laid out on a line, in quartiles, the numbers on the high and low scoring ends (three each) and those in the middle two quartiles (38) somewhat resembled a bell shaped curve. From these, MSI selected only single project evaluations for the pilot, leaving aside multi-project evaluations. In all, MSI selected eleven projects to include in the pilot: the one project with a low SOW score that was also a single project evaluation; two projects with high scores that were single project evaluations and eight from the middle two quartiles, selected at random.

While this was a small sample, the inclusion of all available high and low SOW rating cases created an opportunity to see if a clear association existed between high/low rated SOWs and high/low evaluation report ratings. MSI then scored the evaluation reports to which each of the projects in this pilot sample was linked. To ensure inter-rater reliability, each report was read and scored by two different team members. The resulting score sheets were checked against one another in order to verify that scoring was being done consistently. If conflicting scores arose then the scorers discussed their findings, referenced the score-sheet handbook, and determined the appropriate score. MSI then analyzed the scores for these eleven evaluations, using the same type of visual tool it used in Figure 1 to present the results of its scoring of 44 SOWs. Findings from the evaluation ratings on this small sample of eleven evaluations using MSI's updated scoring instrument are presented in Figure 2 below.

As Figure 2 illustrates, evaluations rated high on some factors and lower on other factors. Also of note, the updated evaluation rating checklist worked as had been hoped, providing a more detailed picture of the types of evidence evaluations use to reach conclusions and recommendations.

---

[2] Bollen, Kenneth, Pamela Paxton, and Rumi Morishima. 2005. Assessing International Evaluations: An Example from USAID's Democracy and Governance Program. *American Journal of Evaluation* 26 (2):189-203.

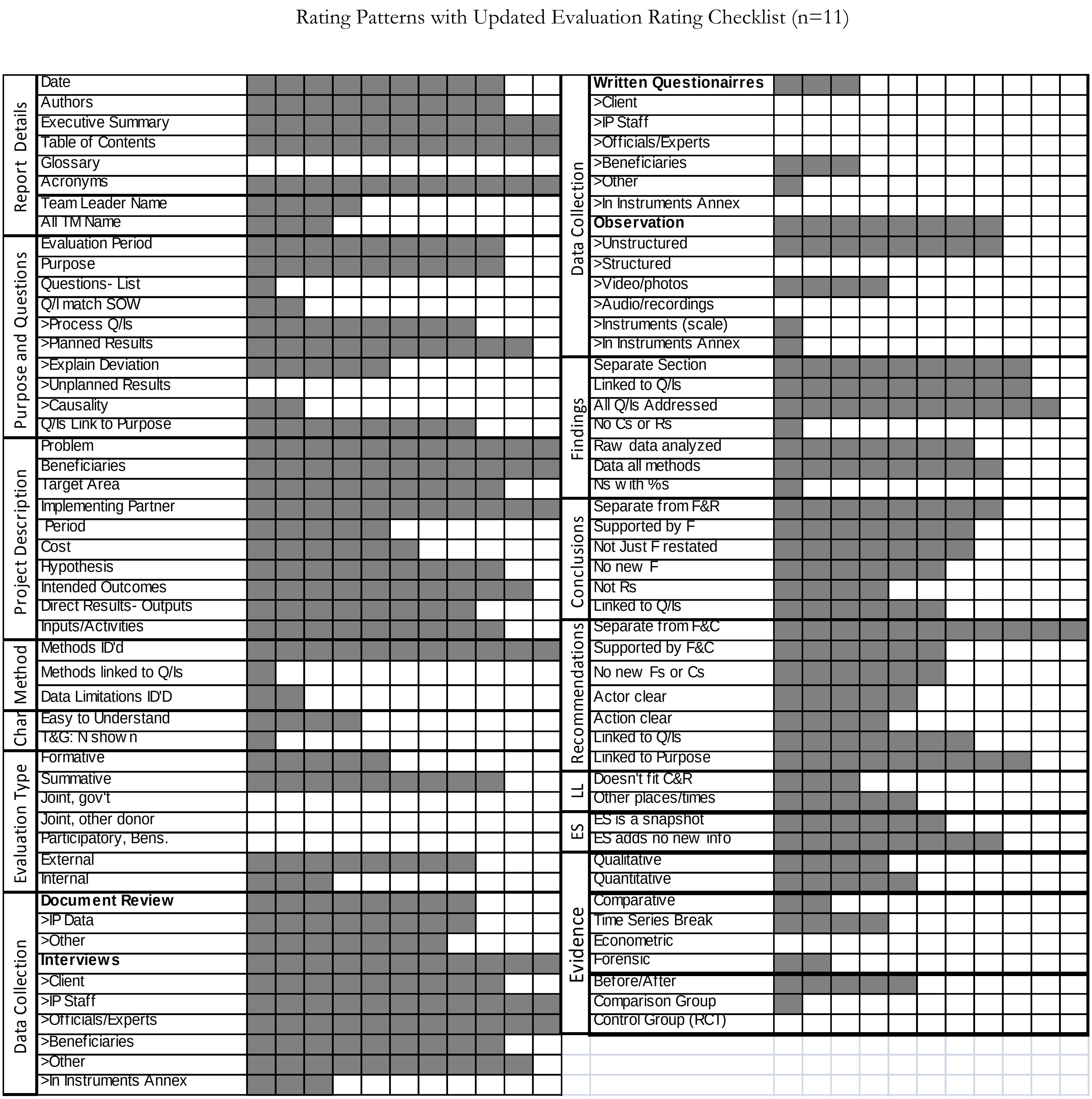
For purpose of this report, however, the focus was not on evaluation scores per se, but rather on the relationship between SOW ratings and evaluation report ratings.[3] When completed, the evaluation ratings for these eleven evaluations were compared to the ratings given to the SOWs that led to them.

