

***USAID/Europe and Eurasia/Program Office***

# **Evaluation Handbook**

**Revised June 2004**

# Table of Contents

Preface: Agency Notice on Evaluations from Andrew Natsios

<b>I.</b>	<b>Evaluation Basics .....</b>	<b>4</b>
A.	Background.....	4
B.	Monitoring vs. Evaluation .....	4
C.	When to Conduct an Evaluation.....	5
<b>II.</b>	<b>Evaluation Design .....</b>	<b>7</b>
A.	Evaluation Design Overview.....	7
B.	Decide on the Type of Evaluation.....	7
C.	'Big Picture' Evaluation Questions.....	8
D.	Identify Stakeholders.....	10
E.	Focus the Evaluation Questions.....	11
F.	Develop an Evaluation Planning Matrix.....	12
G.	Data Collection Methods.....	13
<b>III.</b>	<b>The Evaluation Scope of Work .....</b>	<b>20</b>
A.	Outline of a Scope of Work .....	21
B.	The Evaluation Team.....	22
C.	Time and Money .....	23
D.	Evaluation Report.....	23
E.	Utilizing Evaluation Results .....	24

## Appendices

- A. USAID Evaluation Mechanisms
- B. EvalWeb: A PPC-Sponsored Tool for Evaluation Practitioners
- C. Program Evaluations Meta-evaluation Checklist
- D. Guidelines for Conducting Meetings
- E. Using the Flipchart as an Effective Facilitation Tool

## Agency Notice Message:

**ADMINISTRATOR**

USAID/General Notice  
A/AID  
10/20/2003

**SUBJECT: Program Evaluations - Suggestions for Improvement**

Improving the effectiveness of foreign-aid remains my top priority. In order to remain at the forefront of both international development theory and practice, we must continue to learn from field experience. Unfortunately, over the past years the number of evaluations submitted to the Agency's document repository has fallen to only one fourth of prior levels. We need to improve this situation.

ADS 203.3.6 provides Agency guidance on program evaluations. Additional assistance to operating units with the design and production of assessments and evaluations is now available on EvalWeb. This is a one-stop-web-shop on how to conduct evaluations produced by the Policy and Program Coordination Bureau's Center for Development Information and Evaluation. The current URL on USAID's internal website is <http://cdie.usaid.gov/evalweb/>, and it will be accessible from USAID's external website later this month. (Please copy and paste the web address into the internet address bar to download the information.)

Also, regional and pillar bureaus and Missions have the responsibility to produce evaluations that inform decision-making and better tell USAID's story. Evaluations should be commissioned not only to help the specific operating unit but also to generate and share broader lessons learned that can be factored into Agency policies and programs.

I have asked PPC to institute procedures that assure compliance with the above, and also assure that lessons learned and best practices are shared across USAID.

Andrew S. Natsios

Attachment: ADS 203 Guidance on Assessing and Learning

Notice 1044

Attachment 1 [ADS 203 Guidance](#)

## **I. Evaluation Basics<sup>1</sup>**

### **A. Background**

*"An evaluation is a relatively structured, analytical effort undertaken to answer specific program management questions regarding USAID-funded assistance programs or activities." (ADS 200.6.2)*

#### **Evaluations:**

- Systematic analytical efforts that ask why certain results are being achieved.
- Planned and conducted in response to specific management questions about performance of USAID-funded development assistance programs or activities.
- Occasional -- conducted when needed.
- Focused on why results are or are not being achieved.
- Interested in issues such as relevance, effectiveness, efficiency, impact, or sustainability.
- Provide management with lessons and recommendations for adjustments in program strategies or activities.

An evaluation provides a systematic way to assess program performance and impact. Program performance includes:

- Effectiveness -- to what extent were the intermediate results achieved (such as increased agricultural yields by poor farmers);
- Efficiency -- are there ways to achieve the results for less cost or in less time;
- Sustainability -- have the institutions, finances, and personnel requisite for the continued success of the activities been established; and
- Replicability -- to what extent are the circumstances surrounding the intermediate results typical.

An evaluation verifies the continued validity of the development hypothesis or assumptions; identifies impacts (both positive and negative, intended and unintended); gathers lessons learned; and makes recommendations for either the operating unit, or USAID programming and policy writ large.

### **B. Monitoring vs. Evaluation**

#### **Performance Monitoring:**

- Tracks and alerts management as to whether actual results are being achieved as planned.
- Is built around a hierarchy of objectives logically linking USAID activities and resources to intermediate results and strategic objectives through cause-and-effect relationships.

---

<sup>1</sup> Unless otherwise noted, material in the Evaluation Resource Handbook comes from USAID/CDIE EvalWeb at <http://www.dec.org/partners/evalweb/>

For each objective, one or more indicators are selected to measure performance against explicit targets (planned results to be achieved by specific dates). Performance monitoring is an ongoing, routine effort requiring data gathering, analysis, and reporting on results at periodic intervals.

### **Monitoring<sup>2</sup>**

- Focuses on **whether** results are being achieved
- Translated SOs and IRs into measurable performance indicators
- Identifies indicators, baselines and targets
- Routinely collects data on the indicators and compares actual results with targets
- Usually, though not always, quantitative

### **Evaluation**

- Focuses on **why and how** results are, or are not, being not achieved
- Format varies – often qualitative
- A structured analytical tool to answer management questions about: the validity of the hypothesis; unexpected progress; stakeholder needs; sustainability; unintended impacts; and lessons learned.

## **C. When to Conduct an Evaluation**

In a project's lifecycle several events might "trigger" an evaluation. The word trigger should be used cautiously. There is nothing automatic about conducting an evaluation. Use all the following points as tests for deciding to evaluate in your specific situation. Think through your potential evaluation in terms of how its purpose would be classified, whether the management need behind it is associated with positive or negative signs, and whether it is generated by one of the situations below.

- ☑ Performance monitoring indicates **unexpected results** (positive or negative) that need to be explained.
- ☑ A key management decision must be made and there's **inadequate information**.
- ☑ Annual performance reviews have identified **key questions that need to be answered**.
- ☑ Customer or partner feedback suggests that there are implementation **problems or unmet needs**.
- ☑ The contribution of USAID activities to results is questioned.

---

<sup>2</sup> Adapted from "Monitoring and Evaluation –What's the Difference?" by Steve Gale and Annette Binnendijk, USAID.

- ☑ Issues of **sustainability, cost-effectiveness, or relevance** arise.
- ☑ The **validity of results** frameworks hypotheses and **critical assumptions** is questioned.
- ☑ **Recommendations** for actions to improve performance are needed.
- ☑ Extracting **lessons learned** for the benefit of other operating units or for future programming.

## II. Evaluation Design

### A. Evaluation Design Overview

An **evaluation** provides a **systematic way to assess program performance and impact**. Program impact is really looking at the question, “To what extent have the intermediate results lead to achievement of the strategic objective?” Program performance includes:

- effectiveness—to what extent were the intermediate results achieved (such as increased agricultural yields by poor farmers);
- efficiency—are there ways to achieve the results for less cost or in less time;
- impact - what lasting results have been achieved through this activity;
- sustainability—have the institutions, finances, and personnel requisite for the continued success of the activities been established; and
- replicability—to what extent are the circumstances surrounding the intermediate results typical.

An evaluation verifies the continued validity of the development hypothesis or assumptions; identifies impacts (both positive and negative, intended and unintended); gathers lessons learned; and makes recommendations for either the operating unit, or USAID programming and policy writ large.

### B. Decide on the Type of Evaluation

As evaluation is integrated into the management processes of an organization, it is natural that a "short-hand" develops to refer to different types of evaluation. Below is a list of terms that are commonly used; some of these are specific to USAID, while others are used outside of USAID.

- **Impact Evaluation:** An evaluation where the management need being answered is explicitly development impact. The evaluation's purpose is to determine if and why an activity produced the desired result.
- **Midterm/Final Evaluation:** This terminology was common when evaluations were required for every program at USAID. That policy required every program to be evaluated at a mid-point and at program termination.
- **Sectoral/Program Evaluation:** An evaluation which examines various activities with similar goals simultaneously, rather than focusing on one specific program. For example, a sectoral evaluation might examine four different activities to develop small and medium enterprises in one country together, rather than examining them individually.
- **Ex-Post Evaluation:** An evaluation which seeks to learn lessons after a program has terminated to understand long-term impacts of the activity. This type of evaluation satisfies management's need for knowledge.

