Meta-Evaluation of USAID Evaluations: 2009-2012

Molly Hageboeck and Micah Frumkin

Management Systems International

November 25, 2013

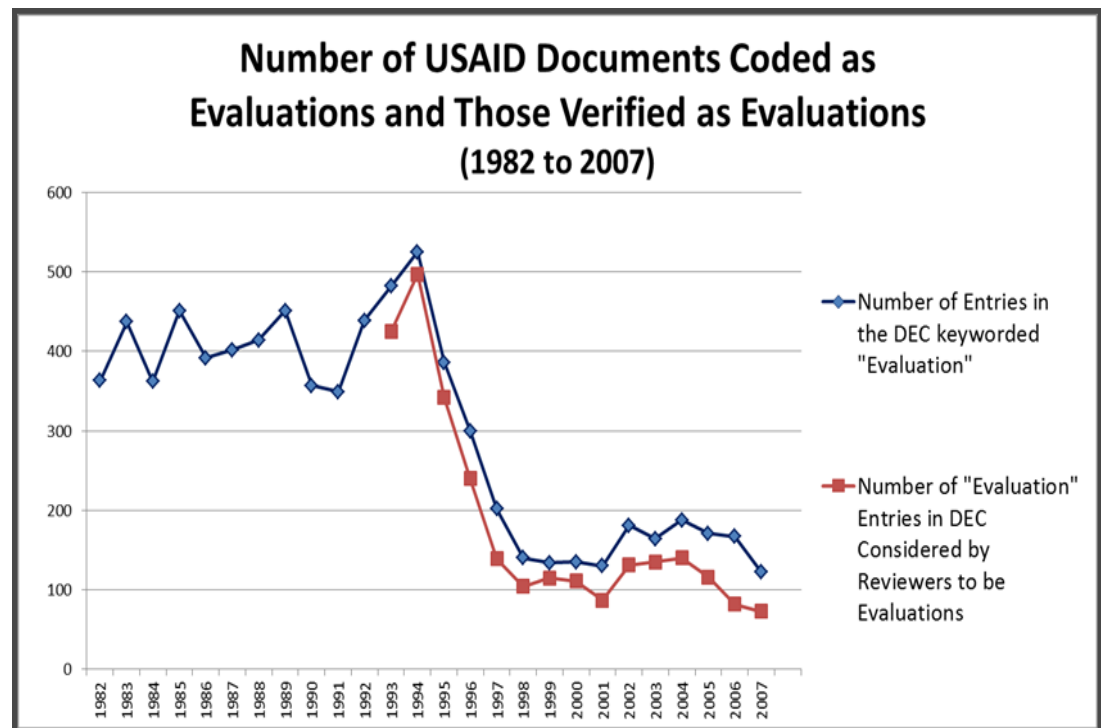# USAID Meta-Evaluation Context

Early Evaluation Leader – 1970 - 1994

- **1970s --** Developed/adopted the Logical Framework to focus designs and evaluations on a project's "theory of change"

- **1980s**
    - -- Meta-evaluations on a regular basis to monitor evaluation report quality
    - -- Early and well received version of an Impact Evaluation series introduced – focused on outcomes: failures as well as successes
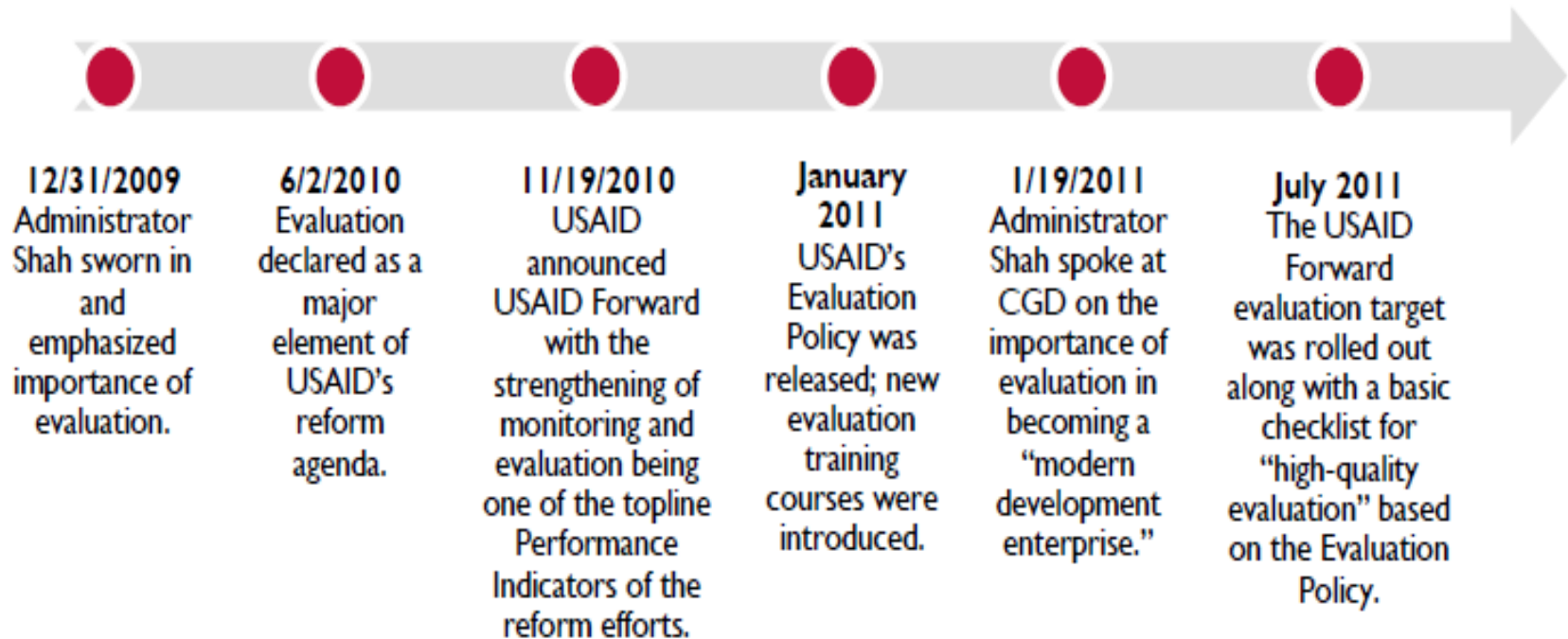
USAID Meta-Evaluation Context