For each of the eleven reports for which both SOW and evaluation reports ratings were available, the team looked at evaluation narratives as well as ratings to determine whether and how the quality of a SOW affected evaluation report ratings. What this analysis found, was that there is no clear pattern or predictive relationship between SOW and evaluation ratings. An evaluation report that receives a high rating can emerge even when the rating for the SOW that launched the evaluation receive a low rating. Conversely, a SOW with a high rating does not automatically yield an evaluation report that receives a high rating. Average and good SOWs in this small sample were not good predictors of evaluation report quality. This finding does not support the hypothesis that this pilot review, of the relationship between SOW and evaluation report ratings, set out to test. A repeat test on a larger sample might more-readily detect patterns in relationships among individual scoring items, but a larger sample would not negate the basic finding of no clear association at an aggregate score level already evident from this pilot.

The counterintuitive finding from this small pilot effort raises more questions than it answers. A narrower examination of the relationship between the SOW its evaluation product, focused primarily on evaluation questions, may be more revealing in this regard.

[3] MSI plans to further analyze its results using this updated instrument compared to results using earlier versions, and other evaluation report rating approaches, but this will be done outside the context of this report.

Figure 2: Table of Evaluation Strengths and Weaknesses



ANNEXES

ANNEX A: SOW SCORE-SHEET

Statement of Work Checklist Keyed to USAID ADS 203.3.6.3

Project or Program to be Evaluated: ____

Person Who Reviewed the SOW:________________ Date of Review: ____________________

SOW Elements and Sub-Elements	How Well is the SOW Element Addressed[4]					Issues Noted by SOW Reviewer
	Complete		*Partial*	*Incomplete*		
	Above	***Acceptable***		***Applicable \Missin***	***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.						

[4] 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.

SOW Elements and Sub-Elements	How Well is the SOW Element Addressed[4]					Issues Noted by SOW Reviewer
	Complete		***Partial***	***Incomplete***		
	Above	***Acceptable***		***Applicable Missing***	***Not Applicable***	
8. Is SOW clear and specific about existing 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. [count question marks]:
13. Is the SOW list of evaluation questions 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 each evaluation questions, e.g., are they in priority order or are "top priorities" identified?						
15. As a group, do the evaluation questions 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						

SOW Elements and Sub-Elements	How Well is the SOW Element Addressed[4]					Issues Noted by SOW Reviewer
	Complete		Partial	Incomplete		
	Above	Acceptable		Applicable Missing	Not Applicable	
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						

SOW Elements and Sub-Elements	How Well is the SOW Element Addressed[4]					Issues Noted by SOW Reviewer
	Complete		*Partial*	*Incomplete*		
	Above	***Accept able***		***Applica ble Missin***	***Not Applica ble***	
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, timeline and budget for this evaluation clear and reasonable?	Yes		No	Insufficient Information		

ANNEX B: SOW SCORE-SHEET HANDBOOK

Statement of Work Handbook Keyed to USAID ADS 203.3.6.3

SOW Elements and Sub-Elements	How Well is the SOW Element Addressed[5]
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?	Does the SOW provide a big-picture description of the project, touching on who funds the project and the funding levels, what is the project focused on, and what type of project is it.
2. Is the duration of the project or program stated in the SOW, i.e., start and end years?	If the project has not come to an end, then it is acceptable for the SOW to address the expected end dates, or project duration, provided that the start date was provided, i.e. a five-year project that began in 2006.
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?	This should discuss the reason that an intervention is necessary, be it extreme corruption, high levels of HIV/AIDS or malaria, etc. It should be clear what exactly the project will 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?	We are looking for an if-then statement clarifying what the result of a specific intervention will be. If we do _____ then _____ will occur. Visual depictions such as results frameworks or logical framework diagrams also work.
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)	This takes the development hypothesis to the next level and describes the exact nature of the intervention (technical assistance or trainings, etc) and the expected results preferably at outcome levels such as increased capacity for ______.
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 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?	To receive full credit, the SOW must provide information on how the documents will be accessed/provided. Simply listing the names of documents is not sufficient.
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).*	To say that an evaluation is being conducted since it is required, or to evaluate so that it is evaluated is not sufficient. Management purposes include: to redesign an existing activity, to provide insight for a follow-on activity, to determine impact, etc.
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?	A single audience is sufficient, since there are not always multiple audiences. This information may come out in the section discussing report dissemination.
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.]	These must be questions with a question mark at the end to be counted as a specific list of questions. Every question mark should be counted as a question.
13. Is the SOW list of evaluation questions 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 each evaluation questions, e.g., are they in priority order or are "top priorities" identified?	