- **Formative Evaluations** are conducted in the middle of a program. Their purpose is to determine lessons from the project thus far and determine how to refine the program going forward.
- **Summative Evaluation:** These evaluations are the "autopsy" of a program after it has concluded. They seek to find out "what happened" and draw conclusions for future projects.

It is important to clarify at the very beginning whether the primary purpose of the evaluation is to make an overall judgment about the effectiveness of a program, or to make recommendations for improving or refining the program.

### C. 'Big Picture' Evaluation Questions<sup>3</sup>

The following questions are illustrative and should be used as a "model" not as a "blue print". They should be adapted to the characteristics and features of each evaluation. Not all of the evaluation concerns have to be examined in every evaluation. The final choice will depend on the purpose of each evaluation. However, it should be clear that the larger the number of concerns, the more comprehensive (or less partial) the assessment of the project impact will be. If one or more of these concerns are assessed in isolation from the other ones, the evaluation results can be biased or misleading (or both).

#### Validity of Design

- Are the objectives clearly stated, describing the solutions to the identified problems and needs?
- Are the indicators of achievement clearly defined, describing the changes to be brought about?
- Have the external factors affecting project implementation been identified and assumptions proven valid?
- Is the project document logical and coherent linking the inputs, activities and outputs to each immediate objective?
- Are the roles and commitment of the various partners clearly defined?
- Is there any reason to revise it?

**Delivery Process:** Has the overall execution of the project focused on the achievement of the objectives?

- Have the various partners contributed to project implementation as planned?
- Have the main partners interacted and coordinated as planned?

**Relevance:** Examines the usefulness of the project's results in solving the identified problems and meeting the needs of the target group(s). The analysis ascertains whether the project continues to make sense and identifies any changes that may have occurred in its context during implementation. The initial problems and needs may no longer exist. New problems and needs may have emerged as a result of

---

<sup>3</sup> Source: <http://www.ilo.org/public/english/bureau/program/guides/indpen/annex4.htm>

political, economic, social and other factors, or even because of the project's activities. Ultimately, the analysis determines whether the objectives are still valid or should be reformulated.

- Do the problems/needs that gave rise to the project still exist, have they changed or are there new needs that should be addressed?
- Was the project an appropriate response to the problems/needs that existed when it started?
- Is it still appropriate to the problems/needs?
- Have the priorities given to the basic components of the project, i.e. institutional development versus direct support, changed? If so, why?
- Are the objectives still valid or should they be reformulated?

**Effectiveness:** Assesses the extent to which the project has achieved its objectives and reached its target group(s). The analysis determines if the expected results have been or will be accomplished and, if not, whether the statement of objectives should be modified (mid-term evaluation) or the project extended (final evaluation). This assessment is facilitated if the objectives and indicators of achievement adequately describe the desired situation at the end of the project. The changes brought about by the project are ascertained by comparing information regarding the pre-project situation (of the beneficiaries) with the existing situation at the time of the evaluation.

- To what extent has the project achieved its immediate objective(s)?
- Have data been collected by the project on the indicators of achievement? Do they provide adequate evidence regarding the effects and impact of the project? Is it necessary to collect additional data?
- Did the target group participate in the formulation and implementation of the project?
- Have the benefits of the project accrued to the target group?

**Efficiency:** Assesses the results obtained in relation to the expenditure incurred and resources used by the project during a given period of time. The analysis focuses on the relationship between the project's inputs, including personnel, consultants, travel, training, equipment and miscellaneous costs, and the quality and quantity of the outputs produced and delivered. It ascertains whether there was adequate justification for the expenditure incurred and examines whether the resources were spent as economically as possible. This assessment should also determine whether the actions by the various partners were complementary and identify alternative strategies to deliver more and better outputs with the available inputs.

- Do the expected project's results continue to justify the costs incurred?
- Have the resources been spent as economically as possible?
- Were the actions of the various partners complementary?
- Are there more efficient ways and means of delivering more and better outputs with the available inputs?

**Sustainability:** Ascertain the extent to which the project's results have had or are likely to have lasting effects after the termination of the project and the withdrawal of external resources. The factors affecting sustainability are examined on the basis of

the priority assigned to the project by the direct recipients and/or intended beneficiaries. Their readiness to continue supporting or carrying out specific activities, or even replicate the activities in other regions or sectors of the country, is particularly relevant. The analysis also assesses the availability of local management, financial and human resources that would be needed to maintain the project's results in the long run.

- What is the likelihood that the project's benefits will be sustained after the withdrawal of external support?
- Do conditions exist to ensure that the project's results will have lasting effects?

**Causality:** Examines the factors or events that have affected the project's results. If the inputs needed to carry out the planned activities and deliver the expected outputs were available on time, the project's implementation and performance would be successful. If, on the other hand, there were significant deviations from the planned schedules, the analysis would determine the reasons for such changes. The assessment should also analyze the effect of other factors such as technical, administrative or managerial constraints, inadequate inputs, failed commitment by project partners, insufficient funds, a faulty assumption or the effect of an unexpected external factor.

- What particular factors or events have affected the project's results?
- Were these factors internal or external to the project?

**Unanticipated Effects:** Identified during the evaluation to ascertain if a project is having any significant unforeseen positive or negative effects. Once identified, appropriate action can be taken to enhance or mitigate them for a greater overall impact.

- Is the project having any significant (positive and/or negative) unforeseen effects?
- What could be done to either enhance or mitigate them so that the project has a greater overall impact?

**Alternative Strategies:** Evaluations examine whether alternative approaches might have had greater impact or might have been more cost-effective. This analysis is especially valuable when follow-up activities are being planned.

- Is there, or would there have been, a more effective way of addressing the problem(s) and satisfying the needs in order to achieve the objective(s)?
- Is the project's strategy still valid or should it be reformulated?

#### **D. Identify Stakeholders -- whose questions will be addressed in the evaluation?**

All evaluations have multiple stakeholders. A stakeholder is defined as any person or group who has an interest in the project being evaluated or in the results of the evaluation. Stakeholders include funders, project staff and administrators, project

participants or customers, community leaders, collaborating agencies, and others with a direct, or even indirect, interest in program effectiveness.

It is important to remember that evaluators (whether internal or external) are stakeholders, and not neutral third parties, as is often assumed. Evaluators have a vested interest in what they are doing and care about doing it well.

To ensure that you have gathered multiple perspectives about the salient issues, involve as many stakeholders as possible in initial evaluation discussions. Otherwise, the evaluation is likely to be designed based on the needs and interests of only a few stakeholders--usually the ones with the most power--and may miss other important questions and issues of stakeholders who are not included at the table. Of course, involving every stakeholder may not be realistic. However, try to consult with representatives from as many stakeholder groups as possible when designing or redesigning the evaluation plan, and provide them with timely results and feedback.

Involving many stakeholders will help ensure that the evaluation process goes more smoothly: more people are invested and willing to work hard to get the necessary information; project staff concerns about evaluation are reduced; the information gathered is more reliable and comes from different perspectives, thus forcing the team to think through the meaning of contradictory information; and the recommendations are likely to be accepted by a broader constituency and implemented more fully and with less resistance.

### **E. Focus the Evaluation Questions**

Good questions, first and foremost, are the ones that will meet your evaluation's purpose. Good questions should be:

- Clear
- Answerable
- Fair
- Bounded
- Based on existing knowledge

This checklist can help you see if your questions meet these criteria. Do the questions specify:

- What is being studied (e.g., the program)?
- What is the timeframe being covered?
- What is the location you're concerned about?
- What is the unit of analysis (e.g., program participants, NGOs, etc.)
- Are all terms defined and will both stakeholders and external readers understand them?
- Are the criteria for normative terms like timely, adequate, satisfactory, reasonable, etc., defined specifically? Are these definitions based on appropriate sources, like regulations or statutes, or international definitions?
- Are the questions objective and fair?
- Are questions neutrally stated or do they inappropriately imply an answer?

- Is the scope of questions too broad or too narrow?
- Can the questions be answered within the evaluation's timeframe and budget?
- If making a choice between questions, have you chosen those with the biggest impact, for example on budget or program improvement?
- Do the questions reflect previous research or management practice?
- Do the questions build on existing knowledge?

Once you have listed the evaluation questions, weed out the ones that are either too broad or too narrow and prioritize those that are left. To improve the list of questions, try clustering them according to whether or not they are:

- Impact questions
- Design questions
- Cost questions

In general, it is important to remember that:

- There is no correct evaluation design.
- It all depends on the purpose of your evaluation and the questions it must answer.
- Complexity for its own sake is not desirable.
- Keep the design as simple as possible.