Neglect and Decline – 1995 to 2007/8

- Emphasis on performance monitoring and "Success Stories"

- Evaluation function shifted to State for a few years (now reversed)

Number of USAID Documents Coded as Evaluations and Those Verified as Evaluations (1982 to 2007)

Number of Entries in the DEC keyworded "Evaluation"

Number of "Evaluation" Entries in DEC Considered by Reviewers to be Evaluations

USAID Meta-Evaluation Context

Since 2009/10 – New Emphasis on High Quality Evaluations

12/31/2009	6/2/2010	11/19/2010	January 2011	1/19/2011	July 2011
Administrator Shah sworn in and emphasized importance of evaluation.	Evaluation declared as a major element of USAID's reform agenda.	USAID announced USAID Forward with the strengthening of monitoring and evaluation being one of the topline Performance Indicators of the reform efforts.	USAID's Evaluation Policy was released; new evaluation training courses were introduced.	Administrator Shah spoke at CGD on the importance of evaluation in becoming a "modern development enterprise."	The USAID Forward evaluation target was rolled out along with a basic checklist for "high-quality evaluation" based on the Evaluation Policy.

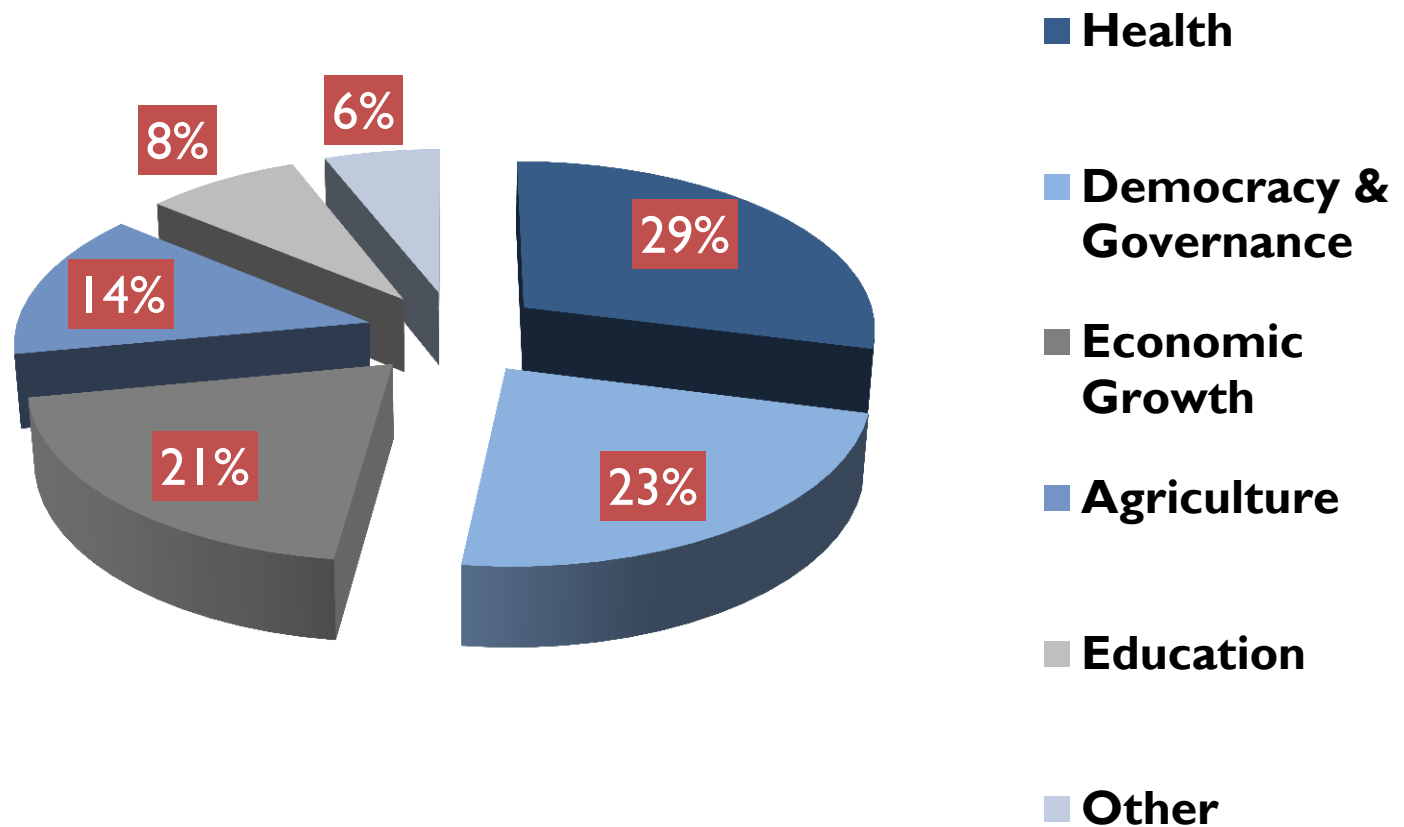**February 2013** – Meta-evaluation (2009-2012) initiated

Meta-Evaluation Sample & Process

- Independent samples for each of four years: 2009-2012

- 85% confidence level, +/- 5%   (budget constrained)

- Total number of evaluations scored: 340

- 37 point checklist used – based on evaluation policy and prior meta-evaluations

- Inter-rater reliability training and spot checks

- 10 of 37 checklist items used to create an overall "score"

- Supplementary group interviews with USAID and contractor staff as well as a small survey of recent evaluation team leaders