15. As a group, do the evaluation questions appear to be consistent and supportive of the evaluation's purpose?	This relates directly to the Management Purpose addressed earlier. Not every question must address the purpose, but the majority should.
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?	We are looking for the designation of authority. Who makes the decision as to who determines methodology? Even if recommending methods, decisions can still be up to the team.
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?	Even if the methodology is up to the team, often the SOW will provide recommendations on how data should be collected or analyzed. If one or the other is included then it is acceptable.
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?	If any of these issues are touched on, they must be explained in the SOW to be deemed acceptable.
Specify evaluation deliverable(s) and the timeline	
20. Are the deliverables for which the evaluation team is responsible clearly specified in the SOW?	Details for each deliverable are not necessary, but rather the deliverables should be included and obvious that they are deliverables that the team is required to produce.
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?	Additional or custom deliverables must be described so that the team knows exactly what they are responsible for, particularly for uncommon deliverables.
22. Does the SOW include information about expected start and completion dates for the evaluation?	This can be in the format of a start date and end date, a start date or end date and the duration, or the duration and the general time period during which it is to be conducted.
23. Are dates provided for all of the deliverables specified as evaluation requirements?	Specific dates are not necessary, and dates can be referenced as: due within three days from the beginning of field work, or final report due within seven business days of receiving comments.
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?	One or the other is acceptable, both are above average.
25. Are specific positions and/or skills the team is expected to include clearly defined, e.g., specific positions and associated qualifications including	A list of skills or qualifications is acceptable even if not associated with a specific position.

technical, geographic, language and other skill/experience requirements?	
26. Is the SOW explicit about requiring that one team member be an evaluation specialist?	Specialist can include expertise or significant experience and does not mean that they are a specialist solely in evaluation.
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)?	Accompanied by staff etc. is not sufficient, but rather there should be some comment as to how they will be involved in the evaluation process such as data collection or analysis.
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.	These could be any dates around which the evaluation team must schedule their evaluation
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.	This would include anything above and beyond the delivery of a single electronic report.
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, timeline and budget for this evaluation clear and reasonable?	If even one of the three is missing then there is insufficient information. If all are present, then best judgment is required.

ANNEX C: EVALUATION SCORE-SHEET

Meta Evaluation Quick Check Form

	Yes	No		Number
Date of Report Visible			Executive Summary Pages	
Authors' Names Visible			Report Pages (without annexes)	
Executive Summary			Team Size	
Table of Contents			Evaluation Questions in Report	
Glossary				
List of Acronyms				True
			Type = Formative	
Team Leader Name			Type = Summative	
Team Members Names			Type = Joint (government)	
			Type = Joint (other donor)	
Evaluation Period (Dates)			Type = Participatory (beneficiaries)	
Evaluation purpose stated			Team = External (Outsiders)	
Evaluation questions - list			Team = Internal (Insiders – Staff/IPs)	
Q/Is on list match SOW			***Details on Methods Used***	**Yes, No, Number**
• Process Q/Is			***Document Review***	
• Planned Results Q/Is			• IP performance data	
• Explain Deviation Q/Is			• Other sources	
• Unplanned Results Q/Is			***Interviews***	
• Causality Q/Is			• Evaluation client	
• Q/Is link to Purpose			• IP Staff	
Program Description			• Officials/Experts	
• Problem addressed			• Beneficiaries	
• Intended beneficiaries			• Other	
• Target area (map)			• In Instruments Annex	
• Implementing Partner			***Written Questionnaires***	
• Intervention period			• Evaluation client	
• Intervention cost			• IP Staff	

• Causal Hypotheses			• Officials/Experts	
• Intended outcomes			• Beneficiaries	
• Direct results - outputs			• Other	
• Inputs/Activities			• In Instruments Annex	
Methodology			***Observation***	
• Methods identified			• Unstructured	
• Methods linked to Q/Is			• Structured (form)	
• Data limitations ID'd			• Video/Photos	
Tables or Graphs			• Audio/recordings	
• Easy to Understand			• Instruments (scale)	
• N shown on Ts & Gs			• In Instrument Annex	
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				