#### **F. Develop an Evaluation Planning Matrix<sup>4</sup>**

Although by all appearances, the matrix is a very simple tool, it has a powerful purpose. It helps you to consider a range of questions, stakeholders, and the data collection methods that will be used to elucidate the answers. Evaluators sometimes get into the habit of using one or other data collection method, e.g., an end-of-training questionnaire, without considering the advantages of alternative methods. This tool prompts you to consider each evaluation question and to decide which of the many data collection methodologies have the greatest potential for providing the desired information.

<b>Questions (examples)</b>	<b>Source of Data</b>	<b>Methodology</b>	<b>Analysis Plan</b>
1. How do the Ministries of Health and Education cooperate on the HIV/AIDS program?	Ministry officials	Key informant interviews	Synthesis

<sup>4</sup> <http://www.ilo.org/public/english/bureau/program/guides/indpen/annex4.htm>

2. What is the impact of the HIV/AIDS education program?	Beneficiaries	Sample survey and observation	Statistical program
3. What are the constraints to achieving impact?	Implementation team	Interviews and document review	Pattern analysis
4. Have the trainees learned the course material?	Trainers	Interviews and course material; before/after student testing	Frequency distribution

### G. Data Collection Methods<sup>5</sup>

There are many data collections methods to choose from – the key is to choose the methods that work best for the intended users of the evaluation report. Be aware that there are pros and cons to each method – some are more costly and more reliable, while others are cheaper but less reliable. Recognizing the trade-offs is key to identifying which data collection methods to use. Most data collection strategies draw on a mix of methods – document review, interviews, and observations.

Before identifying the methods, it is important to think about what the unit of analysis is. The unit of analysis is the major entity that you are analyzing in your study. For example, the unit of analysis can be:

- Individuals
- Groups
- Geographic units (communities, towns, states)
- Social interactions (arrests, divorces)

Evaluation methods and procedures are the processes you use in gathering information and analyzing it to answer your evaluation questions. The information we are interested in is known as “empirical evidence,” that is, it is based on observation and it is *out there*, wherever our evaluation will occur. Remember that an evaluation is not unlike a mystery investigation and like any good detective, you need to go to the scene of the crime, talk to witnesses, and gather evidence through observation. Good detectives may rely on hunches, too, but ultimately they need to gather empirical evidence to solve the mystery.

The second point is that the methods and procedures that are right for your evaluation depend on the questions you are answering. Selecting the appropriate methods and procedures for your evaluation questions can be a complex task. Therefore, this section will provide you with some issues to consider in making that decision, guidance on seeking the help of experts in choosing methods, and common terms that are used in discussing evaluation methods.

---

<sup>5</sup> USAID’s EvalWeb

## Major Methods

Data collection methods can be broken into two basic types: data collection methods and data analysis methods.

**Data collection** is where you go out and get the data or information that will answer your evaluation questions. You need to have a pretty good idea of what you will do with the data, so that you don't collect the wrong information.

**Data analysis** is the process of drawing conclusions from a stack of completed questionnaires or focus group flip charts or a spreadsheet full of numbers. That information needs to be examined and analyzed in a way that will persuasively arrive at a conclusion or answer to the evaluation question. Again, what you can do in the data analysis stage largely depends on the information you have collected. For example, can you say anything of value about one subgroup of your population through your survey data? It all depends on whether you consciously collected a sample large enough to be representative of that part of the population (see "stratified sampling" below).

Therefore, you should develop your methods for data collection and analysis hand-in-hand, so that you have the right information to perform the analysis you have planned. You may see unexpected patterns once the data are collected, which suggest other analysis you would like to perform. That's a positive development and commonly happens. But it is important to anticipate as many needs as you can during the evaluation design, in order to give yourself the flexibility to pursue these potential analyses.

A second division we can make in methods is between **quantitative** and **qualitative** methods.

- Quantitative methods refer to anything that can be summarized by numbers.
- Qualitative methods rely on language and explanation.

You must choose whether quantitative methods, or qualitative methods, or both, best meet your goals.

The matrix below represents these two divisions in methods. In each quadrant some major methods are listed. Below we will explore each of these methods, some of the major issues associated with their use, and examples of the sorts of evaluations they would be appropriate for.

## Some Major Evaluation Methods Organized By Type

	Data Collection	Data Analysis
<b>Quantitative</b>	<ul style="list-style-type: none"> <li>• Questionnaires</li> <li>• Censuses</li> <li>• Probability samples</li> <li>• Physical measurement</li> </ul> <div style="text-align: right; border: 1px solid black; width: 30px; height: 30px; margin-left: auto; display: flex; align-items: center; justify-content: center;">I</div>	<ul style="list-style-type: none"> <li>• Descriptive statistics</li> <li>• Hypothesis testing</li> <li>• Indicators</li> <li>• Correlation/regression</li> </ul> <div style="text-align: right; border: 1px solid black; width: 30px; height: 30px; margin-left: auto; display: flex; align-items: center; justify-content: center;">III</div>
<b>Qualitative</b>	<ul style="list-style-type: none"> <li>• Questionnaires</li> <li>• Semi-structured interviews</li> <li>• Focus groups</li> <li>• Key informants</li> <li>• Desk reviews</li> <li>• Literature searches</li> </ul> <div style="text-align: right; border: 1px solid black; width: 30px; height: 30px; margin-left: auto; display: flex; align-items: center; justify-content: center;">II</div>	<ul style="list-style-type: none"> <li>• Content analysis</li> <li>• Syntheses of existing documents</li> <li>• Comment coding</li> <li>• Case Studies</li> </ul> <div style="text-align: right; border: 1px solid black; width: 30px; height: 30px; margin-left: auto; display: flex; align-items: center; justify-content: center;">IV</div>

### Quadrant I: Quantitative Data Collection Methods

**Questionnaires** or surveys, as they are also known, are familiar to all of us from daily life. Surveys may ask us to respond with information that is either quantitative or qualitative in nature. For instance, if a survey asks the number of children we have, we can respond with a number. The analyst can use those responses and average them or find its correlation with another response or perform other quantitative analyses.

Surveys often get at qualitative information, which is then analyzed quantitatively. Some surveys provide a series of pre-defined answers, where the responses can then be summarized quantitatively. Another common format is to present a statement and then ask the respondent to mark whether they “strongly agree”, “agree”, “disagree”, or “strongly disagree”, also known as a **Likert scale**.

Finally, surveys may employ open-ended questions. Responses are written out on the survey and often must be coded later during analysis.

The questionnaire itself is often known as the **survey instrument**. While surveys often look fairly simple and straightforward, much thought should go into designing the instrument so that the wording will be clear, the order of the questions is effective, the length is not burdensome, and, most importantly, the survey gathers data on the questions of interest to the study. Often, to make sure that the survey design is satisfactory before spending the money to implement it, a survey will be **pre-tested**.

Often we administer a questionnaire to only a sample of a population. This is because, if done right, we can get an excellent idea of the characteristics of a much larger group by just looking at a sample. Looking at a small number can have a great advantage in terms of time and cost, especially if we are administering the survey by phone or in person. Where we allow respondents to choose to fill out a survey, such as with surveys on the Internet or at the front counter of a business, we run into a problem with selection bias. That is, the people who filled out the survey are not representative of the whole population of interest. This same problem can affect surveys that are mailed to respondents, known in this case as, nonresponse bias.

If you intend to use a survey, read the article entitled, “20 Questions a Journalist Should Ask About Poll Results” by the [National Council on Public Polls](#). Because surveys are so common, it is important to make sure that you conduct your survey using standard social scientific practices, otherwise you will invalidate your results:

*“The major distinguishing difference between scientific and unscientific polls is who picks the respondents for the survey. In a scientific poll, the pollster identifies and seeks out the people to be interviewed. In an unscientific poll, the respondents usually “volunteer” their opinions, selecting themselves for the poll.”*  
-- 20 Questions

**Censuses:** A census examines every member of a population to come to conclusions, rather than just looking at a sample. The U.S. national census is a questionnaire to collect data about every resident of the United States.

**Probability samples** use probability theory to choose a group which is statistically **representative** of the population as a whole. Probability samples are representative because they choose members of the sample randomly. Such a technique can make surprisingly conclusive statements based on a relatively small sample. Basic requirements to implement a probability sample are to have a list of all members of the population—which sounds like an easy requirement but can often be quite challenging. This list is known as the **sampling frame**. The other requirement is that each member of the population has an equal chance of being selected.

