



**USAID**  
FROM THE AMERICAN PEOPLE

# TRENDS IN INTERNATIONAL DEVELOPMENT EVALUATION THEORY, POLICY AND PRACTICES

17 AUGUST 2009

This publication was produced for the United States Agency for International Development. It was prepared by Molly Hageboeck, Management Systems International.

# TRENDS IN INTERNATIONAL DEVELOPMENT EVALUATION THEORY, POLICY AND PRACTICES

Contracted under RAN-M-OO-Q4-00049-A-FYOS-072  
IMR2 Assignment Plan 72: Development Evaluation Theory and Practice

**August 17, 2009**



600 Water Street, SW  
Washington, DC 20024



## **DISCLAIMER**

The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

# CONTENTS

<b>PREFACE</b> .....	<b>ii</b>
<b>1. EXECUTIVE SUMMARY</b> .....	<b>1</b>
<b>2. STUDY PURPOSE AND SCOPE</b> .....	<b>2</b>
<b>3. METHODOLOGY</b> .....	<b>3</b>
<b>4. SIGNIFICANT DEVELOPMENT EVALUATION TRENDS</b> .....	<b>3</b>
A. TRENDS IN EVALUATION THEORY .....	4
B. TRENDS IN EVALUATION POLICY AND PRACTICE.....	5
C. NEW THINKING -- THE ROLE AND PLACEMENT OF EVALUATION OFFICES.....	29
<b>5. OPPORTUNITIES FOR IMPROVING EVALUATION AND ITS UTILIZATION IN USAID</b> .....	<b>33</b>
A. UTILIZATION OF EVALUATION RESULTS.....	34
B. STAKEHOLDER “OWNERSHIP” OF THE EVALUATION PROCESS .....	37
C. STRENGTHENING EVALUATION QUALITY AND EVIDENCE .....	38
D. IMPROVING EVALUATION FOLLOW-UP .....	41
<b>ANNEX A. PATTERNS IN RECENT USAID EVALUATIONS IN THE DEVELOPMENT EXPERIENCE CLEARINGHOUSE</b> .....	<b>44</b>
<b>ANNEX B. SAMPLE USAID EVALUATION SUMMARY (FORM AID I330- 5 (10/87))</b> .....	<b>47</b>
<b>BIBLIOGRAPHY</b> .....	<b>56</b>

## PREFACE

This study was one of two reports commissioned by USAID’s Office of Evaluation. Molly Hageboeck served as the principal investigator for this study. She was supported by Micah Frumkin who managed an analysis of USAID evaluations carried out between 2005 and 2008, which is included as Annex A. In parallel, MSI conducted a review of the broader U.S. foreign assistance context and institutional environment in which USAID currently operates and its implications for evaluation, entitled *Trends in U.S. Foreign Assistance Over the Past Decade*. This second study was led by Keith Brown and Jill Tirnauer.

# I. EXECUTIVE SUMMARY

## STUDY PURPOSE AND METHODS

This study was undertaken to provide the U.S. Agency for International Development (USAID) with a portrait of trends in international development evaluation over the past decade. Coverage includes the findings of previous reviews of trends in evaluation; information on other donor evaluation policies and guidelines; and an analysis of trends in USAID evaluation reports for the period 2005 through 2008. The paper includes options for USAID's new Office of Evaluation that are consistent with current donor practices and USAID's own history.

## TRENDS IN EVALUATION THEORY

Evaluation theory, which focuses on the way in which evaluation interacts with the social programs it examines, has evolved slowly since the late 1960s when it was first introduced in U.S. Government agencies. Advances in evaluation theory have focused on the purposes served by evaluations; options with respect to evaluation timing; the role of the evaluator, including evaluation competencies and evaluation ethics; what constitutes "good enough" evaluation evidence; what methods are best suited for various evaluation tasks; and when and how evaluation is utilized. Trends in evaluation policy and practice draw on evaluation theory, but their reach is just as likely to be to the entire body of evaluation theory as it is to be the most recent innovation. For example, current interest by some donor agencies in improving evaluation rigor has not looked to new developments in evaluation theory but, rather, doubled back to classic works produced in the mid-1970s. This study found no strong linear link between the progression of evaluation theory and current trends in international development evaluation policy and practice.

## TRENDS IN EVALUATION POLICY AND PRACTICE

Changes in donor agency evaluation policy and practice described in this paper have been triggered by multiple forces. Trends in the way development assistance is conceptualized, funded, and managed constitute one of these forces. Criticism of development assistance donor evaluation practices, such as inadequate evidence of evaluation use and impact, is another of these factors.

Broadly speaking, shifts in the way evaluations are conducted in donor organizations over the past decade have a common goal: *enhancing aid effectiveness*. Donor explanations for encouraging specific evaluation practices recognize that evaluation utilization is the path through which evaluation impacts development effectiveness. Some of the trends identified by this study focus directly on improvements in post-evaluation follow-up and utilization while others focus on improving evaluation ownership, enhancing evaluation quality, and the optimal placement for an evaluation office within a donor agency.

Donor initiatives aimed at improving evaluation use during the decade have focused primarily on the immediate use of evaluation findings and recommendations for improving existing programs and planning follow-on interventions, rather than on the cumulative impact on multiple evaluations, over time, or on the way practitioners think about the development process. Innovative practices in donor agencies include efforts to spotlight where the utilization of evaluation has had an impact, e.g., case examples in the World Bank's publication, *Influential Evaluations*. Interest in improved access to and use of evaluation organizations has also driven donor efforts to create electronic libraries accessible through agency websites and, at least in AusAID, targeted dissemination plans for every evaluation. Improved utilization is also the focus of donor efforts to synthesize evaluation findings on a topical basis, e.g., in the UNDP's *Essentials* series and CIDA's *What are We Learning?* series. A number of donors have also strengthened their guidelines for post-evaluation follow-up and several organizations (SIDA, World Bank, DFID and UNDP) are monitoring the frequency with which evaluation recommendations are accepted and, once accepted, whether or not they are implemented and the degree of thoroughness of the implementation.

Donor initiatives aimed at enhancing a sense of “ownership” among potential evaluation users, while not new, have expanded over the decade with the greatest shift in this period being towards more joint evaluations. This shift is linked in part to an emphasis on mutual accountability in the Paris Declaration, to which many donors subscribe. New guidance in this area has been issued by the Development Assistance Committee (DAC), and European donors are accelerating their use of joint evaluations, e.g., half of all DANIDA evaluations are now joint evaluations. While this trend builds on earlier efforts that highlight the importance of stakeholder participation in the evaluation process, joint evaluations vary considerably in the degree to which they utilize highly participatory evaluation methods, e.g., direct stakeholder involvement in data collection.

Among the strongest trends in evaluation policy and practice during the decade are those that focus on improving the quality of donor-funded evaluations. New policies and practices in this area have, in part, been stimulated by critiques of the quality of evidence found in donor-funded evaluations, including those identified in the best known of these reviews, the 2006 Center for Global Development (CGD) report titled *When Will We Ever Learn? Improving Lives Through Impact Evaluation*. This and other critical reviews link weak utilization of evaluation reports to their credibility. Advocating more rigorous evaluation, CGD has taken up the issue not only with USAID but also with a number of other donors, primarily through a conference in Bellagio, Italy. Responding to these critiques, some donors have issued new evaluation policies that call for at least some evaluations that meet rigorous standards for evidence, e.g., AusAID, NZAID, the World Bank and the Millennium Challenge Corporation in the U.S. In parallel, donors that were already investing in evaluations that involved experimental designs (randomized controlled trials) and quasi-experimental designs using innovative econometric techniques, as done by the Inter-American Development Bank and the World Bank, have made a greater effort to highlight these types of evaluations.