ANNEX D: EVALUATION SCORE-SHEET HANDBOOK

Meta Handbook

Date of Report visible: Is the date of the report provided anywhere in the report (including cover pages), specified to at least the month and year?
Authors' names visible: Are the names of the author(s) provided in the report? This will usually be on a cover page and may or may not be the same as the evaluation team members.
Executive Summary: Is there an executive summary present?
Table of Contents: Is a table of contents present? It should indicate sections of the report and associated page numbers.
Glossary: Is a glossary or section dedicated to the definition of words provided? This should not include acronyms.
List of Acronyms: Is a list of all acronyms and their meanings provided in the report?
Team Leader Name: This is one individual (unless explicitly stated as two) assigned to lead the team. It should be very apparent which team member is assigned this position.
Team Members Names: Every member of the team should be named as part of the evaluation team. Two of four members being named is not sufficient. Also, a report claiming "this report prepared by …" is not sufficient. They must be explicitly named as evaluation team members. This may be found in the cover pages or in a section near/in "methodology."
Evaluation Period (Dates): The dates and duration of the evaluation should be provided, or a combination thereof, so as the reader knows either 1) when it began and ended 2) when it began and how long it lasted for, or 3) the length of the evaluation and the end date.
Evaluation purpose stated: This is the management purpose and should indicate how the client will use the information, e.g., to inform future decisions or a redesign, etc. Conducting an evaluation to conduct an evaluation is not sufficient.
Evaluation questions – list: Is a list of the evaluation questions provided? These should be questions and not instructions or directives, meaning they should all have a question mark. Also, it should be a list of one question after another with no text or narrative in-between.
Q/Is on list match SOW: If the list is provided, does it match the list from the SOW exactly (word for word)? Even if not questions, does the team claim they will address the exact issues outlined in the SOW, be they questions, instructions, directives etc.?
• **Process Q/Is:** These are questions or instructions pertaining to the management/implementation of the project. These may look at efficiencies or relationships with stakeholders/partners, etc.
• **Planned Results Q/Is:** These questions or instructions look at whether the intervention is reaching indicated goals or objectives such as IRs or AOs. For example, was the project effective at reaching the targeted change in agriculture production?

• **Explain Deviation Q/Is:** These questions or instructions address ***WHY*** targets/ objectives were or were not reached.
• **Unplanned Results Q/Is:** These questions or instructions look at whether or not there were unexpected results from the intervention, and perhaps why.
• **Causality Q/Is:** These questions or instructions look at whether or not the project or intervention was the cause of the observed results. Or, could these results have been caused by other donors, other projects, or something to do with the external environment?
• **Q/Is link to Purpose:** Do the questions as a group reflect the management purpose stated earlier? 'As a group' indicates that not all questions must address this purpose, but rather that the majority of them do.
Program Description This refers to the intervention itself and not the evaluation
• **Problem addressed:** Is the problem that the project or intervention was created to address clearly described?
• **Intended beneficiaries:** Are the intended beneficiaries clearly indicated so that the reader knows exactly who the project is meant to help? These are the end users of the outcomes of the interventions.
• **Target area (map):** Is it clear to the reader where exactly the intervention is taking place? A single country can be named, but **only** if it is a nationwide project. A localized project working only in specific areas must identify those areas. In that case, naming a country is insufficient. While a map is extremely useful, it is not necessary.
• **Implementing Partner:** The implementing partner is the organization contracted by USAID and responsible for the management and implementation of the intervention.
• **Intervention period:** This period must include the specific start date of the intervention and the specific (or projected) end date of the intervention. If a contract amendment or modification took place that changed the dates of the project, it should be addressed here.
• **Intervention cost:** This is the total value of the cost of the intervention. If a contract amendment or modification took place that changed the funded value of the project, it should be addressed here.
• **Causal Hypotheses:** This is the theory under which the intervention was implemented and through which it will accomplish its stated objectives. Often an "if-then" statement, it could also be displayed visually as a results framework or logical framework. Trigger words include: accomplish ___ through, if…then, in order to, etc.
• **Intended outcomes:** These are the higher level objectives that the intervention is striving to reach. They may include IRs and AOs. They are not outcomes.
• **Direct results – outputs:** Outputs are the direct results of inputs and may include the number or people trained or the amount of medication provided.
• **Inputs/Activities:** Inputs are the direct efforts put forth through the intervention, be they hosting trainings for X number of people or providing technical assistance or medical services to X number of people.
Methodology
• **Methods identified:** Is it clearly stated what methods the evaluation team used both for data collection and data analysis? This should include interviews, surveys, document reviews, observations, and any samples that may have been used. A separate section in the report is most likely dedicated to this topic.
• **Methods linked to Q/Is:** Does the team indicate what methods were used to collect or analyze data for each specific question or instruction? This could be provided in either the methodology section or in the section write-up for each specific question/instruction.