There are several types of probability sampling. One that is often used is **stratified sampling**, which divides the total population into groups or strata. This technique allows us to make more conclusive statements about certain subgroups of the population than we could with a simple random sample. Other sampling techniques are called **nonprobability sampling**, which are used when we cannot meet the requirements of probability sampling or we only need an informal estimate.

**Physical measurement** is often used in health or environmental studies. The instruments in these cases may be scales to weigh babies, or calipers to measure trunk widths, or counts of agricultural pests.

## **Quadrant II: Qualitative Data Collection Methods**

**Semi-Structured Interviews:** Similar to surveys administered face-to-face, semi-structured interviews are an opportunity to ask a series of questions. This technique is used when the data you seek is more qualitative and would not be captured in a set of pre-defined responses. It is also a preferable technique when dealing with high-level interviewees who may feel put-off by answering a survey. Such interviews are not simply conversations; rather they use as their instrument an **interview protocol**, or list of questions which you ask through the course of the interview.

**Focus groups** bring a group of people together to answer open-ended questions. They are similar to semi-structured interviews, in that they are best suited to gather qualitative data, which would be difficult to gather through a survey. They differ from semi-structured interviews in that the interaction of participants with each other can elicit a different response than speaking one-on-one in an interview format. Similar to

semi-structured interviews, focus groups use an **interview guide**, for the facilitator to guide the group's discussion.

**Key informants** allow a researcher to get “inside” a community, especially when the nature of what is being studied would make it difficult to use other methods. If, for instance, we are researching whether or how program beneficiaries sell the condoms they receive through the condom distribution program, we might need a key informant to explain to us how the actual transactions take place, who buys them, how much they sell for, etc.

**Desk review** analyzes previously collected information, often information collected in the course of program operations. We then check these files for program compliance or other patterns and relationships. If for instance, we are interested in knowing if program administrators are following guidelines, we may analyze a sample of program files and draw conclusions from what we see there.

**Literature searches** use libraries and electronic databases to find the extent of previous research on a subject. If we are interested in knowing potential unintended effects of a condom distribution program in a developing country, we may perform a literature review to find documented problems with other similar projects.

### **Quadrant III: Quantitative Data Analysis Methods**

**Descriptive statistics** are the most common form of statistics in our daily lives. Descriptive statistics are used to understand better a given set of data, or distribution. Measures of central tendency are commonly used and include the:

- **Mean:** what we normally call the average, calculated by summing the data and dividing by the number of data element
- **Median** is the middle piece of data
- **Mode** the piece of data which shows up the most

We also generally want to know how spread out or packed in the data are with respect to the mean. **Standard deviation** is simply the average difference between the data and the mean. We might also be interested in the **range**, the difference between the lowest and highest number in the data.

**Hypothesis testing:** We mentioned hypothesis testing above as being important in deductive studies. Hypothesis testing is used to test an interpretation of data when we have gathered that information through a sample. We want to know whether it is possible that what we think we see in the data could have occurred just through random chance. There are a variety of hypothesis tests we can use, the most common of which are **Z tests** and **T tests** and **Chi square**. When we successfully test our hypothesis, we can say that our findings are statistically significant at a certain confidence level.

**Indicators** include numeric values, percentages, scores, and indices. At USAID, they are commonly used to observe progress and to measure actual results compared to expected results. In evaluations, indicators are also often used to facilitate

comparisons between whatever our **units of analysis** are—countries, people, NGOs, municipalities, etc.

**Correlation/regression** is used to find associations between certain characteristics and certain outcomes. Knowing the strengths of these relationships can help predict what future results will be, based on known characteristics. These techniques are powerful, but complex to apply and often need lots of detailed historical data.

#### **Quadrant IV: Qualitative Data Analysis Methods**

**Content Analysis** is a powerful technique to deal with masses of qualitative, usually written, information. This technique might be used to analyze the portrayal of a new governmental program in the country’s national media. It is many times better than a “general impression” when making assertions about the content of writing. In a nutshell, the technique relies on counts of words, ideas, or assertions in a sample of written material. Computerized programs like NUD\*IST are often commonly employed to help track where such references occur and how often.

**Syntheses** of existing documents: This technique may rely on a literature search to gather relevant data or it may simply involve the analysis of an already known group of documents. Its advantage is that it leverages already collected information for use in the evaluation.

**Comment coding** may be part of the analysis of data gathered through many techniques, such as surveys, focus groups, semi-structured interviews, or participant workshops, or it may be part of a content analysis. The basic concept is to create a manageable number of categories of comments that sound similar. Then you go through the written data marking each with a code representing one of the categories. This coding process is often done on a spreadsheet. Such coding facilitates some quantification, helping you to get a handle on what was said. For instance, if 40% of the comments dealt with some form of dissatisfaction with the project, this might be an important finding. For example,

*Responses to Question 21:*

Respondent 1: XXXXXXXXXXXXXXXXXXXXX  
Respondent 2: UUUUUUUUUUUUUUUUUUUUU  
Respondent 3: CCCCCCCCCCCCCCCCCCCC  
Respondent 4: CCCCCCCCCCCCCCCCCCCC  
Respondent 5: UUUUUUUUUUUUUUUUUUUUU  
Respondent 6: CCCCCCCCCCCCCCCCCCCC  
Respondent 7: XXXXXXXXXXXXXXXXXXXXX  
Respondent 8: CCCCCCCCCCCCCCCCCCCC  
Respondent 9: UUUUUUUUUUUUUUUUUUUUU  
Respondent 10: JJJJJJJJJJJJJJJJJJJJJ

4 Respondents said C = 40%  
3 Respondents said U = 30%  
2 Respondents said X = 20%  
1 Respondent said J = 10%

**Case studies** can be helpful in providing an in-depth picture of a few examples of whatever you are studying. Often the data for case studies is collected through semi-structured interviews with individuals. Case studies can be powerful and are relatively easy to produce, in that you do not need to speak to a large number of members of the population. They may be misused or misinterpreted as being representative of the whole.

### **III. The Evaluation Scope of Work**

The SOW is the plan that lays out in a fair amount of detail all of the specifics regarding the evaluation—the who, what, where, why, how, and when of the evaluation. The SOW helps the individual or team managing the evaluation as well as the team conducting the evaluation. The SOW's structure allows you to organize all of the information about the evaluation from all of the evaluation's stakeholders.

The second role of the SOW is to communicate. The SOW is the key communication tool for the evaluation's stakeholders around the world. It also serves to inform team members about the evaluation, whether they are internal or contractors. The SOW, used in a competitive bidding process, is the basis for contractors' proposals.

A good evaluation scope of work makes clear what is expected and what questions need to be answered. Points to consider when you draft your evaluation scope of work include:

- Why is the evaluation being done? Who is the audience and what will be done with the findings?
- What problem did the project address? What are the assumptions and hypotheses about the problem and the solutions?
- What is the scope of the program or project being evaluated? What is included and what is not? [The SO description, project design documents, program reviews and related material need to be provided to the evaluator.]
- What are the research questions that the evaluation should address?
- What has been learned from other USAID projects in other countries? What other literature could help answer the questions – other development agencies' experiences, academic research?
- What methodology will be used?
- What variables and indicators can be used to measure performance, impact, cost-effectiveness, efficiency, and sustainability?
- What type of data collection and analysis will be used?
- What is the evaluation team composition? What skills are needed, and why?
- What are the work plan, schedule, personnel, logistical and budget requirements?
- What "deliverables" will the evaluator produce? When will they be completed?
- What is the Mission's marketing or utilization plan for the evaluation findings?

## **A. Outline of a Scope of Work**

The standard outline with approximate page lengths is as follows:

### I. Title and Results Framework Linkage (½ page)

### II. Purpose of the Evaluation (½ page)

- Why the evaluation is being done, who is the audience, and what will be done with the findings?

### III. Background (1-2 pages)

- What problem did the project address?
- What are the assumptions and hypotheses about the problem and the solutions?
- What has been learned from other USAID projects in other countries?
- What other literature could help inform the evaluation?

### IV. Evaluation Focus (1-4 pages)

- What is the scope of the program or project being evaluated? What is included and what is not? The SO description, project design documents, program reviews and related material need to be provided to the evaluator.
- What are the research questions that the evaluation should address?