Geographic Distribution of 2009–12 Evaluations
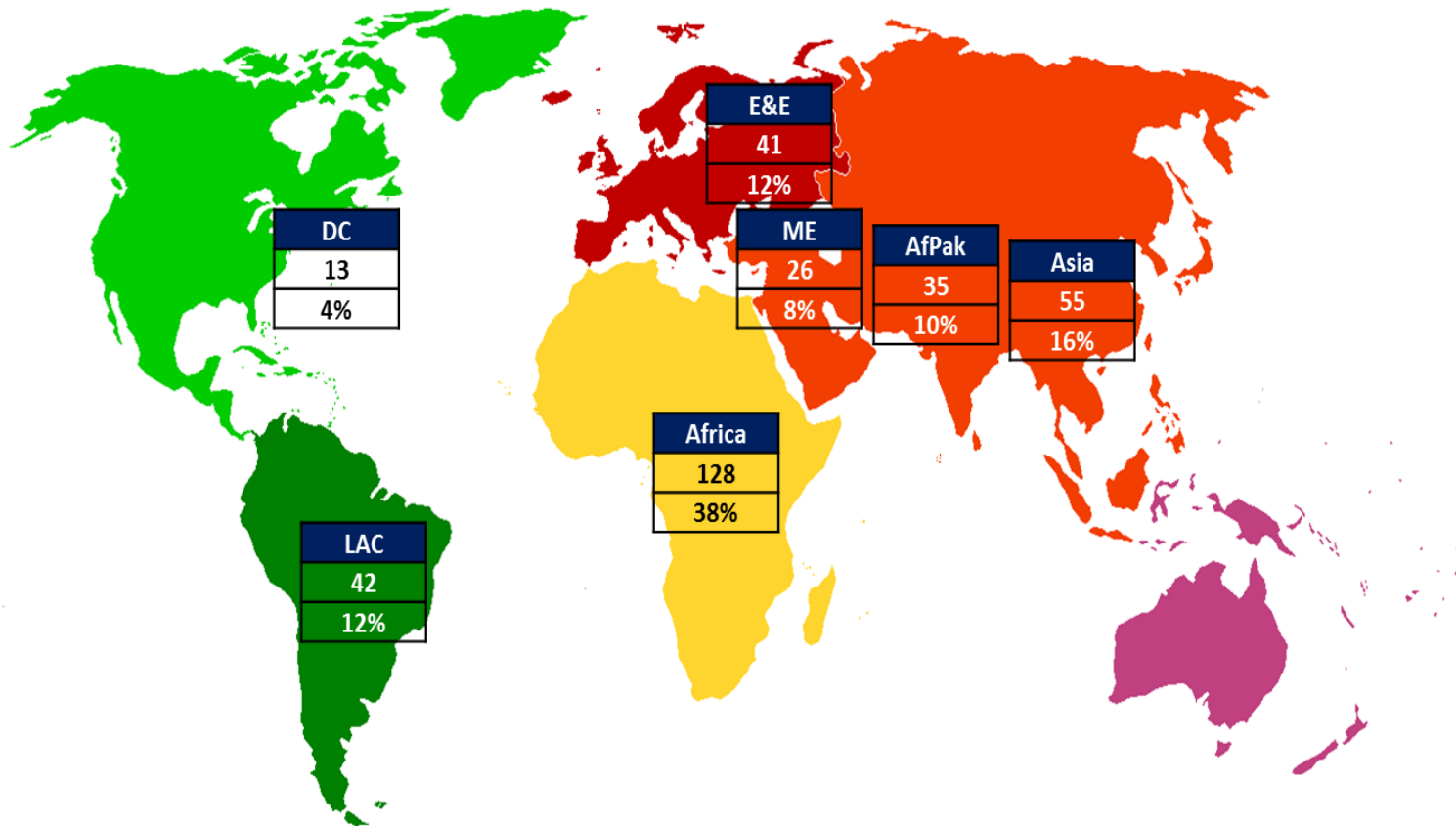
Sector Representation
in the Meta-Evaluation Sample

Legend:
- Health — 29%
- Democracy & Governance — 23%
- Economic Growth — 21%
- Agriculture — 14%
- Education — 8%
- Other — 6%

Percentage of Sample that were USAID Forward Evaluations

20%

80%

■ USAID Forward Evaluation

■ Non-USAID Forward
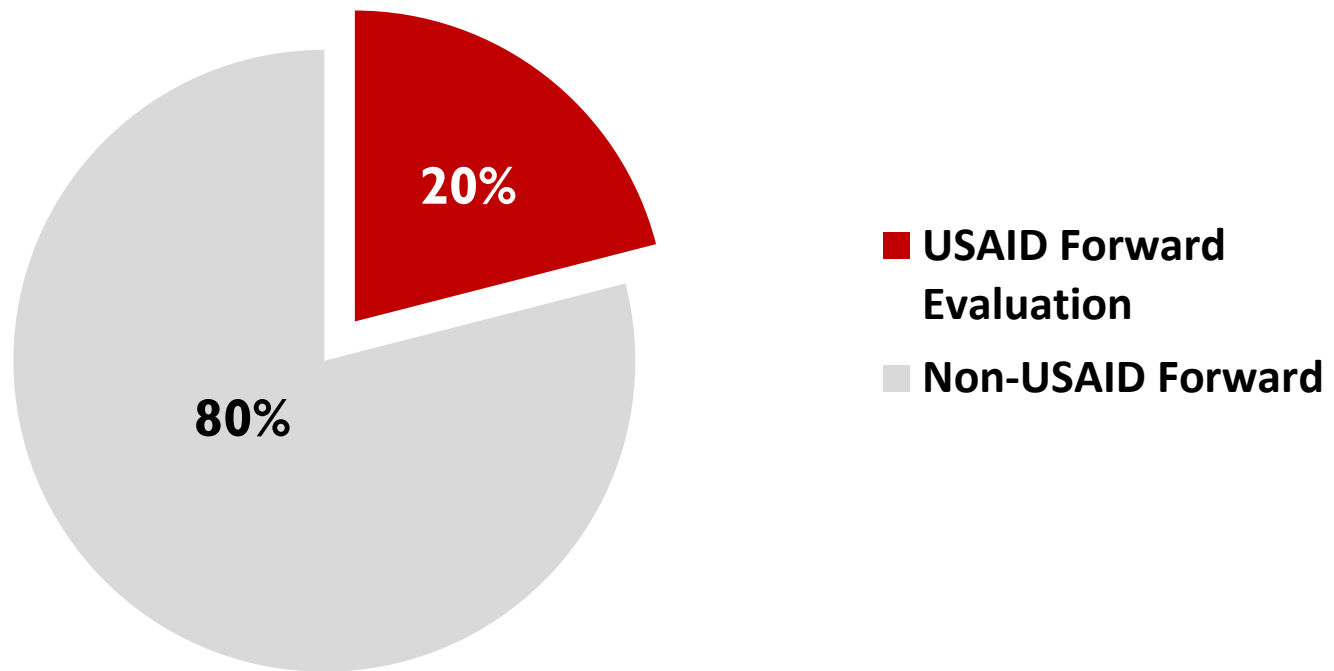
Important Study Limitation

In the 1980s, USAID meta-evaluations looked at cost and evaluation duration when assessing the quality of evaluations.