With quality improvement in mind, several donors have updated their evaluation guidelines during the decade and some have introduced processes that monitor compliance with evaluation standards. The European Commission, *Agence Française de Développement* and NZAID all use rating systems to score evaluations against published evaluation standards, while UNICEF has a system for scoring the quality of evaluation reports. In addition, UNDP, DFID and CIDA have begun to introduce minimum competency for evaluation staff and heads of evaluation offices. Evaluation training opportunities for both donor staff and host government evaluation personnel have also expanded over the decade, as have national evaluation networks and other evaluation interest organizations with worldwide membership, reflecting the increasingly international character of this field.

Also of note in the donor evaluation community, are shifts in the way evaluation offices relate to agency management. Some agencies have elevated, or are considering elevating, their evaluation offices so that they report either to the head of the agency or its board of directors, e.g., DFID and the World Bank, as a means of ensuring their independence. Further, at least one government, Sweden, has created a separate agency that conducts evaluations of program funded by SIDA, the country’s main organization for the delivery of foreign assistance. In contrast, France’s *Agence Française de Développement*, after establishing its evaluation office as a relatively independent entity, concluded that the evaluation unit was too isolated from the rest of the institution, and moved it into the organization’s management division.

The final section of this paper highlights options for USAID in each of the aspects of donor evaluation in which trends and innovative practices were found.

## **2. STUDY PURPOSE AND SCOPE**

Evaluation has been an integral part of the U.S. Agency for International Development (USAID) approach to managing foreign assistance from the beginning. Its first Office of Evaluation was established within the Policy and Program Coordination Bureau (PPC) in 1968. In the early 1980s, PPC consolidated evaluation and several related functions in a new Center for Development Information and Evaluation (CDIE). In

2006, as part of an effort to integrate the work of USAID and the Department of State, CDIE was dismantled, and responsibility for overseeing evaluation in USAID shifted to the Bureau of Foreign Assistance (F Bureau) in the Department of State. USAID is now in the process of reestablishing its own central evaluation office. This study was undertaken to support that effort.

Rather than simply recreating an earlier version of its Office of Evaluation, USAID is interested in defining the role and agenda for its new evaluation unit to be consistent with current needs. To this end, this study was undertaken to provide USAID with a portrait of the way in which international development evaluation is currently conceptualized and practiced. The mandate for this study was to scan the horizon and synthesize how international development evaluation theory, policy, and practice have changed over the past decade and formulate options for USAID that are consistent with current practice and USAID's own history.

### **3. METHODOLOGY**

No specific methodology was prescribed for this study. Initial expectations were that a mix of desk research and interviews would cover bilateral and multilateral development agencies and, as warranted, foundations, private voluntary organizations, and other entities working in international development. In practice, more attention was paid to the evaluation policies and practices of bilateral and multilateral donors, given their direct relevance for USAID, than to the internal practices of other types of international development organizations.

Document review was the primary data collection method used. Donor websites, for the most part, include up-to-date information about their evaluation policies and practices. Reports on several earlier efforts to describe the state-of-the-art in evaluation practice were also reviewed. Among them, an *Agence Française de Développement* comparative study on Evaluation Policies and Practices in Development Agencies (Foresti 2007), UNICEF's Working Paper on New Trends in Development Evaluation (Segone 2006), Robert Picciotto's article on International Trends in Development Evaluation (Picciotto 2003, 227-234), and Andrea Liverani and Hans Lundgren's article on Evaluation Systems in Development Aid Agencies analyzing the results of OECD Development Assistance Committee (DAC) peer reviews (Liverani 2007), were particularly useful.

To complement the study's broad review of evaluation trends, an analysis was undertaken of USAID evaluation trends from 2005 to 2008. This analysis extends trend lines that USAID has been tracking for about ten years and is included as Annex A.

Given the approach taken by this study, some things will have invariably been missed that might have come out of extended interviews or perhaps a written survey of donor organizations. Despite these possible losses, the study's examination of extensive documentation about the policies and practices of most donor organizations, as well as numerous articles in evaluation journals and on the Internet, provides a sound basis of evidence for the options suggested in this report.

### **4. SIGNIFICANT DEVELOPMENT EVALUATION TRENDS**

Unlike applied sciences, such as medicine, where theory, policy, and practice tend to move sequentially, trends in evaluation theory, policy, and practice exist in a common space and influence each other, but directionality varies.

## A. TRENDS IN EVALUATION THEORY

Evaluation theory has evolved slowly. While its impact on evaluation policy and practice has been significant, it has not necessarily been linear. With respect to the current decade, critiques of evaluation quality and utilization have played a greater role than has evaluation theory in fostering changes in policies and practices.

The evaluation of social programs, including international development programs, emerged as a distinct field in the late 1960s and early 1970s. This is when practitioners working in a variety of subjects, using techniques drawn from a wide range of disciplines, began to interact with each other. While evaluation is largely a practice-driven enterprise, it has over the years expanded its attention to questions about the relationship of evaluation to the social programs that constitute its focus. Early thinking, or theoretical constructs, about the causal process through which evaluations impact on social programs presumed that high quality evaluations would be used to improve social programs. The fact that this is not always true in practice prompted introspection and the evolution of evaluation theory.

As a collection of ideas, evaluation theory is multi-faceted. Most of the dimensions along which evaluation theory developed, and continues to evolve, were at least visible by the late 1970s. These efforts to explain or predict how evaluation interacts with social programs consider, among other things, (a) the nature of the social programs on which evaluations focus, e.g., whether they are sufficiently well defined to be evaluable, and how context may affect their performance as well as their evaluability; (b) options with respect to evaluation timing, evaluation focus (process versus outcomes), the role of the evaluator, and whether those involved with a program do or do not participate in evaluations; (c) evaluation evidence, i.e., what constitutes “good enough” evidence for decision-making, how knowledge is acquired including what methods are best suited for various evaluation tasks; (d) evaluative judgment, including cultural perspectives and bias; and (e) utilization, including types of use (immediate and cumulative), the role of institutional processes, and user interest and sense of ownership (Shadish 1991).

Over the past four decades, much of the thinking about these aspects of evaluation has been iterative, with new publications suggesting variations on previous approaches. This process has spawned new terminology that suggests a somewhat greater diversity of ideas than actually exists. Some issues have sparked debates within the evaluation community, e.g., the merits of theory-driven, outcome-oriented, hypothesis-testing evaluations, which often depend heavily on quantitative data, versus more responsive, process-oriented, and methodologically fluid, or evolutionary, evaluations. Other aspects of evaluation theory development have been supportive of consensus-building among a diversity of practitioners and institutions about the nature of the field and what constitutes good practice. Overall, the evolution of evaluation theory has fostered an acceptance of diversity as well as increased interest among practitioners in the use of a pragmatic mix of approaches and methods.