• **Data limitations ID'd:** Data limitations should be provided in a report as a caveat to what the team was unable to do and why. This could indicate the unavailability of materials they were expecting to receive, unwillingness of beneficiaries to comment, or even unexpected weather conditions impacting travel.
Tables or Graphs
• **Easy to Understand:** A reader should be able to look at a table or graph and understand exactly what it is portraying without forcing the reader to look in the text for context. This means clearly labeled titles and axes.
• **N shown on Ts & Gs:** It is crucial that any table or graph representing data that the team collected display the "N," or the number of people from which the data was collected. •
Executive Summary pages: This is the number of pages of the executive summary. The pages, if not numbered, should be counted. Any text (even one word) counts as a page. Only whole numbers should be provided.
Report pages (without annexes): Count the number of pages in the body of the report up until the annexes. The Executive Summary does not count and should not be included. Only whole numbers should be provided.
Team Size (evaluators): How many people were considered part of the evaluation team? To be considered part of the team, they must have participated in data collection, data analysis, report writing or another significant part of the evaluation process beyond joining the team at site visits with no indication of their role.
Evaluation Questions in Report: How many evaluation questions did the team claim they addressed? These are questions and not instructions or directives. If not questions, do not count them.
Type = formative: A formative evaluation asks questions about how to look forward to a new intervention or to redesign an existing intervention. While it typically is conducted in the middle of the intervention, this is not always the case, and it should not be called a mid-term evaluation. An evaluation may be solely formative or may also be summative in nature.
Type = summative: A summative evaluation typically asks questions about whether an intervention reached stated objectives or may ask if the effects are sustainable without the intervention present. Possibly also dealing with scaling up, lessons learned or best practices. While it likely takes place at the end of an intervention it is not always the case and should not be called a final evaluation. An evaluation may be solely summative or combined with a formative evaluation.
Type = joint (government): This type of evaluation is conducted and/or commissioned collaboratively by a donor and the host country government.
Type = joint (other donor): This type of evaluation is conducted and/or commissioned collaboratively by two or more donors.
Type = participatory (beneficiaries): This type of evaluation utilizes the local beneficiaries' skills and resources in the process of conducting the evaluation. This may take place in the form of drafting the SOW, collecting/analyzing data, or providing input in the recommendations.

Team = external (outsiders): An external evaluation team consists solely of professional evaluators and experts contracted by the evaluation client to conduct the evaluation. An evaluation team may consist of both internal and external members, though both categories should then be checked.
Team = internal (insiders – staff/IPs): An internal evaluation team consists of solely USAID staff or Implementing Partners of the intervention. These are people that have intimate knowledge of the Agency and/or the intervention. An evaluation team may consist of both internal and external members, though both categories should then be checked.
Document Review
• **IP performance data:** These documents include all intervention reports such as PMPs and annual or quarterly reports. Also included are reports or publications created under the auspices of the intervention. Please provide the number of these documents reviewed. If it cannot be determined, please put a "?"
• **Other sources:** Other sources may include general country data, reports or data from other donors or development agencies, specialized sector specific reports, or any additional information the team deems relevant that was not created by the intervention. Please provide the number of these documents reviewed. If it cannot be determined, please put a "?"
Interviews: Interviews are direct interactions between an evaluator and a person or people of interest often based around a specific set of question. For the purpose of this tool, interviews may include individual interviews, group interviews or focus groups.
• **Evaluation client:** For this exercises general purposes, this will include USAID staff. Please provide the number of these interviews conducted. If it cannot be determined please put a "?"
• **IP Staff:** This is anyone working directly for, or subcontracted through, the Implementing Partner. Please provide the number of these interviews conducted. If it cannot be determined, please put a "?"
• **Officials/Experts:** Officials or Experts (aka Key Informants) are people that can speak at a higher level about subjects that the general person could not. This could be a Minister of Health or an expert in Female Reproductive Health Policy, for example. Please provide the number of these interviews conducted. If it cannot be determined, please put a "?"
• **Beneficiaries:** These are the people who are receiving the benefit of the interventions. They may be the people trained or those who receive the benefits of people being trained. They are the end users of the results of objectives being reached. Please provide the number of these interviews conducted. If it cannot be determined, please put a "?"
• **Other:** This category includes people who have valuable input, but are not beneficiaries or officials/experts. They may be employees in a ministry of health (but are not the minister), or other donors or staff of other interventions, etc. Please provide the number of these interviews conducted. If it cannot be determined, please put a "?"
• **In Instruments Annex:** Are all of the instruments used to collect data provided in an annex?
Written Questionnaires: Questionnaires are surveys or other forms of questioning persons of interest in a written manner (even if responses are obtained orally). Methods may include direct mailing or internet, among other options.
• **Evaluation client:** For this exercises general purposes, this will include USAID staff. Please provide the number of these questionnaires distributed. If it cannot be determined please put a "?"