Somewhere along the way USAID stopped systematically collecting time and cost data on its evaluations – thus these two factors were not examined.

Evaluation Questions

Evaluation Quality

Time

Cost
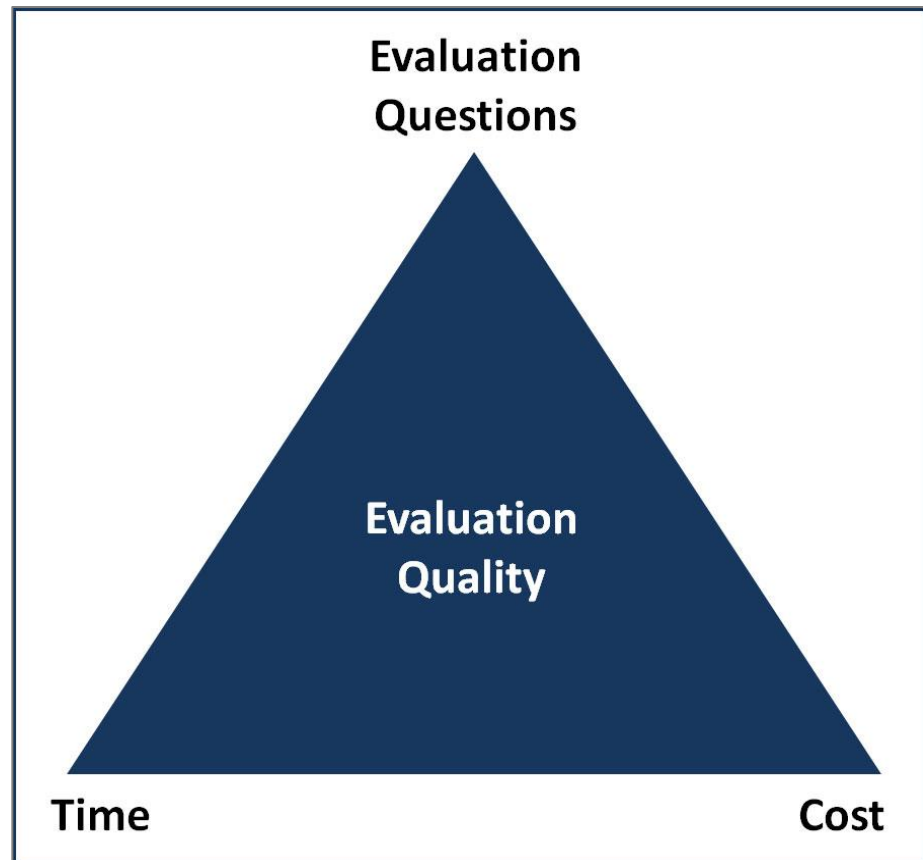
USAID Meta-Evaluation Questions

1. To what degree have quality aspects of USAID's evaluation reports, and underlying practices, changed over time?

2. At this point in time, on which evaluation quality aspects or factors do USAID's evaluation reports excel and where are they falling short?

3. What can be determined about the overall quality of USAID evaluation reports and where do the greatest opportunities for improvement lie?

1. Did USAID's evaluation practice, particularly evaluation quality, change over the study period?

Trends in Number of Documents Submitted to DEC and Number Deemed Evaluations by Reviewers
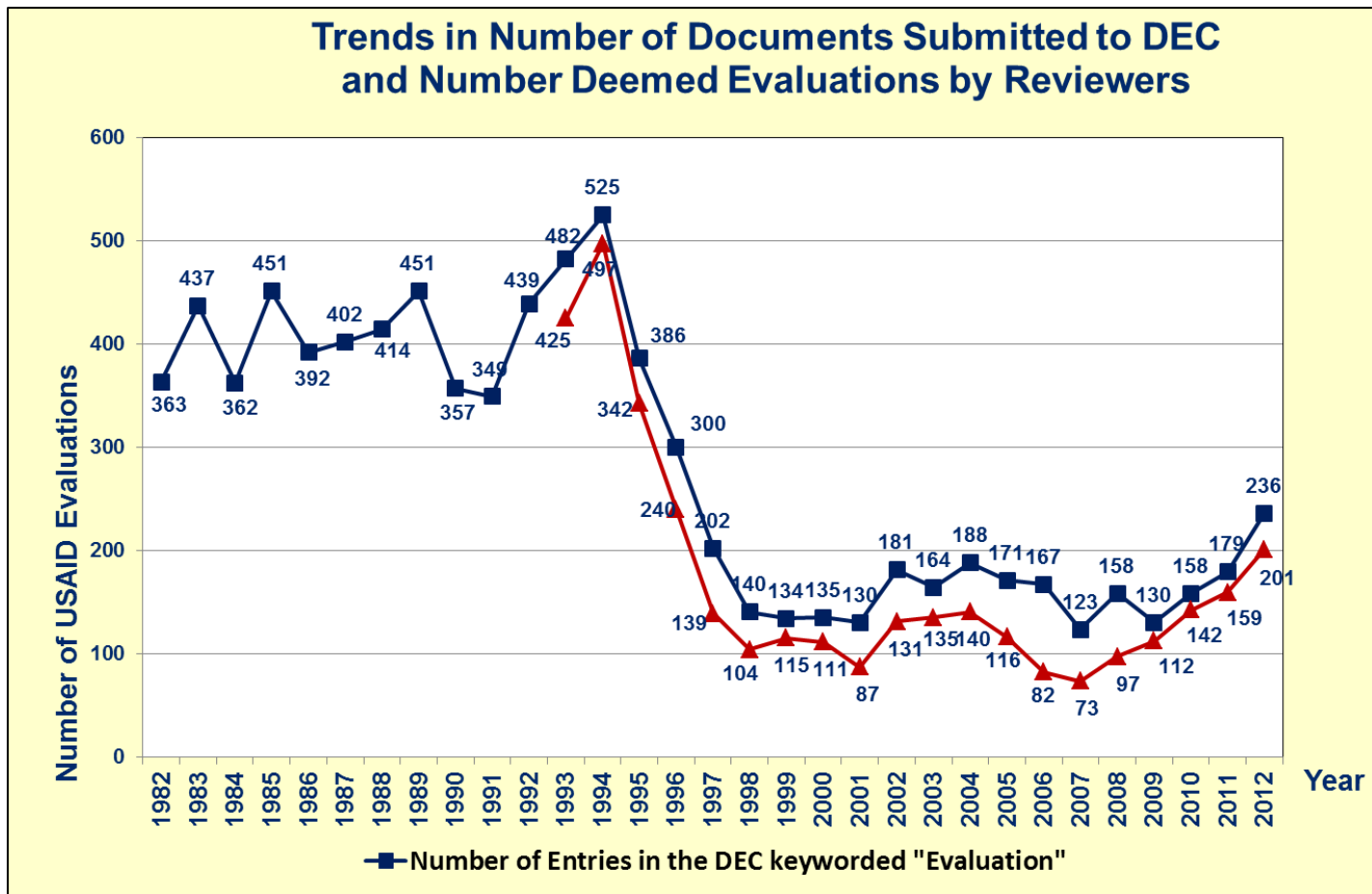
1. Did USAID's evaluation practice, particularly evaluation quality, change over the study period?