Trends in evaluation policy and practice draw on evaluation theory but their reach is just as likely to be to the entire body of evaluation theory as it is to be the most recent innovation. For example, 1980s advancements in thinking about the relationship between evaluation utilization and the earliest steps in the evaluation process (e.g., the role of intended users in defining the questions an evaluation would address) arguably influenced practice in a linear way. On the other hand, current interest across the evaluation community in enhancing utilization by improving evaluation rigor has not looked to new developments in evaluation theory, but rather doubled back to classic mid-1970 evaluation theory about gradations in the quality of evaluation evidence (i.e., from employing experimental versus quasi-experimental and other approaches).<sup>1</sup> Thus, while evaluation theory continues to evolve, and some new contributions are cited in later sections of this paper, no strong linear link was found by this study between the progression of evaluation theory over the past decade and obvious trends in evaluation policy and practice.

---

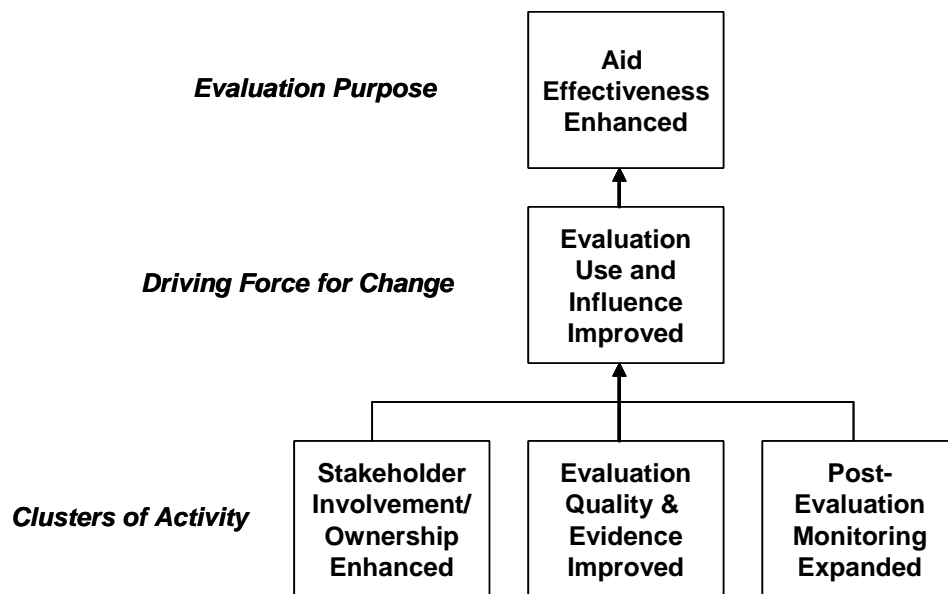
<sup>1</sup> Current initiatives aimed at improving the quality of evaluation evidence have also reached laterally to new developments in econometric methods, as discussed in later sections of this paper.



## B. TRENDS IN EVALUATION POLICY AND PRACTICE

Evaluation policies and practices in development assistance organizations respond to a range of stimuli, including political demands as well as new developments within the field of evaluation. Over the past decade, changes in donor agency evaluation policy and practice described in this paper appear to have been triggered by multiple forces. Broad international trends in the way development assistance is conceptualized, funded, and managed, as outlined in the text box below, constitutes one of these forces. Criticism of development assistance donor evaluation practices (e.g., inadequate evidence of impact, weak utilization of evaluation findings and recommendations) is another of these factors. For USAID, trends in the U.S. approach to the funding, delivery and management of foreign assistance -- the subject of a companion paper, entitled *Trends in U.S. Foreign Assistance Over the Past Decade* -- also play a role, as do trends in U.S. Government-wide guidance on performance reporting and evaluation quality.

Broadly speaking, shifts in the way evaluations are conducted in donor organizations over the past decade have a common goal: *enhancing aid effectiveness*. Moreover, donor explanations of the reasons for encouraging certain evaluation approaches recognize that evaluation utilization – whether immediately or through cumulative knowledge - is the path through which evaluation practice impacts on development effectiveness. Some of the trends below focus directly on improvements in evaluation utilization. As shown in the following text box, other trends in evaluation practice and policies are viewed as contributing to improved utilization. To that end, a number of trends in evaluation practice focus on improving evaluation ownership, enhancing evaluation quality and post-evaluation follow-up.



## **Illustrative Trends in Development Assistance that Affect Evaluation Policy and Practice**

### *Decentralization and Devolution of Development Management Responsibilities*

The decentralization of development management responsibilities in donor organizations is not new. USAID, for example, was to some degree “born decentralized” as a function of its field operations. In other donor organizations decentralization is a trend that has expanded more gradually with donors shifting responsibility for project management purposes including evaluation to the field or to decentralized technical offices. Decentralization has also involved a steady trend toward handing over key responsibilities to host nationals, including USAID Foreign Service Nationals (FSNs) who now serve as Mission Evaluation Officers.

Devolution, to tweak a term, has also been part of this process as demonstrated by the movement of a significant degree of responsibility for the design of programs and their performance monitoring and evaluation systems to implementing partners, including local partners in developing countries. Devolution in this sense also encompasses the transfer of development management responsibilities to host countries, many of which now have their own evaluation units and systems. While transfers of this sort are limited in the USAID portfolio, they have proliferated significantly in the portfolios of other bilateral donor agencies. The result of these shifts from an evaluation perspective is that the responsibility for conducting evaluations is now widely dispersed not only in USAID but throughout the donor community. As DAC peer reviews of donor evaluation systems have begun to point out, one result is that central evaluation offices in donor agencies no longer have total knowledge about the lessons from or control over the quality of evaluations being undertaken in their agencies.

### *Heightened Interest in Aid Effectiveness*

The 2002 International Conference on Financing Development in Monterrey, Mexico initiated a new dialogue between donors and developing countries aimed at creating more effective partnerships for development. The consensus of this meeting recognized the leadership role of developing countries with respect to their development plans, fostered donor support of those plans and simultaneously called for improvements in developing country governance. This consensus was expanded in 2005, through the Paris Agreement, which set forth an agenda for reforming aid delivery and management with the aim of improving its effectiveness. The principles of this agreement focused on host country ownership of the development process; improved harmonization of donor practices and their alignment with country-led strategies; better use of managing for results approaches; and mutual accountability for the use and results of aid.

Out of these interactions also has emerged a broad commitment to the Millennium Development Goals and an international monitoring system for “keeping score” that is accessible to donors and developing country citizens alike. These developments have also fostered a new level of interest in joint evaluations and developing country evaluation capacity building, as discussed below.

### *Harmonization or “Getting on the Same Page”*

Harmonization, broadly understood, aims at making it simpler for developing countries to succeed by consolidating and streamlining the donor footprint. On the ground, this is leading to more joint evaluations; an important trend discussed in this report. But more importantly, it has produced an explosion of efforts among donors to “get on the same page” with respect to a wide array of evaluation policies and practices. The pace of work in the DAC Network on Development Evaluation under the OECD’s Development Assistance Committee (DAC) has accelerated, bringing out a wave of new common guidance and standards. Bolstering the harmonization of evaluation work in the DAC are efforts by two newer umbrella organizations, the first of which is the Evaluation Coordination Group (ECG), established in 1996 by the multilateral development banks. The second organization is the United Nations Evaluation Group (UNEG). Set up in 2004 to supersede an earlier interagency working group on evaluation, UNEG is charged with the implementation of the evaluation aspects of the massive One UN initiative begun in 2006 (UNEG 2009b). With all three of these organizations at work and clearly aware of what the others are doing, progress on standards and norms has been rapid. Implicitly sharing the workload, these entities have confronted such issues as evaluation quality and what constitutes a complete and adequate evaluation report.