• **IP Staff:** This is anyone working directly for or subcontracted through the Implementing Partner. Please provide the number of these questionnaires distributed. If it cannot be determined, please put a "?"
• **Officials/Experts:** Officials or Experts (aka Key Informants) are people that can speak at a higher level about subjects that the general person could not. This could be a Minister of Health or an expert in Female Reproductive Health Policy for example. Please provide the number of these questionnaires distributed. If it cannot be determined, please put a "?"
• **Beneficiaries:** These are the people who are receiving the benefit if the interventions. They may be the people trained or the ones who receive the benefits of people being trained. They are the end user of the results of objectives being reached. Please provide the number of these questionnaires distributed. If it cannot be determined, please put a "?"
• **Other:** This category includes people who have valuable input, but are not beneficiaries, nor are they officials/experts. They may be employees in a ministry of health, but are not the minister. They may be other donors or staff of other interventions etc. Please provide the number of these questionnaires distributed. If it cannot be determined, please put a "?"
• **In Instruments Annex:** Are all of the instruments used to collect data provided in an annex?
Observation: Observations are the silent collection of data through watching and listening without necessarily engaging people or disrupting normal habits or routines
• **Unstructured:** This is an informal approach where the team will conduct site-visits or watch interactions or activities to gain a general knowledge and not any specific qualitative or quantitative information, even if there is an underlying goal of the observations. Please provide the number of these observations conducted. If it cannot be determined, please put a "?"
• **Structured (form):** A structured observation looks at specific quantifiable or qualitative items/interactions/activities. This type of observation should be accompanied by a form or checklist. For example, it may look at how many beds are in a clinic, how many people are in a room, or the nature of the interaction between patient and doctor. Please provide the number of these observations conducted. If it cannot be determined please put a "?"
• **Video/photos:** If a photo of the intervention is included in the body of the report, this is sufficient. Otherwise specific mention of photos taken would have to be included. Videos are the same and may even be included in some documents. Please provide the number of these observations conducted. If it cannot be determined, please put a "?"
• **Audio/recordings:** While audio recordings can be included in a report, it is rare and most likely would have to be specified in the report. Please provide the number of these observations conducted. If it cannot be determined please put a "?"
• **In Instrument Annex:** Are all of the instruments used to collect data provided in an annex?
Findings (F): Findings are irrefutable and empirical data collected and/or analyzed by the evaluation team.
• **Separate section:** Is there a specific section entitled "Findings" in the report? There may be multiple findings sections (one for each question) or one large findings section
• **Linked to Q/Is:** Are the findings linked directly to specific evaluation questions or instructions? Is it clear and evident how these findings will answer one of the questions or instructions?
• **All Q/Is addressed:** Are Findings provided for ALL of the questions or instructions outlined in the report?

• **No Cs or Rs in section:** Is the Findings section devoid of any conclusions or recommendations? This means that there are only findings in this section. If one or two conclusions are found here you can mark it a 'yes,' but if happening regularly mark it 'no.' If there is not a Findings section, then this should be marked 'N/A.' Please be aware of the potential false negative. Marking a "yes" means yes, there were no…. and "no" means no, there were …
• **Raw data analyzed:** This means that data provided is analyzed in some manner and is not in raw form. For example, raw data is baseline = 1 and Yr 1 =3 and Yr 2 = 10; analyzed data would show the change per year or the percentage increase, etc. Also, it should be demonstrated that data was extrapolated from observations, interviews, surveys and document reviews.
• **Data all methods used:** Data from all methods should be provided. Often observations are made, but the knowledge gained through them is never provided in the report. Make sure that somewhere in the report data from all data collection methods is reported.
• **N with %s in text:** Whenever data is provided as a percentage in the text of the report, the "N" or number of pieces of data from which the percentage was created should be clear in the report.
Conclusions (C): A conclusion is a value or judgment made from interpreting findings.
• **Separate from F&R:** Is there a specific section entitled "Conclusions" in the report? There will most likely be multiple conclusions sections (one for each question)
• **Supported by F:** Are all conclusions supported by findings? There should be no conclusions that seem to come out of nowhere. If this happens once or twice and generally Cs are supported by Fs then mark yes. If it is happening regularly it is marked no.
• **Not just Fs restated:** Cs should be more than the restatement of Fs, as they must interpret the findings and add value or judgment to them. If a finding is "50% of respondents liked the program," a C cannot be "we concluded that the program was liked by half of the respondents" and but should instead be "we conclude that there is room for improvement in the program." Please be aware of the potential false negative. Marking a "yes" means yes, there were no…. and "no" means no, there were …
• **No new Fs:** Fs should only be introduced in the Findings section. No new Fs should appear in the C section. Once or twice can be considered acceptable. Please be aware of the potential false negative. Marking a "yes" means yes, there were no…. and "no" means no, there were …
• **Not Rs:** Cs should be the judgment of a F and should not prescribe a course of action to fix a problem. A C is not an R and no Rs should be found in the C section. Please be aware of the potential false negative. Marking a "yes" means yes, there were no…. and "no" means no, there were …
• **Linked to Q/Is:** Are all Cs directly linked to specific questions or instructions outlined earlier in the report?
Recommendations (R): Recommendations are specific courses of action that should be taken by a specific audience and are founded on connected Fs and Cs
• **Separate from F&C:** Is there a specific section dedicated to, and entitled, Recommendations.
• **Supported by F&C:** Are all Rs supported in a clear and specific way to both Fs and Cs? Every R should have at least one C that was interpreted from at least one F. No "blue sky" or unsubstantiated Rs should be included in this section.