### V. Methods and Procedures (1-2 pages)

- What type of evaluation is needed and what methodology will be used?
- What variables and indicators can be used to measure performance, impact, cost-effectiveness, efficiency, and sustainability?
- What type of data collection and analysis will be used?

### VI. Team Composition (1 page)

- What is the evaluation team composition? What skills are needed, and why?

### VII. Reporting Requirements (½ page)

- What "deliverables" will the evaluator produce?
- When will they be completed?
- What is the Mission's marketing or utilization plan for the evaluation findings?

### VIII. Logistics (1 page)

- What are the work plan, schedule, personnel, logistical and budget requirements?

### IX. Budget (1 page)

## **B. The Evaluation Team**

**Contracting a Team:** Sometimes, evaluations will be conducted by teams made up exclusively of USAID personnel. However, the long term trend at USAID involves hiring contracted personnel on evaluation teams. These contractors include USAID/W selected consulting firms, other US-based consulting firms, host-country consulting firms, or Operating Unit-selected independent consultants. For these reasons, knowing how to handle contracting tasks is an important part of managing an evaluation. Some issues to keep in mind include:

- Contracting processes may take up a significant amount of time.
- Delays can have a significant impact because often the evaluation questions are tied to a decision which has real external constraints (e.g., budget cycles) which cannot be pushed back.
- It is important to know the most appropriate vehicle for contracting to make the process as quick as possible.
- Conflicts of interest. Sometimes the best people to do a job will not or cannot do the evaluation because by doing so they will preclude themselves from future work. You also need to make sure that regardless of the contractor's willingness, you are aware of the general conflict of interest issues associated with performing the evaluation.

**Data Ownership:** "Rights to data" refers to ownership of that data after the evaluation has been conducted, as data created through the evaluation is important beyond its use for the evaluation at hand. It is important to clarify with the contractor what their rights and responsibilities are with regard to the data before entering into a contract.

**Selecting a Team:** Depending on the evaluation and its size, there are a variety of vehicles through which you might contract personnel. The most common is for a Mission to use an existing program support contract to bring an existing contractor on board for the evaluation. This process is relatively quick because it does not require a competitive bidding process; the Mission simply issues a task order under an existing contract. Even when such a vehicle is used, the Mission can request that the contractor subcontract to a particular individual consultant. Such an arrangement can be an effective way to get a known consultant on a job, but you should keep several caveats in mind as discussed below in Team Considerations.

Indefinite Quantity Contracts (IQC) are used to mobilize US-based contractors relatively rapidly. With IQCs, contractors have already been pre-approved through a competitive award process for such work. IQCs are much faster ways to secure contractor support than by issuing a new Request for Proposals and reviewing bids.

Lastly, in some limited cases, independent consultants might be hired for short jobs through purchase orders. While purchase orders are relatively quick, their use is limited, however, because they are intended for small value goods and services.

### **C. Time and Money**

Break down the evaluation into manageable parts and then make estimates which are "rolled up" into a total. The work is broken down into a series of tasks and subtasks, much like an outline. Then estimate the number of labor hours it will take to perform each subtask (or the smallest unit in the outline). Sum these up by task and then arrive at a grand total. That is the level of effort or total number of labor hours required.

**Direct costs:** In addition to labor, there usually are other sorts of costs: travel, hotel, per diem, and other direct costs (ODCs).

**Other direct costs:** Usually include costs like photocopying, report binders, or database charges for a literature review-think of it as a miscellaneous column. These costs and labor should be provided by task; that way, you can look at the cost for each task and if one looks way too expensive for what you get, you can comfortably cut it out and know how it will affect your bottom line.

In order to plan for the evaluation, it is important to have an idea of the different phases of the evaluation, and the amount of time devoted to each phase. This will make it much easier to budget time and resources

Having a clear idea of how much time and how many people are needed at each stage will help you create a budget for the evaluation, and will help the evaluation team manage their time.

**Time:** When preparing for an evaluation, it is important to recognize that each step in the process (planning, developing interview guides, training interviewers on how to use them, collecting and analyzing data, and writing the report] takes time. It is important to remember that the time plan for an evaluation also includes identifying the evaluation's stakeholders, drafting of the SOW, and identifying the evaluation questions. Once the evaluation has been completed, time is needed for disseminating the results and making sure the results are utilized.

### **D. Evaluation Report**

Good evaluations achieve the following:

- Provide new knowledge and contribute to organizational learning rather than merely confirm what is already known.
- Are carried out from an independent perspective, using innovative approaches.
- Have credibility, which is dependent on the quality of the evaluator, survey methods used, the expression of both positive and negative findings, and the degree of open and critical discussion concerning the evaluation implementation and conclusions between the various interested parties.
- The results are presented at a point when this type of information is needed. This demands advanced planning on the part of those designing the evaluation.

- Results are utilized. This requires that conclusions and possible recommendations are relevant, that credibility is high, that they are presented clearly and understandably, and that they are disseminated to interested parties.

Good evaluation reports should be organized by finding, conclusion and recommendation.

A **finding** is a "factual statement" about the project results based on the collected information, e.g. "90% of the loan beneficiaries were men".

A **conclusion** is a "synthesis of findings" corresponding to a specific circumstance based on the evaluator's appreciation, e.g. "the credit scheme failed to benefit women owned enterprises".

A **recommendation** is a "prescription" of what should be done based on the conclusions about a specific circumstance, e.g. "define a quota system for the approval of loan applications to increase the access to credit to the women target groups".

A **lesson learned** is a "generalization" about a specific circumstance (a project) which could be applied to a "type of situation" (other projects), e.g. "a clear selection criteria is necessary to ensure reaching the intended beneficiaries in setting-up projects' credit schemes for the rural poor in the highlands".

## **E. Utilizing Evaluation Results**

Evaluation should be used to inform strategy, generate lessons learned, and increase our understanding of development issues. To underscore the point that evaluations should be used, Administrator Natsios noted in a recent Agency Notice on Evaluations that,

*“Regional and pillar bureaus and Missions have the responsibility to produce evaluations that inform decision-making and better tell USAID's story. Evaluations should be commissioned not only to help the specific operating unit but also to generate and share broader lessons learned that can be factored into Agency policies and programs.”*

The OECD makes the following recommendations for improving the utilization of evaluation results

#### **BOX 7. ACTION POINTS TO IMPROVE EVALUATION FEEDBACK**

- Understand how learning happens within and outside the organization (identify where the blockages occur);
- Assess how the relevance and timeliness of evaluation feedback can be improved, and ensure that this happens;
- Be explicit in identifying key audiences for evaluation feedback and the reasons for wanting to reach them, both in general and in specific cases;
- Get to know target groups better to learn what they want from evaluations, how they use evaluation information, and how feedback systems can respond better to these demands;
- Develop a more strategic view of how feedback approaches can be tailored to the needs of different audiences;
- Make sure the quality of evaluation outputs is up to standard—particularly in terms of brevity, clarity and presentation;
- Consider diversifying the range of approaches used to communicate with audiences, using innovative methods where appropriate;
- Improve evaluation websites and intranets, recognizing that ease of access and user-friendliness are key factors;
- Ensure that full disclosure of evaluation reports becomes the norm and that proper approval and notification processes are in place so that senior management or key partners are not caught unawares by controversial findings;
- Put more effort into finding better ways of involving country-level stakeholders in evaluation work, including the feedback of evaluation lessons, recognizing that language barriers are a key constraint;
- Recruit specialist staff where necessary to fill skills gaps, particularly in communications work.

## Appendix A

The evaluation IQC contract is a valuable resource to help USAID Washington Bureaus and Field Missions monitor program performance and evaluate impact. The IQCs will help USAID to:

- Undertake activity-, program-, results- and goal-level evaluations and related performance monitoring.
- Re-examine or test the validity of hypotheses and assumptions embedded in strategic objectives and results frameworks.
- Determine whether the needs of intended customers are being met.
- Determine whether conditions for sustainability exist.
- Distill lessons learned, which may be useful elsewhere in the Agency.
- Provide evaluation methodology design and training.

To access the IQCs, an office or mission will first check with Joseph Lieberson of PPC, who will review the scope of work to make sure the requirements are consistent with the IQC objectives. The next step is to have your own contracting officer determine how an individual IQC firm will be selected. After a firm is selected, the contracting officer will negotiate an IQC Task Order.

The four Evaluation Support Service IQC firms are listed below.