Many of the trends in evaluation practice that are visible today evolved from changes in evaluation practice that date back to the mid-1980s. In contrast, new policies and practices aimed at strengthening the evidence that evaluations produce concerning cause-and-effect relationships are relatively new in donor organizations,

though many of the methods advocated for doing this were well understood decades ago. Internal guidance recommending post-evaluation action monitoring is not new, but approaches for ensuring that it occurs are. Trends directed at each of these evaluation practice results are discussed further below.

## (1) EVALUATION USE AND INFLUENCE IMPROVED

Evaluation use or influence is the purpose cited by all development assistance organizations for undertaking evaluation; this has long been the discipline's Achilles' heel. USAID has always understood this, as the text box below from its very first Evaluation Handbook shows. While evaluation use is not discussed extensively in donor evaluation policies, or in guidelines issued by the Development Assistance Committee or the United Nations Evaluation Group, evaluation use and influence are perpetually under discussion in evaluation journals and at conferences where evaluators gather.

Over the past four decades, significant efforts have been made to unbundle what evaluators mean by utilization, and numerous prescriptions have been written about how to increase evaluation use. On the definitional side, there is fairly wide agreement on several potentially important types of evaluation utilization, building on ideas introduced by Carol Weiss (Weiss 1979), Michael Quinn Patton (Patton 2008), and others over the course of several decades (Bayley 2008). Uses include:

- conceptual or “enlightenment” use – improved understanding of how programs work
- instrumental use – direct use, e.g., to modify a program or design a new one
- symbolic use – as justification for actions taken or to be taken
- process use – gains made by those who participate in evaluations, regardless of the evaluation results

**Utilization at Center Stage**  
Program evaluation...is valuable only when its findings are applied. If used properly, evaluation findings should permit Mission management (and USAID/W) to materially improve the quality of its performance; if not, the work isn't worth the effort despite its historical interest.  
–USAID Evaluation Handbook 1970

In a 2000 address to the American Evaluation Association, Karen E. Kirkhart challenged these formulations as being too narrow and proposed instead an integrated theory of evaluation influence which, in turn, has spawned a literature of its own. Equally active is the focus on factors that influence evaluation use, which divide roughly into propositions that focus on the goodness of the evaluations themselves. These range from evaluator skills in communicating results and framing actionable recommendations to the environment (evaluation culture) into which reports are delivered and the level of engagement of the immediate client for an evaluation (Kirkhart 2000).

Despite this advanced thinking about evaluation use (Caracelli 2000), empirical research aimed at validating theory has, since the 1970s, resembled a stalled train. Around the same time, a flurry of studies of evaluation use in federal agencies were undertaken, including one by Robert Yin and Carol Weiss for USAID that examined evaluation use in a sample of 33 evaluations and found some level of instrumental use with half of these evaluations (Yin and Weiss 1988). For more recent utilization constructs, such as Kirkhart's integrated theory of influence, empirical data is even thinner. Exceptions in this regard, which have high relevance for the development community, are two utilization-focused activities undertaken by the World Bank.

- The first, the landmark report *Influential Evaluations*, identified eight cases of evaluation utilization and documented the way they were employed and the impact of the use of evaluation results (WB 2004). For several of these cases, the cost of the evaluation was compared, in dollar terms, to the benefit stream produced. While this study focused on only a few “best case” examples, it demonstrated that important questions about utilization and its consequences are in fact answerable.

- More recently, the World Bank instituted a new set of procedures for monitoring utilization across a large set of evaluations. The Bank has identified two types of evaluation use as performance objectives for its central evaluation unit and the data are being collected for two performance indicators, as shown in the text box below (IEG 2008, 95-105). In the narrative discussion of these data, the Bank’s evaluation unit credited efforts to improve the clarity and actionability of evaluation recommendations for improvements in the percentage of evaluation recommendations accepted by managers.

**Utilization as an Evaluation Performance Measure at the World Bank**

***Intermediate Outcome: Evaluation Recommendations Incorporated in to Bank Operations and Policy***

- 96 percent of recommendations accepted; comparable to prior years
- 95 percent adoption/implementation (of those accepted) up from 85 percent the prior year

***Intermediate Outcome: Use of Evaluation Findings by Bank Staff at the Operational Level***

- 72 percent of those surveyed reported using findings as a basis for advice or comments
- 54 percent reported using findings for designing new operations
- 45 percent report using findings for modifying ongoing operations

Annual Review of Development Effectiveness  
World Bank, 2008

Improvement in the monitoring evaluation utilization in the development community, if not more generally in the field of evaluation, is something of a double-edged sword. On the one hand, advances in utilization theory and methods make it clear that much more rigorous research in this area is possible. On the other hand, doing this kind of research might draw more attention to non-utilization, and its implicit cost, than evaluation offices in most donor organizations might be comfortable examining.

**(2) STAKEHOLDER INVOLVEMENT AND OWNERSHIP ENHANCED**

The notion of “ownership” or “buy-in” is central to much of what has been written about the use of evaluation. Individual and institutional agreement on the relevance of evaluation questions and acceptance of their findings are considered to be important precursors to action. When decision makers do not value what evaluations offer, implementation of their recommendations is unlikely. “Ownership” is viewed as being somewhat distinct from questions about the quality of evaluation products and the readiness of organizations to learn and change. Both of these are important topics, discussed separately below.

Generally speaking, approaches that are employed to increase the likelihood of evaluation use by enhancing a sense of ownership among intended users are participatory in nature. Involvement in the process of developing evaluation questions, gathering and analyzing data, working through what has been learned to reach conclusions, and formulating action recommendations, are steps that engage intended users. Participatory evaluation approaches are not new; they emerged as an alternative to approaches that relied exclusively on independent external evaluations in the 1970s. Where participatory approaches have made new inroads over this decade is on the development planning side. This shift is closely associated with the

release of a comprehensive framework by the World Bank in 1999, which set the stage for a significant reorientation in its approach to development planning. Some senior World Bank officers now take the time to consult with beneficiaries and ensure that they, and not simply their governments, are committed to realizing and sustaining intended project benefits.

Acceptance of requirements for more participation, or greater democratization of the development process, has expanded appreciation of the intent of participatory evaluation even if it has not dramatically expanded the number of evaluations that use these methods. While fully participatory evaluations are the exception rather than the norm in USAID and other donor organizations, donor policies continue to stress the importance of building a sense of ownership among evaluation users.

#### **a. Fostering Beneficiary/Local Ownership of Evaluation Results**

Closely associated with the work of Robert Chalmers in the development arena and J. Bradley Cousins in the field of education, participatory evaluation gained acceptance in development assistance donor agencies in the 1980s and 1990s as a mechanism for engaging local partners and program beneficiaries. From these roots grew strands in evaluation theory that focus on whose priorities and perceptions of reality should dominate in an evaluation (Guba 2001). Likewise, theories emerged around prescriptions for linking participatory evaluation to other empowerment techniques for the purpose of ensuring that evaluation experiences will yield longer-term, self-determination capacities among local participants (Empowerment Evaluation 2009). In the development arena, learning-by-participating evaluation techniques gained early popularity among non-governmental organizations that worked directly with communities and local partners. This is not an area where new trends tend to stand out.