Evidence found of changes in quality between 2009 and 2012:

- Net gains on 25 (68 percent) of 37 evaluation quality checklist factors scored

- On 6 quality factors the improvement exceeded 15 percentage points

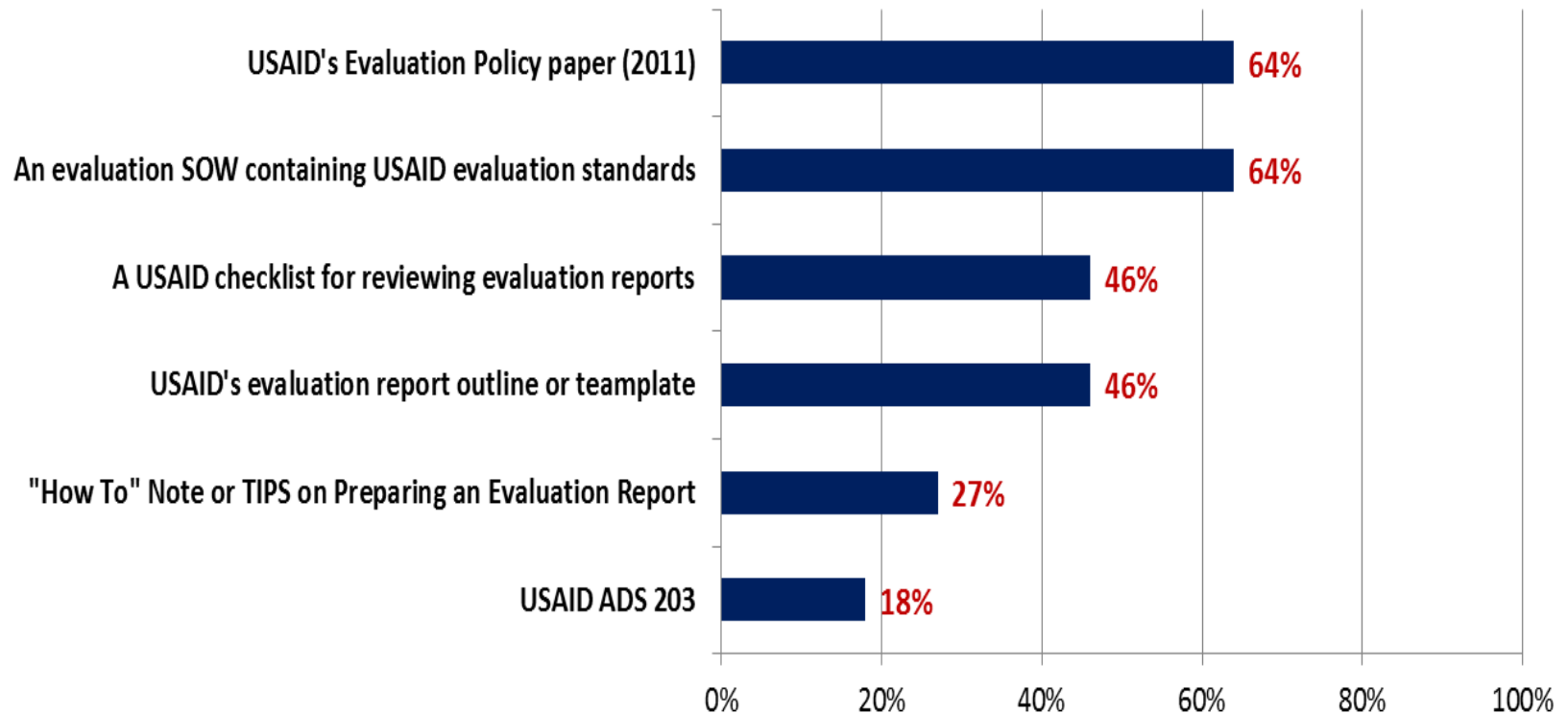| #	Evaluation Report Quality Factors Description	2009–12 Net Change	Percentage Rated Positively in 2012
Net Improvement of More Than 15 Percent on These Quality Factors Between 2009 and 2012			
6	Questions in report same as in SOW	57%	69%
33	SOW is included as a report annex	29%	74%
16	Study limitations were included	26%	64%
35	Annex included data collection instruments	25%	81%
12	External team leader	19%	82%
30	Recommendations—specific about what is to be done	19%	77%

Change over the Meta-Evaluation Period

Some improvements were dramatic and seemed to respond to 2011 Evaluation Policy

Figure __ Percentage of Evaluations that Included a Statement of Study Limitations