### **1. Checchi/Louis Berger Joint Venture AEP-I-00-00-00022-00**

1899 L Street, NW - Suite 800  
Washington, DC 20036  
Charles Bell: Louis Berger, Project Manager, Technical Management, Patricia McPhelim: Checchi, Deputy Manager, Financial Management  
(202) 452-9700  
FAX (202) 466-9070  
pmcphelim@checchiconsulting.com

#### ***Sub-Contractors:***

CRI Consult  
IncDevTech Systems, Inc.  
Datex, Inc. TVT Associates, Inc.

### **2. Development Associates, Inc. AEP-I-00-00-00023-00**

1730 North Lynn Street  
Arlington, VA 22209  
Peter Davis, President  
(703) 276-0677  
FAX (703) 276-0432  
Mal Young, [Myoung@devassoc1.com](mailto:Myoung@devassoc1.com)  
Jack Sullivan  
John Garcia

***Sub-Contractors:***

Nathan Associates, Inc. and  
International Programs Consortium, Inc.

**3. Management Systems International** AEP-I-00-00-00024-00

600 Water Street, SW  
Washington, DC 20024  
Marina Fanning, Project Director

Roberta Warren, IQC Manager  
(202) 484-7170  
FAX 488-0754  
Rwarren@MSI-inc.com  
Keith Brown (X124), Stacy Stacks, Technical and Administrative Managers

***Sub-Contractors:***

Academy for Educational Development (AED)  
Juarez and Associates Associates for International Resources and Development  
(ARID)  
MetaMetrics, Inc.  
International Resource Group (IRG)  
Carolina Population Center

**4. The Mitchell Group** AEP-I-00-00-00025-00

1816 11<sup>th</sup> Street, NW  
Washington, DC 20001  
Jenkins E. Cooper, Director of Operations  
(202) 745-1919  
FAX 202)234-1697  
[Jcooper1@erols.com](mailto:Jcooper1@erols.com)  
Ken Kornher, IQC Manager,  
John Pittman, Program Associate

***Sub-Contractor:***

The South-East Consortium for International Development

## Appendix B

EvalWeb is a PPC-sponsored tool for evaluation practitioners:

<http://www.dec.org/partners/evalweb/>

EvalWeb contains:

- tools and resources to help you with evaluations;
- contact information and links to other evaluators and evaluation organizations;
- lists and links to recent and upcoming evaluations; and
- a collaborative online community where practitioners and others can exchange ideas and resources.

An impetus for this site is the decline in evaluations and assessments that are submitted to USAID's document repository, the Development Experience Clearinghouse.

# Appendix C

## Program Evaluations Meta-evaluation Checklist

### PROGRAM EVALUATIONS METAEVALUATION CHECKLIST (Based on *The Program Evaluation Standards*)

[pdf version](#)

Daniel L. Stufflebeam

2000

*This checklist is for performing final, summative metaevaluations. It is organized according to the Joint Committee [Program Evaluation Standards](#). For each of the 30 standards the checklist includes 6 checkpoints drawn from the substance of the standard. It is suggested that each standard be scored on each checkpoint. Then judgments about the adequacy of the subject evaluation in meeting the standard can be made as follows: 0-1 Poor, 2-3 Fair, 4 Good, 5 Very Good, 6 Excellent. It is recommended that an evaluation be failed if it scores Poor on standards P1 Service Orientation, A5 Valid Information, A10 Justified Conclusions, or A11 Impartial Reporting. Users of this checklist are advised to consult the full text of *The Joint Committee (1994) Program Evaluation Standards*, Thousand Oaks, CA: Sage Publications.*

#### TO MEET THE REQUIREMENTS FOR *UTILITY*, PROGRAM EVALUATIONS SHOULD:

##### U1 Stakeholder Identification

- Clearly identify the evaluation client
- Engage leadership figures to identify other stakeholders
- Consult stakeholders to identify their information needs
- Ask stakeholders to identify other stakeholders
- Arrange to involve stakeholders throughout the evaluation, consistent with the formal evaluation agreement
- Keep the evaluation open to serve newly identified stakeholders

6 = Excellent

5 = Very Good

4 = Good

2-3 = Fair

0-1 = Poor

---

##### U2 Evaluator Credibility

- Engage competent evaluators
- Engage evaluators whom the stakeholders trust
- Engage evaluators who can address stakeholders' concerns
- Engage evaluators who are appropriately responsive to issues of gender, socioeconomic status, race, and language and cultural differences
- Help stakeholders understand and assess the evaluation plan and process
- Attend appropriately to stakeholders' criticisms and suggestions

6 = Excellent

5 = Very Good

4 = Good

2-3 = Fair

0-1 = Poor

---

##### U3 Information Scope and Selection

- Assign priority to the most important questions
- Allow flexibility for adding questions during the evaluation
- Obtain sufficient information to address the stakeholders' most important evaluation questions
- Obtain sufficient information to assess the program's merit

- Obtain sufficient information to assess the program's worth
- Allocate the evaluation effort in accordance with the priorities assigned to the needed information

6 = Excellent      5 = Very Good      4 = Good      2-3 = Fair      0-1 = Poor

---

#### **U4 Values Identification**

- Consider all relevant sources of values for interpreting evaluation findings, including societal needs, customer needs, pertinent laws, institutional mission, and program goals
- Determine the appropriate party(s) to make the valuational interpretations
- Provide a clear, defensible basis for value judgments
- Distinguish appropriately among dimensions, weights, and cut scores on the involved values
- Take into account the stakeholders' values
- As appropriate, present alternative interpretations based on conflicting but credible value bases

6 = Excellent      5 = Very Good      4 = Good      2-3 = Fair      0-1 = Poor

---

#### **U5 Report Clarity**

- Issue one or more reports as appropriate, such as an executive summary, main report, technical report, and oral presentation
- As appropriate, address the special needs of the audiences, such as persons with limited English proficiency
- Focus reports on contracted questions and convey the essential information in each report
- Write and/or present the findings simply and directly
- Employ effective media for informing the different audiences
- Use examples to help audiences relate the findings to practical situations

6 = Excellent      5 = Very Good      4 = Good      2-3 = Fair      0-1 = Poor

---

#### **U6 Report Timeliness and Dissemination**

- In cooperation with the client, make special efforts to identify, reach, and inform all intended users
- Make timely interim reports to intended users
- Have timely exchanges with the pertinent audiences, e.g., the program's policy board, the program's staff, and the program's customers
- Deliver the final report when it is needed
- As appropriate, issue press releases to the public media
- If allowed by the evaluation contract and as appropriate, make findings publicly available via such media as the Internet

6 = Excellent      5 = Very Good      4 = Good      2-3 = Fair      0-1 = Poor

---

#### **U7 Evaluation Impact**

- As appropriate and feasible, keep audiences informed throughout the evaluation
- Forecast and serve potential uses of findings
- Provide interim reports



- Be efficient
- Make use of in-kind services
- Inform decisions
- Foster program improvement
- Provide accountability information
- Generate new insights

6 = Excellent      5 = Very Good      4 = Good      2-3 = Fair      0-1 = Poor

---

**SCORING THE EVALUATION FOR FEASIBILITY:**

**Add the following:**

Number of *Excellent* ratings (0-3)      \_\_\_ x 4 = \_\_\_

Number of *Very Good* ratings (0-3)      \_\_\_ x 3 = \_\_\_

Number of *Good* ratings (0-3)      \_\_\_ x 2 = \_\_\_

Number of *Fair* ratings (0-3)      \_\_\_ x 1 = \_\_\_

Total Score      = \_\_\_

**Strength of the Evaluation's provisions for Feasibility:**

11 (92%) to 12      **Excellent**

8 (69%) to 10      **Very Good**

6 (50%) to 7      **Good**

3 (25%) to 5      **Fair**

0 (0%) to 2      **Poor**

**TO MEET THE REQUIREMENTS FOR PROPRIETY, PROGRAM EVALUATIONS SHOULD:**

**P1 Service Orientation**

- Assess program outcomes against targeted and nontargeted customers' assessed needs
- Help assure that the full range of rightful program beneficiaries are served
- Promote excellent service
- Identify program strengths to build on
- Identify program weaknesses to correct
- Expose persistently harmful practices

6 = Excellent      5 = Very Good      4 = Good      2-3 = Fair      0-1 = Poor

---

**P2 Formal Agreements**

**Reach advance written agreements on:**

- Evaluation purpose and questions
- Audiences
- Editing
- Release of reports
- Evaluation procedures and schedule
- Evaluation resources