#### **b. Fostering Decision-Maker/Stakeholder Ownership of Evaluation Results**

In contrast to a lull in new participatory initiatives at the beneficiary level, there has been considerable movement aimed at increasing the level of involvement of major program and project stakeholders in the evaluation process. Efforts to increase the number of joint evaluations involving multiple donors, as well as host country representatives, are setting the pace.

Stakeholder involvement in evaluations is closely related to aspects of evaluation theory that focus on decisions, including Michael Quinn Patton's 1979 *Utilization Focused Evaluation*. Efforts aimed at increasing stakeholder participation in evaluation often focus on decision-maker engagement in the process for deciding the questions that an evaluation will address and for determining the timing of evaluations. In UNDP's tripartite review, where UNDP is the primary funding organization, host government representatives and the implementing U.N. organization all participate. In some countries it has conducted its tripartite reviews in project settings (Office of Evaluation UNDP 2009). This U.N. approach provides a useful working model for joint evaluations.

Among European donors, DANIDA appears to lead in its commitment to joint evaluations, which often involve a good deal of negotiation at the TOR development stage and in the framing of conclusions and recommendations. DANIDA estimates that half of the evaluations it undertakes are now joint evaluations. Joint evaluations are gaining in popularity where multiple donors are working with a government on a set of development problems through a shared effort, or one that is funded from pooled resources. Joint evaluations are also consistent with donor and host government commitments to increasing "mutual accountability" (OECD 2008) and evaluation capacity building (Feinstein 2003). To facilitate high-quality joint evaluations, the DAC released new guidelines on this topic in 2006 (DAC Centre 2006).

Participatory evaluation methods have an accepted place in USAID's repertoire and are included in ADS 203, Figure 203B, as one of the types of evaluations USAID conducts. Collaborative or joint evaluations are shown as another recognized type of USAID evaluation in this same table, and both are encouraged by USAID in ADS 203.3.5.6. In addition, USAID has an evaluation TIPS document that helps guide evaluation managers through this type of evaluation process. In practice, participatory evaluations are still relatively rare.

Between 2005 and 2008, 22 percent of USAID evaluations sent to the DEC had participatory features. Of these, 82 percent involved stakeholders, while 48 percent involved beneficiaries.<sup>2</sup> Joint evaluations, while not specifically counted for in 2005-2008, are also relatively rare, though USAID has participated in some very large joint evaluations, such as the 2005 evaluation that examined World Food Program approaches in Bangladesh (Russo 2005).

While much has been written about participatory and collaborative evaluation from ownership, empowerment, and capacity-building perspectives, the literature is relatively thin when it comes to measuring the impact of these approaches. Some of the more insightful analytic looks at participatory processes have been undertaken by the World Bank (Estrella 1998), which is heavily invested in participatory approaches. JICA (JICA 2001) also examines the impact of participatory approaches through short-term case studies, rather than longitudinal or comparative efforts. Perhaps more honest and to the point is a look at participatory evaluation practices by Joseph E. Stiglitz, who simply concludes that it is probably the right thing to be doing, even if measures of impact are ephemeral (Stiglitz 2002).

### **(3) EVALUATION QUALITY AND EVIDENCE IMPROVED**

While research on evaluation utilization has long pointed to evaluation quality as a key determinant of use, it has not always been easy to operationally define what that means. Senior evaluation managers recognize quality, or the lack thereof, when they see it. But more than a sense of quality is needed to support systematic improvement efforts. Fortunately for the development assistance community, that job has been taken on by the DAC, which issued a draft version of a set of evaluation quality standards in 2006. These standards are open for a three-year trial period in donor organizations, following which they will be finalized (DAC Network 2006). This effort on the part of the DAC is both prescient and an acknowledgement of the myriad of evaluation quality improvement efforts already underway in the development community. This section examines a range of initiatives, which together represent a trend and highlight the importance the development community now attaches to evaluation quality. Relevant efforts focus on improving:

- Evaluator Competencies
- The Strength and Credibility of Evidence
- Adherence to Ethical Standards, and
- Evaluation Reports: Standards & Quality Assurance

Precisely which initiatives, or what combinations, will be most effective may not be known for some time.

#### **a. Evaluator Competencies**

The skills and experience of those who conduct evaluations have a direct impact on evaluation quality. While that proposition has long been treated as a given, it is only in the last decade that significant attention has been paid to enhancing the competencies of evaluation professionals and evaluation staff in development assistance organizations, in host governments, and among those in developing countries who have an interest and the requisite skills to put them on the path towards a career in evaluation. In this arena, the last fifteen years has seen a whirlwind of activity on the issue of evaluation competencies, as well as a rapid internationalization of the community of evaluation practitioners (IDEAS 2009).

As a craft, program evaluation has its deepest roots in the United States. While some U.S. Government agencies can trace their evaluation work back to the early 1960s or earlier, as in the field of education, the first real charge to federal agencies in this regard came in a 1970 memorandum from then-President Nixon, which stated that evaluation is one of the most important responsibilities of federal agencies. This memorandum further stated that the kind of program evaluation needed would be dependent on whether the objectives of

---

<sup>2</sup> See Annex A.

programs remain; whether programs are effective in meeting their objectives; and whether alternative approaches would produce “the same or greater benefits at the same or lesser costs” (American Presidency 1970). Government-wide response to this directive quickly demonstrated how little was known about the characteristics of a competent evaluator. Over the next three decades, academics and consultants from diverse backgrounds contributed their labor and their ideas. They generated much of the early literature, created the earliest evaluation research networks, which eventually grew into the American Evaluation Association, and developing courses and, eventually, degree programs in evaluation (Hogan 2007). By the late 1990s, evaluators themselves began to ask questions about evaluator competencies that led to parallel studies in the U.S. and Canada (Huse 2006) aimed at defining evaluator competencies clearly, and yet broadly enough to engage practitioners there and around the world in serious discussions about the practicality and need for credentialing or certification of practitioners in this field.

Over the early part of this period, initiatives beyond North America, starting with the emergence of the demand for professional evaluators in Europe and Australia, brought attention to evaluation in the Development Assistance Committee of the OECD, as well as in the multilateral banks and the United Nations. CIDA and GTZ emerged as early leaders along with USAID in the institutionalization of evaluation practices and their integration into program and project design. Similar in concept, these earlier efforts helped foster donor collaboration in the DAC. By the end of the 1980s, interest in evaluation and learning how to do it well had spread to the developing countries. Evaluation offices emerged in government agencies and the first local evaluation networks emerged. In Africa, various networks grew into the continent-wide African Evaluation Association (AfrEA), which early-on took the initiative to examine and redraft evaluation standards crafted in the U.S. for a culturally diverse world. The International Program Evaluators Network, with its roots in Russia, now serves nearly twenty countries. USAID, other donors, partner organizations, and local governments found themselves turning to citizens of developing countries to manage monitoring and evaluation portfolios. These trends highlighted the need to invest in evaluation capacity. Worldwide, new initiatives started in several places simultaneously, as a 1996 DAC review of donor capacity building activities illustrated (Schaumburg-Muller 1996). A decade later, regional UNDP workshops held in 2007 gathered recommendations for enhancing and reinforcing a range of evaluation capacity-building initiatives (UNDP 2007), including:

- Training as a capacity building mechanism emerged as a key donor intervention. It took several forms, including a World Bank funded summer course in evaluation in Canada for senior government personnel; the translation of its on-line curriculum into French by the African Development Bank; a separate USAID-funded in-country evaluation course delivered primarily to local USAID and partner staff in a number of countries to the recent opening, with support from Japan, of an Asian venue for a course that parallels the World Bank’s Canada-based program. Complementing these longer term courses were a myriad of short-term evaluation trainings focused on specific sectors such as health or offered for and by the private voluntary organization community.
- Evaluation networks emerged as a mechanism for developing a critical mass of evaluators at the country and regional level. Networks have received support from a variety of donor and evaluation community organizations, ranging from the Kellogg Foundation -- which brings heads of evaluation networks together and spawned the development of the International Organization for Cooperation in Evaluation (IOCE) -- to UNICEF, which found ways to support evaluation network development on several continents and produced a volume of case studies that is helping these networks multiply and mature (Segone 2009). Japan is also contributing to these efforts, particularly in Asia. Complementing efforts to grow local evaluation networks, the World Bank has funded the International Development Evaluation Association (IDEAS), a website and membership network that is helping to develop a community of practice that links together international development evaluators from all continents.

- Multilateral development banks are also providing organizational assistance in monitoring and evaluation to host governments. The World Bank has a special program through which it helps countries strengthen their evaluation systems, and has produced publications that summarize these experiences for other governments that are seeking to strengthen their evaluation capacity and systems (WBG 2009b). Similar capacity-building assistance is being provided in Latin America through the Inter-American Development Bank (IADB 2009).

The speed with which interest in evaluation has become global has brought the question of evaluation competency back to center stage in a wide range of donor organizations. Evaluation offices in development assistance agencies are well aware of the discussions happening in literature, in evaluation associations, and at development and evaluation conferences<sup>3</sup> about the need to improve evaluation performance and the importance of enhancing evaluation competency.

Inside donor agencies, some new evaluation initiatives focus on evaluation methods while others focus squarely on the skills of evaluation personnel. Notably, some of the newer quasi-experimental evaluation designs require evaluators with econometric skills. Among donor entities, the United Nations Evaluation Group, through its Evaluation Quality Enhancement Task Force, appears to be the furthest along in operationalizing what evaluation competencies are needed in donor organizations. Recent donor agency initiatives focused on evaluation competencies are summarized below:

- **U.N. System** -- In 2007 the UNEG approved a set of core competencies for evaluators at all levels throughout the U.N. system. (UNEG 2009a) This step is an outgrowth of work initiated in 2005 by the Quality Enhancement Task Force of the UNEG. According to the UNEG, adoption of a detailed set of evaluation competencies was undertaken for the purpose of professionalizing and harmonizing evaluation management and practices within the United Nations.

### Meeting the Evaluation Competency Challenge

Improving the professionalism of evaluation by creating clear expected competencies for evaluators will foster credibility of the evaluation function and evaluation staff. As management needs for evaluative evidence for decision making become more complex, critics point to the lack of consistency and methodological rigor within the evaluation community. Much of the criticism comes from within the evaluation community itself. A lack of core competencies has led to a situation where “evaluation” is conducted by a variety of individuals with a social science or statistics background and other disciplines, resulting in products of uneven quality and credibility. Evaluation involves the rigorous application of research methods, statistical tools, analytical techniques and listening skills. Also critical are sound judgment and effective communication skills. There is a consensus that action is required to create professional competency criteria for evaluators. There is as well a need for a common framework for professional development based on the identification of core competencies and a structured training approach. There is a clear message to the evaluation community that the time has come to find means to ensure that evaluators somehow meet minimum professional standards. A major challenge relates to the absence of internationally recognized competency criteria or a governing body to support enforcement.

–UNEG commentary on the adoption of a set of core competencies for U.N. evaluation personnel, 2007

<sup>3</sup> Evaluation competencies were, for example, the focus of several sessions of the 2005 “Crossing Borders, Crossing Boundaries” joint international meeting of the American Evaluation Association and the Canadian Evaluation Society.



- **CIDA** -- A similar process was recently initiated in Canada, which expects to have a procedure in place for certifying the competency of heads of evaluation units across government by 2010. (Amira 2007)
- **DFID** – An evaluation competency initiative is embedded in the new evaluation policy for DFID, which is currently posted on the Internet for comment. (DFID 2009a) DFID’s draft policy commits the organization to defining the professional competencies and skills expected of those involved in managing evaluations - and for specialist evaluators who are engaged to conduct them.

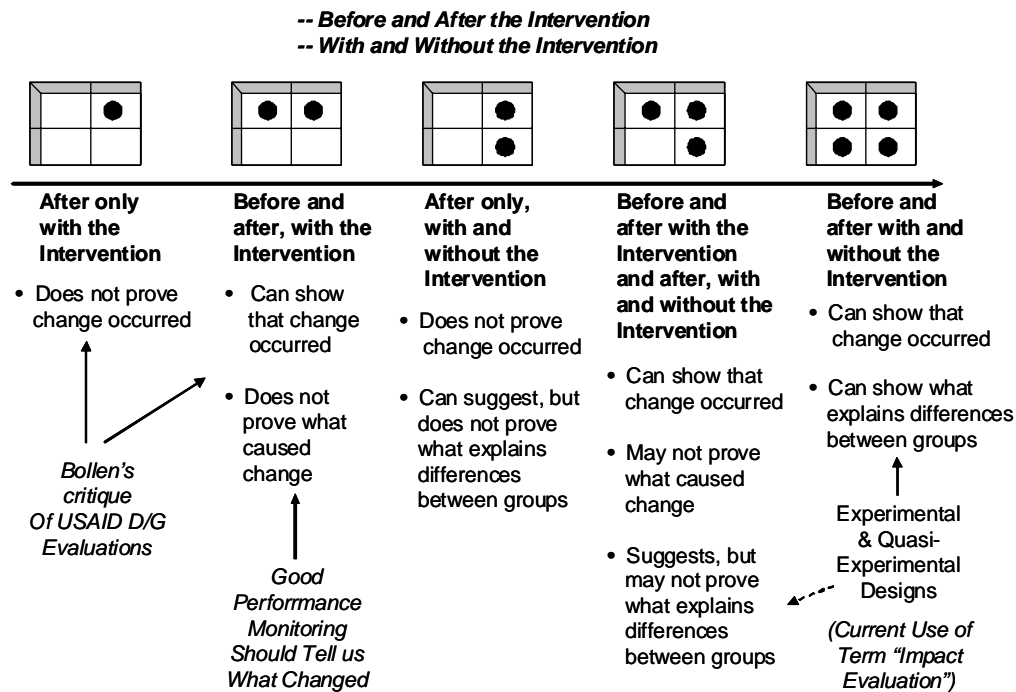
#### **b. The Strength and Credibility of Evidence**

Those who work with evaluations produced by development assistance organizations have long viewed them as being at best spotty in terms of the strength and credibility of the evidence. Even so, serious questions about this issue were not posed in a public way until 2004 to 2006, when a number of key critiques of evaluations and the evidence they present emerged. Among donor agencies, perhaps it was USAID that felt the brunt of this criticism most strongly, as evidenced by the following examples:

- In 2004, along with guidance for the annual OMB agency performance assessment exercise (PART), OMB issued a “white paper” on its website entitled *What Constitutes Strong Evidence of Program Effectiveness?* In this paper, which was a “shot across the bow” for all U.S. Government Agencies, OMB pointed out that most PART evaluations did not really provide information on whether change had occurred or whether federal programs caused those changes. OMB’s presentation of its argument for stronger evaluation evidence focused on the need for evidence about the counterfactual, i.e., What would have happened in a given situation absent the U.S. Government program intervention? The report concluded that “...where it is feasible to measure impact of the program, randomized controlled trials (RCTs) are generally the highest quality, unbiased evaluation [method].” (CEG 2009)
- One year later, in 2005, USAID again faced criticism, this time for the absence of strong evidence of results and causes in its evaluations of its democracy and governance initiatives. An article published in the *American Journal of Evaluation* reviewed 25 USAID evaluations in the democracy and governance field and concluded that, as a group, they lacked information that is critical to demonstrating the results of USAID projects, let alone whether these projects were the real causes of whatever change the evaluations reported. (Bollen 2005)
- In 2006, the results of a two-year examination were produced on why rigorous impact evaluations of social development programs are relatively rare in the international development arena, and what could be done to stimulate the conduct of more and better impact evaluations. The examination was performed by the Center for Global Development, an independent nonprofit policy research organization founded in 2001, with funding from The Bill and Melinda Gates Foundation and The William and Flora Hewlett Foundation. This report, entitled *When Will We Ever Learn? Improving Lives Through Impact Evaluation*, found that the impetus to learn about impact existed in a number of bilateral and multilateral development agencies, as evidenced by the existence of evaluations and other reports that purported to assess program and project impact. Where those studies failed, however, was in their ability to mobilize adequate rigorous evidence to answer the questions they posed. A range of approaches for stimulating the use of more rigorous impact evaluation methods were also suggested.

Structurally, what this collection of critiques demonstrated was weakness at the rigorous end of the evaluation continuum and unwarranted reliance on evaluations which, no matter what they recommended, had failed to convincingly establish that something changed, let alone what caused it.

## Evaluation -- Strength of Evidence Continuum



While all of these studies recommended that this weakness be addressed, none of them clearly defined the circumstances in which experimental designs and other highly-rigorous approaches were most needed. Neither did the studies define when it might be sufficient for a program or project to use less rigorous, though still responsible, methods of demonstrating that change occurred and eliminating most, if not all, other likely causes of that change.

### (1) Investing in Rigorous Evaluation Methods

Rather than going unnoticed, each of the reports cited above stimulated enough reaction to set off a brushfire of conversation and trigger a new wave of investment in remedying the evaluation gap they collectively identified.

- Responses to the OMB paper came quickly from U.S. Government agencies interested in making sure that RTCs would be treated as one way, but not the only way, to pursue program evaluation. Echoing this view, the American Evaluation Association (AEA), which had never before directly engaged government on policy issues, wrote to OMB defending the proposition that a continuum of valid evaluation methodologies exists and no single approach is right for every program evaluation situation. This set off another chain reaction, this time within the evaluation community, in defense of a range of evaluation approaches while, at the same time, stimulating a new wave of interest in RTCs. For example, the European Evaluation Society issued a parallel statement in 2007 indicating that it supports multi-method approaches to evaluation and does not consider any one method to be the “first choice” in every case (EES 2007).
- Hit hard by the criticism of its democracy and governance project evaluations, USAID’s Office of Democracy and Governance turned to the National Research Council of the National Academy of Sciences for an examination of whether evaluation methods could be introduced in this field and the degree of rigor with which they could be performed. That study not only examined the issues, but it

actively engaged with a number of USAID Missions in efforts to design rigorous experimental and quasi-experimental evaluations to test the feasibility of these methods for this type of USAID program. The 2008 report from that effort, *Improving Democracy Assistance: Building Knowledge through Evaluations and Research* (NRC 2008), stimulated sufficient interest in USAID to result in further efforts to introduce and demonstrate the efficacy of these approaches.

- Well-financed in its exploration of the opportunities for more rigorous evaluation in the international development arena, CGD launched its evaluation gap report in Europe and sponsored a conference in Bellagio, Italy, to examine issues and opportunities, inviting a wide range of bilateral and multilateral development assistance donors and their implementing partners to attend. Using the conference as a base for action, CGD sought and received commitments from participating organizations to take action in their own organizations to upgrade impact evaluation as a key element of an overall strategy for enhancing aid effectiveness. Two years hence it is clear that a number of Bellagio participants took that commitment seriously. New policy papers and initiatives that call for more rigorous impact evaluations have been produced in more than half a dozen development assistance organizations including DFID, *Agence Française de Développement* and UNIFEM, among others, over a very short time span.

Helping to solidify a new momentum behind rigorous impact evaluations are well-qualified research institutions that have already conducted a significant number of that type of evaluation overseas. Such institutions include the Jamal Poverty Action Lab at the Massachusetts Institute of Technology (MIT) and Innovations in Poverty Action (IPA), a New Haven, Connecticut-based organization similar to Jamal, with links to Yale.

In addition, new entities have grown up in response to interest in more rigorous methods. The first of these new entities was the Network of Networks for Impact Evaluation (NONIE), a collaboration that includes the DAC Evaluation Group, the UNEG, the ECG and the Kellogg Foundation, which also supported the International Organization for Cooperation in Evaluation (IOCE), an organization of presidents of evaluation associations worldwide. A second and more recent entrant on this stage is the International Initiative for Impact Evaluation (3ie). An outgrowth of CDG's evaluation gap initiative, 3ie receives support from the Gates and Hewlett Foundations as well as from development assistance organizations such as CIDA, DFID, NORAD and Care; the Governments of Mexico and Uganda; Google, Inc.; and a host of developing country research entities that have joined as members. Collectively, these organizations are raising awareness and investing not only in rigorous evaluations but also in developing the capacity needed worldwide to further expand this type of work. Illustrative of the scale of their combined efforts is the March 2009 international conference and set of associated training workshops in Cairo, Egypt, sponsored by NONIE, 3ie, and AFREA, the Africa-wide evaluation network. Notably, parallel domestic efforts are also resulting in a reexamination of how best to integrate rigorous experimental and quasi-experimental evaluations into federal evaluation practice in other parts of the U.S. Government (Simonson 2005).

What is striking about these initiatives, besides the energy that is being invested in them, is that very little of what is being introduced is actually new. With a few exceptions that reflect new work in the fields of statistics and econometrics, the core techniques for impact evaluation were well established within a few years of the emergence of evaluation as a program management tool in the U.S. federal government. *Experimental and Quasi-Experimental Designs for Research*, the bedrock volume by Campbell and Stanley, was published in 1963 and the even more detailed Cook and Campbell collaboration, *Quasi-Experimentation: Design and Analysis Issues for Field Settings* came out in 1979 (revised and updated by Cook and Shadish in 2003). These works were followed by Lawrence B. Mohr's *Impact Analysis for Program Evaluation* in 1988, and a myriad of supporting books and articles that elaborate on and expand the range of applications of these techniques.