2009: 38%
2010: 34%
2011: 61%
2012: 63%
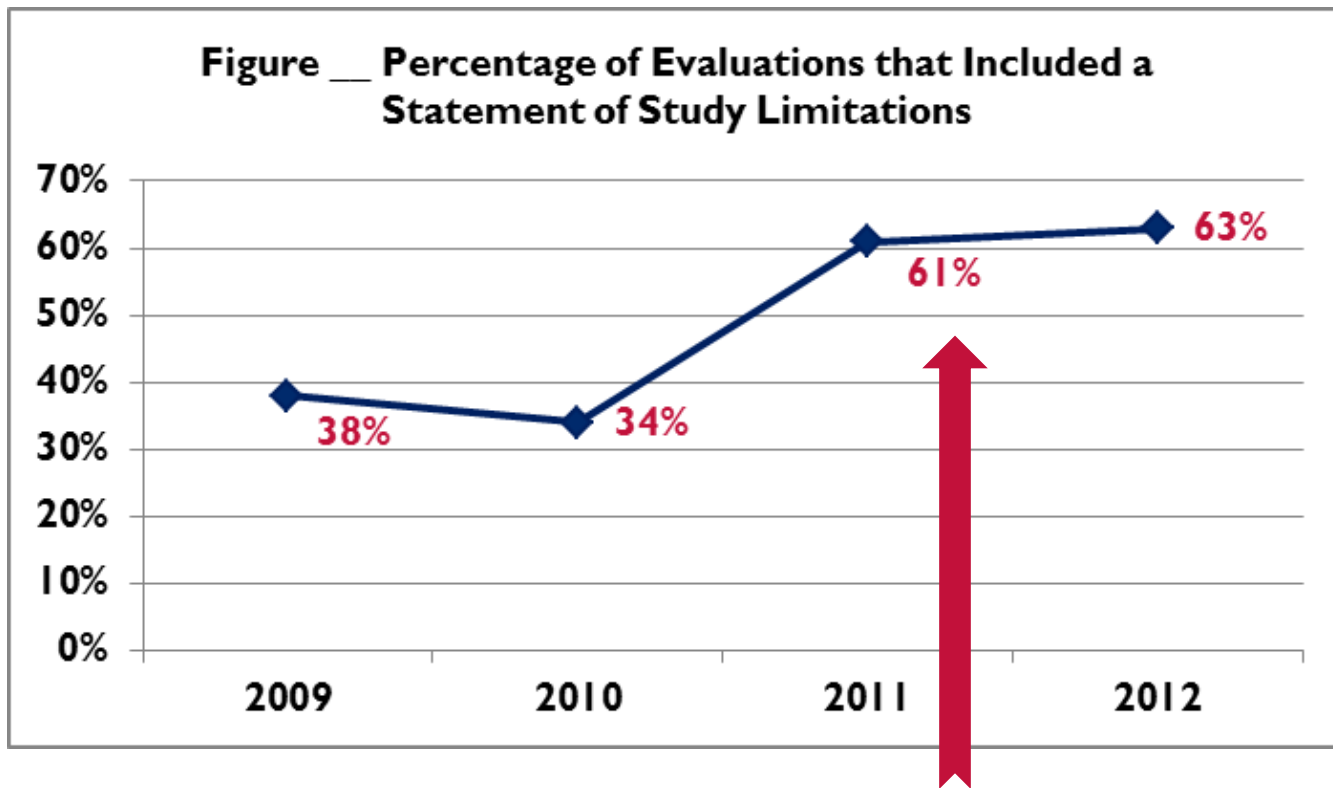
Requirement included in Evaluation Policy in early 2011

Information on USAID Evaluation Standards
Received by Recent USAID Evaluation Team Leaders
(Team Leader Perceptions Survey)
(N = 22)

Standard	Percentage
USAID's Evaluation Policy paper (2011)	64%
An evaluation SOW containing USAID evaluation standards	64%
A USAID checklist for reviewing evaluation reports	46%
USAID's evaluation report outline or teamplate	46%
"How To" Note or TIPS on Preparing an Evaluation Report	27%
USAID ADS 203	18%

2. On which evaluation quality aspects or factors do USAID evaluation reports excel and where are they falling short?

Data on 37 checklist factors plus an extra factor (number of evaluation questions) were sorted by the percentage of evaluations that scored positively on each factor.

- Overall, at least 80% of USAID evaluations met quality expectations for only a few factors (9 out of 38 scored for this question)

Percentage of Evaluations That Met USAID's Quality Criteria in 2012		Evaluation Factors	
Cluster	**Basis for Cluster**	**Number**	**Percentage**
Good	80% of or more met quality criteria	9	24%
Fair	50% to 79% met criteria	11	29%
Marginal	25% to 49% met criteria	6	16%
Weak	Fewer than 25% met criteria	12	32%