• **No new Fs or Cs:** No new Fs or Cs should be introduced in this section, as they have their own section. On occasion an F or C may be restated to clarify the connection to the R. This is acceptable. Please be aware of the potential false negative. Marking a "yes" means yes, there were no…. and "no" means no, there were …
• **Intended Actor Clear:** It should be clear and obvious who is responsible for acting upon this recommendation, be it USAID, the IP, or even the host government
• **Action needed is clear:** It should be clear and obvious what action needs to be taken. This should be detailed to the point that the intended actor knows exactly what needs to be done and in what manner.
• **Linked to Q/Is:** Are all Rs directly answering the specific questions or instructions outlined earlier in the report?
• **Linked to Purpose:** Do all of the Rs come full circle to address the original management purpose of the evaluation?
Lessons Learned
• **Doesn't fit w/ C&R:** Lessons learned should not be relevant to this intervention and therefore do not belong in the C or R sections.
• **For other places/times:** Lessons learned should be useful pieces of information that can be used for future interventions in the region or for similar interventions in other locales.
Executive Summary
• **Is a snapshot of report:** The executive summary should provide all of the relevant information from the report in a several page synopsis.
• **Adds no new info:** No new information should be added to the executive summary that is not provided in the body of the report itself. Everything in the ES should be available in the main text.
Evidence of Change: Does the evaluation prove, through specific empirical evidence that a change occurred in the area of focus of the intervention?
• **Quantitative:** Is this evidence numerically based using statistics and percentages and hard numbers?
• **Qualitative:** Is the evidence qualitative in nature, such as anecdotal evidence, observational evidence?
Evidence of Causality: Does the evaluation provide evidence that the intervention was the cause of the observed change?
• **Comparative:** This approach looks at a comparison group to determine if the intervention may have been the cause of the result.
• **Time Series Break:** This approach looks at the result with the knowledge of where the problem stood prior to the intervention or at the starting point of the intervention. Imagine lowering a speed limit and checking the death rate on that road several years later.
• **Econometric:** This is a statistical approach to determining the relationship between several factors and the end result. Terms such as regression analysis may be used.
• **Forensic (alt. causes):** This approach acknowledges that a change occurred and traces the steps backwards to determine what the cause of the change actually was. Imagine a detective trying to solve a murder by searching for clues.
Comparisons
• **Before and After:** This type of comparison requires baseline data from before the start of the intervention and comparative data at various intervals, (or just after the intervention).

• **Comparison Group:** A comparison group indicates a population that is not being affected by the intervention. This group is decided through a purposive sampling process and is decided based on a number of factors.
• **Control Group (RCT):** A control group is also a population not affected by the intervention; however this sample is created through a randomization process. This is called a Randomized Control Trial.

ANNEX E: EVALUATION REPORTS CONTAINING THE SOWS REVIEWED BY MSI

Austin, Bradley and Ernest Leonardo. 2008. *Assessment of political party programming by USAID in the Islamic Republic of Afghanistan*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM817.pdf.

Baro, Mamadou, and Colette Powers. 2009. *ACDI/VOCA/Uganda : Title II multi-year assistance program -- mid-term evaluation report*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACN710.pdf.

Beyna, Larry S. 2008. *Rapid assessment of USAID/Egypt's administration of justice II project*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACN015.pdf

Boardman, Gerald et al., 2009. *Mid-term evaluation of Afghan e-quality alliance (AeQA) by USAID in the Islamic Republic of Afghanistan*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACN521.pdf.

Boardman, Gerald et al., 2009. *Mid-term evaluation of building education support system for teachers (BESST) by USAID in the Islamic Republic of Afghanistan*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACN522.pdf.

Calavan, Michael, Ashley Barr, and Harry Blair. 2009. *Local administration and reform project: mid-term evaluation*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACO260.pdf.

Capps, Jean and Sandra Wilcox. 2008. *BRIDGE project : final evaluation*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM872.pdf.

Caro, Deborah A. et al., 2008. *Evaluation of the health services program (HSP) in Indonesia: taking stock and looking forward*. Washington DC: US Agency for International Development.

http://pdf.usaid.gov/pdf_docs/PDACM989.pdf.

Chakufyali, Peggy, Jane Chinombwe, and Judith Oki. 2008. *USAID/Zambia assistance to strengthen Ministry of Education support to community schools*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACN014.pdf.

Chopra, Tanja, Susan Pologruto, and Timotio de Deus. 2009. *Fostering justice in Timor-Leste: rule of law program evaluation*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM677.pdf.

Deuster, Paul, Barry MacDonald, and Don Zimmerman. 2009. *Evaluation of USAID/Cambodia's labor and industrial productivity activities*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACO261.pdf.

Dinkin, Donna R. and Robert J. Taylor. 2009. *Evaluation of the leadership, management and sustainability (LMS) project*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACN290.pdf.

Glenn, Russell W. et al., 2009. *Evaluation of USAID's community stabilization program (CSP) in Iraq : effectiveness of the CSP model as a non-lethal tool for counterinsurgency*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACN461.pdf.

Hawkins, Steve, Tsige Gebre-Mariam, and Erwin Lassooy. 2009. *RPM+/SPS (rational pharmaceutical management plus (RPM Plus)/strengthening pharmaceutical systems (SPS) and SCMS (supply chain management system) in Ethiopia : an evaluation*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACO833.pdf.