6 = Excellent      5 = Very Good      4 = Good      2-3 = Fair      0-1 = Poor

---

**P3 Rights of Human Subjects**

- Follow due process and uphold civil rights
- Understand participants' values
- Respect diversity
- Follow protocol
- Honor confidentiality/anonymity agreements
- Minimize harmful consequences of the evaluation

6 = Excellent      5 = Very Good      4 = Good      2-3 = Fair      0-1 = Poor

---

#### **P4 Human Interactions**

- Consistently relate to all stakeholders in a professional manner
- Honor participants' privacy rights
- Honor time commitments
- Be sensitive to participants' diversity of values and cultural differences
- Be evenly respectful in addressing different stakeholders
- Do not ignore or help cover up any participant's incompetence, unethical behavior, fraud, waste, or abuse

6 = Excellent      5 = Very Good      4 = Good      2-3 = Fair      0-1 = Poor

---

#### **P5 Complete and Fair Assessment**

- Assess and report the program's strengths and weaknesses
- Report on intended and unintended outcomes
- As appropriate, show how the program's strengths could be used to overcome its weaknesses
- Appropriately address criticisms of the draft report
- Acknowledge the final report's limitations
- Estimate and report the effects of the evaluation's limitations on the overall judgment of the program

6 = Excellent      5 = Very Good      4 = Good      2-3 = Fair      0-1 = Poor

---

#### **P6 Disclosure of Findings**

- Clearly define the right-to-know audiences
- Report relevant points of view of both supporters and critics of the program
- Report balanced, informed conclusions and recommendations
- Report all findings in writing, except where circumstances clearly dictate otherwise
- In reporting, adhere strictly to a code of directness, openness, and completeness
- Assure the reports reach their audiences

6 = Excellent      5 = Very Good      4 = Good      2-3 = Fair      0-1 = Poor

---

#### **P7 Conflict of Interest**

- Identify potential conflicts of interest early in the evaluation
- As appropriate and feasible, engage multiple evaluators?
- Maintain evaluation records for independent review



## A2 Context Analysis

- Describe the context's technical, social, political, organizational, and economic features
- Maintain a log of unusual circumstances
- Report those contextual influences that appeared to significantly influence the program and that might be of interest to potential adopters
- Estimate the effects of context on program outcomes
- Identify and describe any critical competitors to this program that functioned at the same time and in the program's environment
- Describe how people in the program's general area perceived the program's existence, importance, and quality

6 = Excellent

5 = Very Good

4 = Good

2-3 = Fair

0-1 = Poor

---

## A3 Described Purposes and Procedures

- Monitor and describe how the evaluation's purposes stay the same or change over time
- As appropriate, update evaluation procedures to accommodate changes in the evaluation's purposes
- Record the actual evaluation procedures, as implemented
- When interpreting findings, take into account the extent to which the intended procedures were effectively executed
- Describe the evaluation's purposes and procedures in the summary and full-length evaluation reports
- As feasible, engage independent evaluators to monitor and evaluate the evaluation's purposes and procedures

6 = Excellent

5 = Very Good

4 = Good

2-3 = Fair

0-1 = Poor

---

## A4 Defensible Information Sources

- Once validated, use pertinent, previously collected information
- As appropriate, employ a variety of data collection sources and methods
- Document and report information sources
- Document, justify, and report the means used to obtain information from each source
- Include data collection instruments in a technical appendix to the evaluation report
- Document and report any biasing features in the obtained information

6 = Excellent

5 = Very Good

4 = Good

2-3 = Fair

0-1 = Poor

---

## A5 Valid Information

- Focus the evaluation on key questions
- Assess and report what type of information each employed procedure acquires
- Document how information from each procedure was scored, analyzed, and interpreted
- Report and justify inferences singly and in combination
- Assess and report the comprehensiveness of the information provided by the procedures as a set in relation to the information needed to answer the set of evaluation questions
- Establish meaningful categories of information by identifying regular and recurrent themes in information collected using qualitative assessment procedures

6 = Excellent

5 = Very Good

4 = Good

2-3 = Fair

0-1 = Poor

---

**A6 Reliable Information**

- Identify and justify the type(s) and extent of reliability claimed
- As feasible, choose measuring devices that in the past have shown acceptable levels of reliability for their intended uses
- In reporting reliability of an instrument, assess and report the factors that influenced the reliability, including the characteristics of the examinees, the data collection conditions, and the evaluator's biases
- Check and report the consistency of scoring, categorization, and coding
- Train and calibrate scorers and analysts to produce consistent results
- Pilot test new instruments in order to identify and control sources of error

6 = Excellent

5 = Very Good

4 = Good

2-3 = Fair

0-1 = Poor

---

**A7 Systematic Information**

- Establish protocols and mechanisms for quality control of the evaluation information
- Verify data entry
- Proofread and verify data tables generated from computer output or other means
- Systematize and control storage of the evaluation information
- Strictly control access to the evaluation information according to established protocols
- Have data providers verify the data they submitted

6 = Excellent

5 = Very Good

4 = Good

2-3 = Fair

0-1 = Poor

---

**A8 Analysis of Quantitative Information**

- Whenever possible, begin by conducting preliminary exploratory analyses to assure the data's correctness and to gain a greater understanding of the data
- Report limitations of each analytic procedure, including failure to meet assumptions
- Employ multiple analytic procedures to check on consistency and replicability of findings
- Examine variability as well as central tendencies; identify and examine outliers, and verify their correctness

6 = Excellent

5 = Very Good

4 = Good

2-3 = Fair

0-1 = Poor

- Limit conclusions to the applicable time periods, contexts, purposes, questions, and activities
- Report alternative plausible conclusions and explain why other rival conclusions were rejected
- Cite the information that supports each conclusion
- Identify and report the program's side effects
- Warn against making common misinterpretations
- Whenever feasible and appropriate, obtain and address the results of a prerelease review of the draft evaluation report

6 = Excellent                      5 = Very Good                      4 = Good                      2-3 = Fair                      0-1 = Poor

---

#### **A11 Impartial Reporting**

- Engage the client to determine steps to ensure fair, impartial reports
- Safeguard reports from deliberate or inadvertent distortions
- As appropriate and feasible, report perspectives of all stakeholder groups and, especially, opposing views on the meaning of the findings
- As appropriate and feasible, add a new, impartial evaluator late in the evaluation to help offset any bias the original evaluators may have developed due to their prior judgments and recommendations
- Describe steps taken to control bias
- Participate in public presentations of the findings to help guard against and correct distortions by other interested parties

6 = Excellent                      5 = Very Good                      4 = Good                      2-3 = Fair                      0-1 = Poor

---

#### **A12 Metaevaluation**

- Budget appropriately and sufficiently for conducting an internal metaevaluation and, as feasible, an external metaevaluation
- Designate or define the standards the evaluators used to guide and assess their evaluation
- Record the full range of information needed to judge the evaluation against the employed standards
- As feasible and appropriate, contract for an independent metaevaluation
- Evaluate all important aspects of the evaluation, including the instrumentation, data collection, data handling, coding, analysis, synthesis, and reporting
- Obtain and report both formative and summative metaevaluations to the right-to-know audiences

6 = Excellent                      5 = Very Good                      4 = Good                      2-3 = Fair                      0-1 = Poor

---

#### **SCORING THE EVALUATION FOR ACCURACY:**

##### **Add the following:**

Number of <i>Excellent</i> ratings (0-12)	_____ x 4 = _____
Number of <i>Very Good</i> ratings (0-12)	_____ x 3 = _____
Number of <i>Good</i> ratings (0-12)	_____ x 2 = _____
Number of <i>Fair</i> ratings (0-12)	_____ x 1 = _____
Total Score	= _____

**Strength of the Evaluation's provisions for Accuracy:**

45 (94%) to 48	<b><i>Excellent</i></b>
33 (69%) to 44	<b><i>Very Good</i></b>
24 (50%) to 32	<b><i>Good</i></b>
12 (25%) to 23	<b><i>Fair</i></b>
0 (0%) to 11	<b><i>Poor</i></b>

*This checklist is being provided as a free service to the user. The provider of the checklist has not modified or adapted the checklist to fit the specific needs of the user and the user is executing his or her own discretion and judgment in using the checklist. The provider of the checklist makes no representations or warranties that this checklist is fit for the particular purpose contemplated by user and specifically disclaims any such warranties or representations.*

## **Appendix D**

### **Guidelines for Conducting Meetings<sup>6</sup>**

1. CIRCULATE THE AGENDA PRIOR TO THE MEETING SO PARTICIPANTS CAN COME PREPARED IF NECESSARY.
2. ENSURE APPROPRIATE PEOPLE ARE INVITED TO THE MEETING.
3. START THE MEETING ON TIME.
4. DISCUSS AND GET FINAL AGREEMENT ON MEETING PURPOSE, DESIRED OUTCOMES, AND AGENDA

Begin by quickly sharing your purpose statement, the desired outcomes, the agenda, and the time frame; allow for any final input. Have the agenda on a flip chart or a handout.