80% or more of USAID Evaluations Get it Right on these Nine Factors

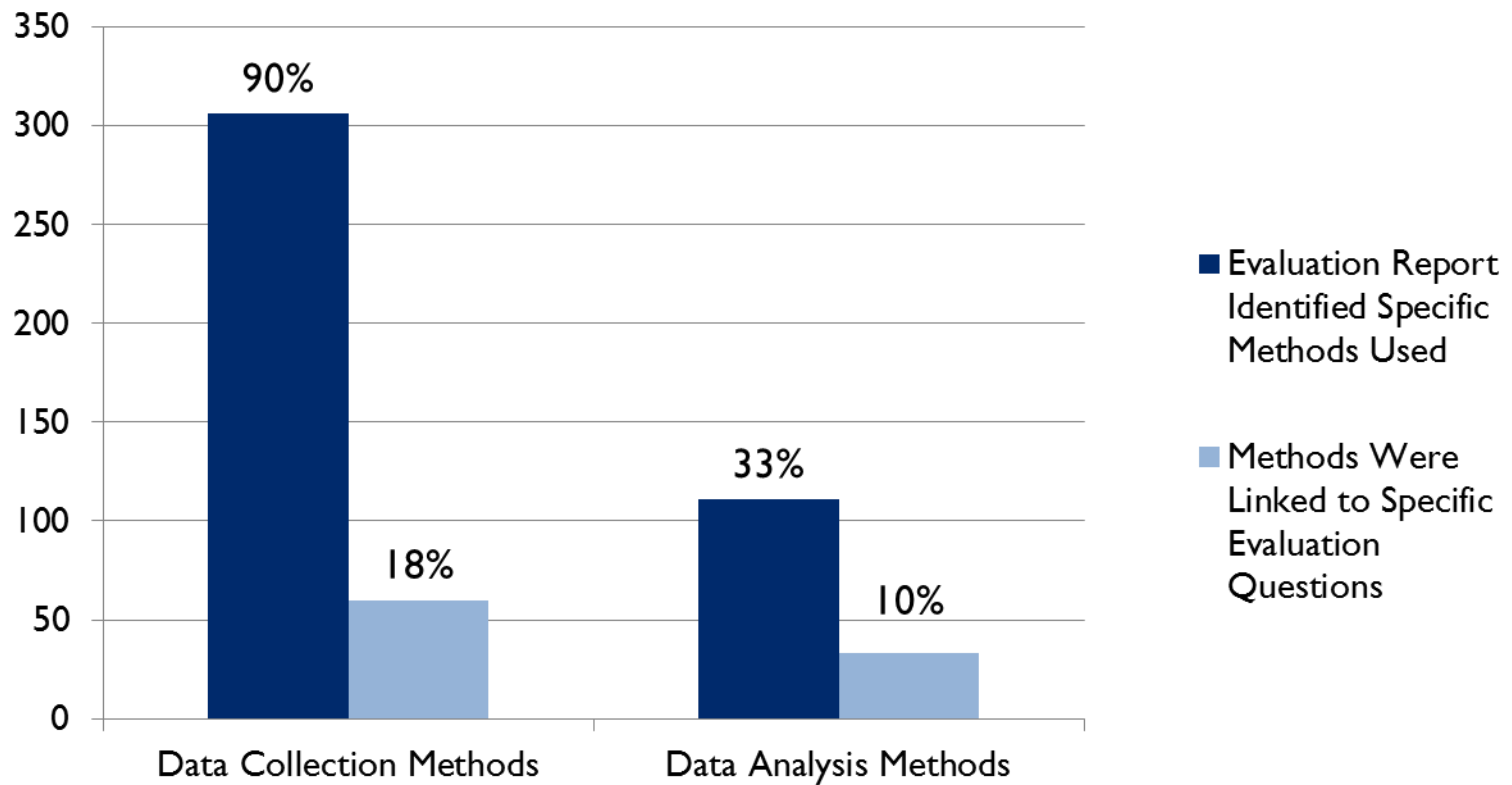| Evaluation Report Quality Factors (Full List)		Rated Positively in 2012	Factors Status in 2012
#	Description		
5	Questions were linked to purpose	98%	Good
8	Data collection methods described	96%	Good
2	Project characteristics described	91%	Good
20	Social science methods (explicitly) were used	84%	Good
34	Annex included list of sources	83%	Good
12	External team leader	82%	Good
4	Management purpose described	81%	Good
35	Annex included data collection instruments	81%	Good
22	Findings supported by data from range of methods	80%	Good

But --- on 29 other quality factors USAID did not reach this level of compliance

Weakest Performance on Rating Factors was Often for the Newest Evaluation Requirements – with Two Important Exceptions
(both of which involve requirements in place since 2008 or earlier)

Evaluation Report Quality Factors (Full List)		Rated Positively in 2012	Factors Status in 2012
#	**Description**		
9	Data collection methods linked to questions	22%	Weak—New
27	Evaluation findings sex disaggregated at all levels	22%	Weak
11	Data analysis methods linked to questions	19%	Weak—New
13	Report said team included an evaluation specialist	19%	Weak
25	Unplanned/unanticipated results were addressed	14%	Weak—May Not Apply
7	Written approval for changes in questions obtained	12%	Weak—New
15	Report indicated conflict-of-interest forms were signed	12%	Weak—New
26	Alternative possible causes were addressed	10%	Weak—May Not Apply
19	Reason provided if some questions were not addressed	9%	Weak—Small N
39	Evaluation SOW includes Evaluation Policy Appendix 1	8%	Weak—New
37	Statements of differences included as an annex	7%	Weak—Small N
38	Report explains how data will transfer to USAID	5%	Weak—New

Presentation of Data Collection & Analysis Methods in Evaluation Reports (N = 340)

3. What can be determined about the overall quality of USAID evaluation reports and where opportunities for improvement lie?

To analyze "overall quality" the study needed an "overall score"

- 10 factors from the 37 point checklist used to create a "score"