Henderson, Keith, Charles Jakosa, and Charles Gibson. 2009. *Evaluation of rule of law programs in Liberia*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACO233.pdf.

Herr, Harvey and Rich Mason. 2009. *USAID community stabilization (CSP) counterinsurgency (COIN): special study on business development component*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACN468.pdf.

Hunt, Jenny et al., 2008. *Final evaluation: USAID/Ethiopia high-risk corridors initiative*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM990.pdf.

International Rescue Committee. 2008. *Integrated assistance to IDPs, returnees & other conflict affected populations in the mid western region of Nepal : final evaluation report*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACL968.pdf.

Jansen, William and Victoria James. 2008. *Zimbabwe HIV/AIDS partnership project & behavior change programme : a joint USAID/DFID assessment*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM725.pdf.

Kropsch, Lia and Jennifer Robinson. 2008. *FORTE Saude mid-term evaluation*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM682.pdf.

LaPin, Deirdre, Sam Amadi, and Victor Adetula. 2009. *Mid-term evaluation of the conflict abatement through local mitigation (CALM) project*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PNADP506.pdf.

Leonardo, Ernest, John Wooten, Jr., and Joseph Lombardo. 2008. *Analysis of USAID's capacity development program (CDP)*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM814.pdf.

Levy, Tennyson, Dinesh Agarwal, and Peter Miller. 2009. *Expanding access and demand for DMPA in Uttar Pradesh, Jharkhand, and Uttarakhand*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACN382.pdf

Lewis, Ann von Briesen, Emergy Brusset, and Chandrakishore Jha. 2009. *OTI Nepal program evaluation, 2006-2009*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACN813.pdf.

Lund, Michael and Jennifer Ulman. 2008. *USAID/Phillipines Mindanao programs evaluation: impact on conflict and peace since 2000*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM822.pdf.

Maynard-Tucker, Gisele, Janet Gruber, and Richard Pollard. 2008. *The enhanced HIV/AIDS prevention and improved family health program (EHAP-IFH), Malawi*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PNADO563.pdf.

McDonald, Wayne, Jennifer Karanian, and Edmund Y. Berzilius. 2008. *Evaluation of the USAID/Sudan 'Sudan transitional environment program (STEP)' implemented by International Resources Group (IRG)*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM088.pdf.

Nelson, Sue et al., 2008. *Liberia elections and political processes program evaluation: final report*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM833.pdf.

Neuse, Margaret et al., 2008. *Sudan health transformation project (SHTP) assessment report*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM390.pdf.

Pollard, Richard and John Davies. 2009. *Evaluation of the COMPRI-A social marketing program by USAID in the Islamic Republic of Afghanistan*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM815.pdf.

Pollard, Richard. 2009. *Mid-term evaluation : the health services support project (HSSP) by USAID in the Islamic Republic of Afghanistan*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACN523.pdf.

QED Group. 2009. *Evaluation of the tuberculosis control assistance program (TB CAP)*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACO835.pdf.

Ravenholt, Betty and William Jansen, William. 2008. *Mid-term assessment of the PSP-One project*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM938.pdf.

*Riggs-Perla, Joy et al., 2008. *ACCESS : safe motherhood and newborn care (SMNC) : project assessment*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACL612.pdf.

*Riggs-Perla, Joy et al., 2009. *Takamol mid-term evaluation*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACO254.pdf.

Rudenshiold, Eric and Keith Schulz. 2008. *Evaluation of the Indiana University parliamentary development program [and] assessment of the development and functioning of the Verkhovna Rada [and] recommendations for future assistance*. Washington DC: US Agency for International Development.

http://pdf.usaid.gov/pdf_docs/PDACM769.pdf.

SENAGROSOL-CONSULT. 2008. *DGL felo program impact assessment*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM341.pdf.

Sibley International. 2009. *Mid-term evaluation of task order no. 1: support for accelerated growth and increased competitiveness IQC -- final report*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACO429.pdf.

*Thapa, Mohan M., Kedar P. Badu, and Shanker R. Pandey. 2008. *Final evaluation of strengthening the NGO sector in Nepal (SNSN) project*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACM275.pdf.

USAID. 2008. *Elections and political processes program in Timor-Leste*: evaluation report. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACL911.pdf.

van Haeften, Roberta et al., 2009. *The final evaluation of the FY2002-2008 Bolivia Title II development assistance program*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACO837.pdf.

Weidemann Associates. 2008. *Final evaluation of USAID/Peru poverty reduction and alleviation (PRA) activity: final report*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACN110.pdf.

Whatley, Jennifer and Andrew Rasmussen. 2008. *Two-part evaluation of the organizational viability and the targeted capacity-building components of the center for victims of torture's capacity building of foreign treatment centers program*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACN697.pdf.

Yates, DeeDee et al. 2008. *President's emergency plan for aids relief track 1.0 orphans and other vulnerable children (OVC) program evaluation*. Washington DC: US Agency for International Development. http://pdf.usaid.gov/pdf_docs/PDACO348.pdf.