5. USE THE AGENDA AS A GUIDE TO RUN THE MEETING

Run the meeting in order of priority of agenda items—unless there is a compelling reason to start with a less important item (for example, you want to tackle an easy item to get going, such as announcing a special event).

6. MONITOR TIME, AND RUN THE MEETING INTENTIONALLY AGAINST TIME AGREEMENTS

Agree on how long the meeting will last. Monitor the time, keep moving, begin and end on time—unless you work out new arrangements with the team.

7. HELP EVERYONE CONTRIBUTE AND KEEP THE MEETING PARTICIPATIVE BUT FOCUSED

Facilitate the discussion, making sure meeting participants are actively involved, staying on the topic, listening to one another, taking turns speaking, etc. Manage the process of the meeting, but avoid dominating it.

8. AS APPROPRIATE, FACILITATE THE TEAM TO MOVE BETWEEN CREATIVE WORK ON ONE HAND, AND ANALYSIS AND DECISION MAKING ON THE OTHER

Lead the group appropriately in both creative, expansive efforts and focused discussions and decisions. Effective teams will at times tap all their resources to be creative and nonjudgmental in order to evoke extraordinary thinking. At other times, they focus, judge, agree on decisions, and move on.

9. KEEP TRACK OF AGREEMENTS AND DECISIONS AS THEY ARE REACHED

---

<sup>6</sup> Source: TRG Inc.

Summarize decisions and outcomes explicitly when they are reached (“So, we have just agreed that ... is the problem and that ... will ...”). Keep a running record of decisions and outcomes, either in your notes or on a piece of flip chart paper or board that everyone can see. (In longer, more complicated meetings, we especially recommend the latter since making agreement summaries explicit will make the summary discussion at the end of the meeting much easier and more effective.)

#### 10.SUMMARIZE AND CHECK IMPORTANT AGREEMENTS CLEARLY AND CAREFULLY AT THE END OF THE MEETING

At the end of the meeting, review decisions, outcomes, and agreements. As appropriate, designate time frames and people responsible for carrying out tasks. If appropriate, capture the agreed upon summary in written form for future referral (e-mail/meeting minutes, etc.)

#### 11.PERIODICALLY, DO A QUICK CHECK TO SEE IF THERE IS ANYTHING THAT COULD BE DONE TO MAKE THE MEETINGS MORE PRODUCTIVE

Check to see if the team thought the meeting was productive. If not, agree on what can be done next time to make it better.

## **Appendix E**

### **Using the Flipchart as an Effective Facilitation Tool, by Steve Joyce**

#### **Why Use the Flipchart?**

The flipchart is an important facilitator tool, not only for making presentations, but also for recording relevant information as it is generated. The most important advantage that flipcharts have over chalkboards and overhead projectors is that flipcharts facilitate group memory.

The concept of group memory involves two elements:

1. **Retention and Reference:** When participants are able to see a presentation's key points listed on a flipchart, and then have visual access to these flipcharts throughout the meeting or discussion, the repeated exposure ensures greater retention of the information.

Taping flipcharts to the walls allows participants to refer back to key points - reinforcing these points with their own examples, and building them with new ideas. Participants become more engaged, and there is greater group synergy and interaction.

2. **Visual Record of Outcomes:** When the facilitator records ideas and suggestions on a flipchart as they are generated, participants have a visual "memory" of key points as the session progresses. Communication is clearer, because of the visual record that the flipcharts provide. And, at the meeting's conclusion, participants have a collective memory of agreements and outcomes, whether these are decisions, next steps, or new ideas.

Chalkboards, overhead projectors, and even the new electronic recording books with photocopy capability cannot match the flipchart for generating group memory. You can leave flipcharts hanging on the walls for days, allowing participants a chance to stay after a session - or to return early the next day - to review them. Chalkboards, on the other hand, are erased as the session continues, and overheads flashed onto the wall quickly disappear into the darkness.

#### **What You Should Chart**

Flipcharts can serve as a useful tool in most situations, but not everything that goes on in a session must be charted. Outlined below are components that in most cases should be charted.

1. **The Agenda**

Whether the agenda is predetermined or developed at the beginning of a session, it should be recorded on a flipchart. Doing so encourages shared

responsibility in achieving the agenda, as well as ongoing assessment of progress.

## 2. Key Presentation Points

Preparing flipcharts to accompany a presentation, as noted earlier, increases the likelihood that participants will retain more, and build on the information as a group. The flipcharts, which can be prepared in advance, should only highlight key points as clearly and succinctly as possible. In determining key points, the session leader should ask himself or herself, “What are the most important messages that I want people to grasp?”

## 3. Proposals/New Ideas

Record on the flipchart new ideas and proposals that are generated during brainstorming sessions or participant presentations. This visual record captures the ideas and proposals so that they are not lost, and it also lets people know that their ideas have been heard.

Meetings often end up stalled (or, worse yet, result in conflict) when participants repeat suggestions and proposals they think haven’t been heard or accepted. Recording ideas on flipcharts acknowledges them, without necessarily requiring any further action.

## 4. Alternate Issues

Issues often surface that are not part of the planned agenda. It is beneficial to capture these issues on the flipchart to acknowledge them (for the same reasons stated before), and, if appropriate, to develop strategies for resolving them.

## 5. Action Items/Next Steps

Although participants may leave sessions agreeing on what needs to be done, they often are not sure about who will do what, by when. Flipcharts help consolidate a group’s agreement on next steps. It is also helpful to type all action items and next steps from the flipcharts and hand them out as a written reminder to participants.

### **Tips for Charting**

Using flipcharts as an effective facilitation tool is not always as easy as it might seem. How you prepare flipcharts and how you record ideas and important points can have an impact upon effective communication. Listed below are some tips for producing flipcharts that enhance communication.

- Words recorded on a flipchart should be large enough for all participants to read comfortably. Use the flat edge of the marker, as opposed to the tip, so that letters have some thickness to them.

- Use words sparingly on a flipchart. Only record major points and key phrases. Ask yourself when preparing flipcharts, “What is essential for participants to remember?”
- When recording participant responses and ideas on a flipchart during a meeting, capture the essence of what the speaker is saying in as few words as possible. Use the speaker’s own words if they are clear and appropriate. If not, paraphrase back what you believe the speaker intended to say, and, if you are correct, then record those phrases.
- After charting, read information back to the group, and ask if any clarification is necessary. It is important that the words recorded on the flipchart reflect what the participants intended to say.
- When preparing flipcharts before a session, use different colored markers to write and highlight words (however, don’t use more than three colors on any one flipchart). Flipcharts thus become more attractive visual aids. Color also helps draw attention to a particular keyword or phrase.
- Don’t use light-colored markers (red, yellow, pink, etc.). Although these colors highlight words well, participants who are more than a few feet away will have difficulty reading words written on them.
- Be creative with your flipcharts. Box in key words, use arrows, and draw figures that illustrate important points.

### **The Flipchart and Facilitator Style**

Finally, a few words should be mentioned about the flipchart and facilitator style. The flipchart is a facilitation tool, not a crutch. It’s not intended to take the place of your lecture notes. The “talking points” on your flipchart serve to focus the group’s attention, and assist you in keeping your thoughts organized. When referring to points on the flipchart, don’t hide behind the flipchart stand. Move to the side or off to the front as you speak. Touch key words on the flipchart once in a while to give them emphasis.

One way to keep people from reading ahead and losing focus is to place a small strip of masking tape at the bottom center of the page, and bring the bottom of the page up to cover the flipchart to the point on which the current discussion is based.

Organize multiple flipcharts on a given topic from right to left on easels or across a wall. If you hang your flipcharts up on the wall prior to the session, keep them covered, and rehearse their location to avoid a frantic search in the middle of your presentation. For flipcharts layered on one easel, a small piece of masking tape makes a good tab.

Keep any task instructions displayed on a flipchart throughout the task so participants can refer to them as needed. If you are using flipcharts to help the group record its thoughts or actions, you can display the charts on the wall as they are developed.