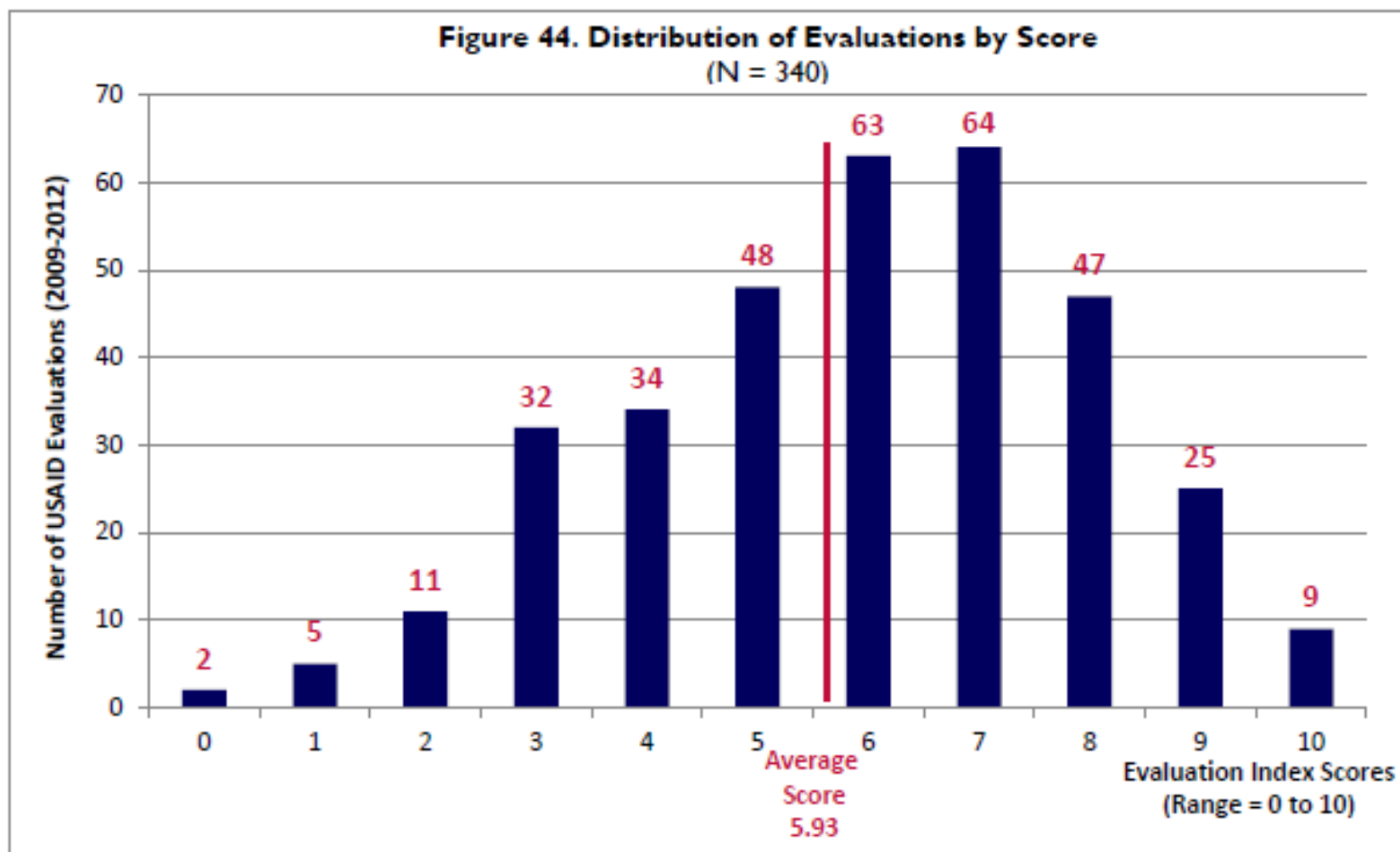- No *a priori* weights were assigned to factors

Average Overall Score: 5.93 out of 10

Historical Comparison: 1983 USAID Meta-Evaluation:
Average Score: 53.8 out of 100

(the only other USAID meta-evaluation that created an overall score)

Frequency of Evaluation Scores Among 2009-2012 Evaluations

Figure 44. Distribution of Evaluations by Score (N = 340)

Important Associations between "Scores" and Key Evaluation Characteristics

➔ **Improvement over Time**

Difference between 2009 (lower) scores and 2012 (higher) scores was statistically significant -- Significant

➔ **Reported presence of an Evaluation Specialist on the evaluation team**

Difference between evaluation scores with and without an Evaluation Specialist was highly statistically significant -- Significant

➔ **Number of Evaluation Questions ---** Not Significant

Meta-Evaluation Recommendations

- Increase the percentage of USAID evaluations that have an evaluation specialist as a fulltime team member with defined responsibilities for ensuring that USAID evaluation report standards are met from roughly 20 percent as of 2012 to 80 percent or more.

- Intervene with appropriate guidance, tools, and self-training materials to dramatically increase the effectiveness of existing USAID evaluation management and quality control processes.

- As a special effort, in collaboration with USAID's Office of Gender Equality and Women's Empowerment, invest in the development of practitioner guidance materials specific to evaluation.

Increase the Percentage of Evaluations that Have an "Evaluation Specialist"

Exhibit 1

	Evaluation Generalist		Evaluation Specialist		
	Novice	**Journeyman**	**Novice**	**Journeyman**	**Master**
Knowledge Dimension	• 40 hour professional evaluation training program, **OR** • Semester undergraduate course involving research design/methods, **OR**	• 40 hour professional evaluation training program, **AND** • Semester undergraduate course involving research design/methods, **OR**	• 80 hour professional evaluation training program, **OR** • Two or more undergraduate or graduate school courses covering research design/methods **AND**	• 80 hour professional evaluation training program, **AND** • Two or more undergraduate or graduate school courses covering research design/methods **AND**	• Two or more undergraduate or graduate school courses covering research design/methods **AND** • Teaches evaluation courses or professional evaluation training programs, **AND**
Practice Dimension	• Full member of one evaluation team involving field data collection, **OR** • Full member of one evaluation design team that produced a design product	• Full member or Team Leader of one or more evaluation team involving field data collection, **OR** • Full member or Team Leader of one or more evaluation design team that produced a design product	• Full member or Team Leader of one evaluation teams involving field data collection, **OR** • Full member or Team Leader of one evaluation design team that produced a design product	• Full member or Team Leader of multiple evaluation teams involving field data collection, **OR** • Full member or Team Leader of multiple evaluation design team that produced a design product	• Team Leader for multiple evaluations, **OR** • Team Leader for multiple evaluation design that produced a product **AND** • Technical quality oversight over a portfolio of evaluations

Increase **USAID** use of Evaluation Management and Quality Control Processes.

Exhibit 2

Evaluation Quality Checkpoints	Timing	What's Different
Evaluation SOW Review	Prior to SOW approval	**Use a SOW Review Checklist**
Evaluation Team's **Document Review** (or Desk Study)	Prior to completion of Detailed Evaluation Design	**Question by Question - what is known & what gaps remain**
Detailed Evaluation Design *(prepared by the team that will actually conduct the evaluation; supersedes proposal stage)*	Prior to approval to start evaluation field work/data collection *(precondition for utilization of LOE allocated for field work)*	**The actual Team that will do the evaluation – with all instruments, sampling plan – oral presentation**
Post Field Work & Analysis **Review of Completeness of Findings, Conclusions, and Recommendations**	Prior to approval for utilization of LOE allocated for writing F-C-R sections of a draft report	**PowerPoint – bullets on as question by question basis – oral presentation**
Review of Draft Evaluation Report & **Approval of Final**	Prior to providing team with feedback on draft and prior to approval of final evaluation report	**Use an Evaluation Report Review Checklist**

Thank You!