Newer quasi-experimental methods aim to overcome problems that arise when comparison groups are selected too casually. Propensity score matching, which measures rather than guesstimates the closeness of treatment and comparison groups to each other on relevant variables, reduces the chances of selection bias in

quasi-experimental design. This technique was central to the methodology used in a multi-country World Bank study of the impact of social funds to compare the status of communities that did and did not implement social fund projects. The study employed several of the newer econometric models in use today in rigorous impact evaluations. These include, among others:

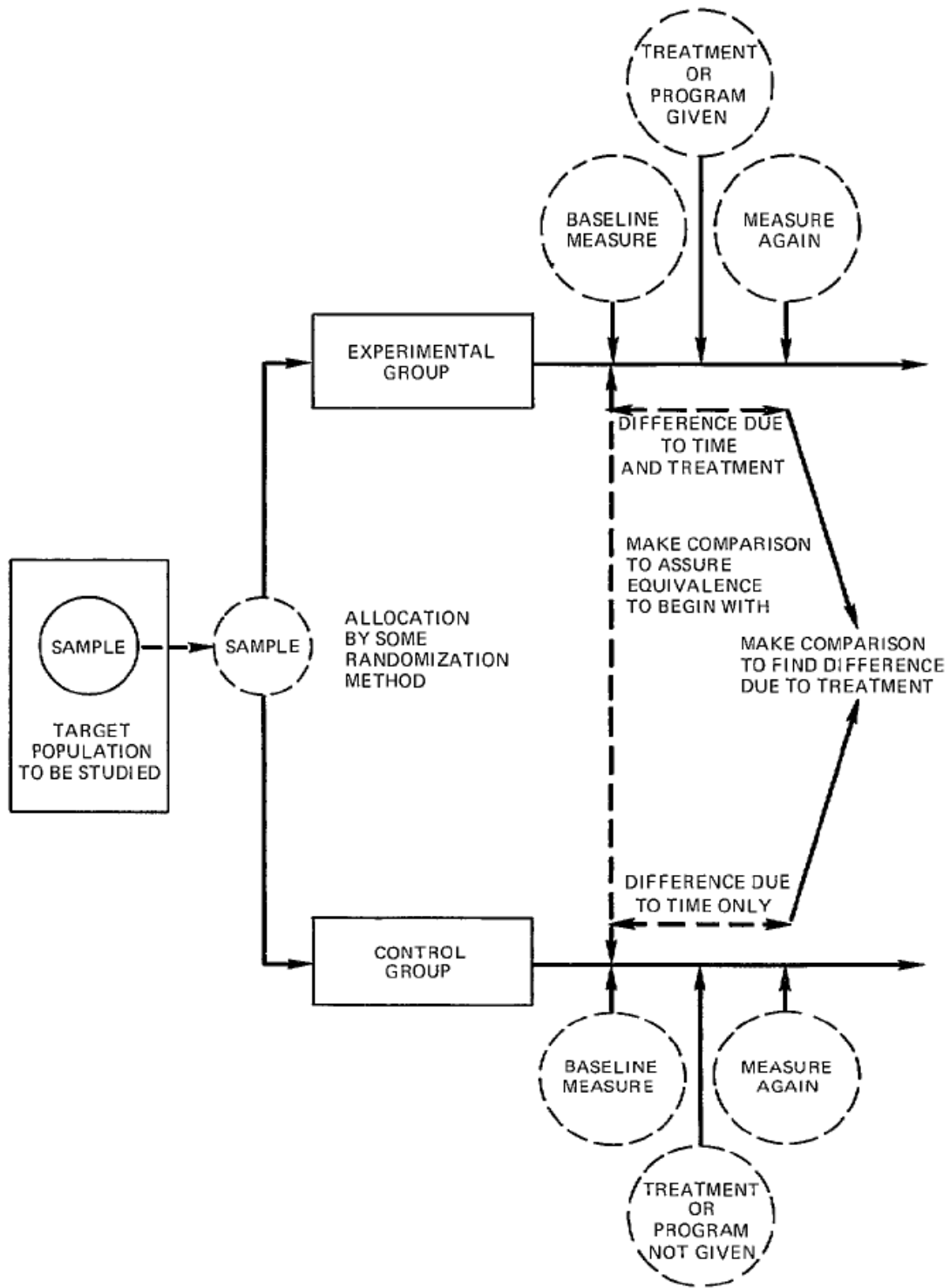
- *regression-discontinuity* designs, of the type used to measure Head Start program results in the U.S., i.e., the difference a program makes among individuals below a cut-off line, which in the case of Head Start is based on income
- *difference in differences* designs – another econometric method which is used to compare the before and after status of those who received a program intervention to those in a control group; this does not involve a field experiment, e.g., to examine the employment impact of minimum wage changes in one state to employment changes in another state where no wage change occurred

Striking, too, is how well grounded USAID was in the logic of experimental and quasi-experimental design when it built its evaluation practice. As the figure below, from the first edition of USAID's Evaluation Handbook in 1970, shows, randomized assignment to intervention (treatment) and control groups was well understood. It was also viewed as being "the ideal study" for making comparisons and determining "whether results are clearly attributable to one or the other of the factors involved." (USAID 1970, 40)

Given USAID's early recognition of the suitability of experimental designs for producing strong evaluation evidence, "What happened?" is a natural question to ask when looking for guidance on the use of these kinds of designs in USAID's ADS or its associated TIPS. The answer seems to be two fold. In terms of its evaluation guidance, USAID's interests have shifted over the years and documents that discuss evaluation reflect those shifts. By the late 1970s, the diagram shown below had been dropped from editions of the USAID Evaluation Handbook, making space for information about other types of techniques such as community based interviews and other rapid appraisal methods, reflecting a wider interest throughout the development community in quick turn around, cost-effective evaluation tools.

# “IDEAL” STUDY DESIGN FOR MAKING COMPARISONS

(USAID 1970, 41)



By 1980, recognizing the need for information about what development projects leave behind, USAID introduced a new Impact Evaluation series that involved short duration ex-post assessments of the impact of USAID projects and programs, rather than rigorous field trials. This pattern of moving some evaluation approaches and techniques off the “front burner” in USAID evaluation guidance to make room for new information about new techniques continued in the 1990s. The Agency began to emphasize performance monitoring as a mechanism for garnering information on progress and results earlier and more consistently than evaluations tended to provide.

In terms of what USAID actually does to examine the impact of its interventions, rigorous experimental and quasi-experimental studies have long been part of the spectrum of evaluation and research in which the Agency engaged. USAID, to the degree it used these kinds of rigorous tools, did not stop using these sorts of designs in the 1970s, but it may have stopped thinking of them as evaluations, relegating them to the realm of research instead. It is also fair to say that in the 1970s, as in the 2000s, these types of studies represent a small fraction of the evaluation and research efforts USAID undertakes. A review performed as part of this study analyzed evaluations in USAID’s DEC for the years 2005 to 2008, identifying nine evaluations that used rigorous designs of the type that today’s proponents of impact evaluation would accept as reasonably strong sources of evidence. Of these, one study used an experimental design involving randomized assignment while the others were quasi-experimental in nature. Also of note in this regard, an Internet search using the terms “USAID” and “experimental design,” “RCT” and/or “randomized assignment” yielded nearly a dozen experimental design studies that were not found among the reports key-word coded “evaluation” in the DEC.

Trying to learn from experience, the World Bank has inventoried all of its rigorous evaluations and made them electronically searchable by country and type of methodology employed, e.g., regression discontinuity designs, experimental design, and difference in differences (WBG 2009b). In comparison, USAID has done little to document and categorize those rigorous impact evaluations it has conducted over the years, beyond a recent effort focusing on health program evaluations and obstacles to more frequent use of experimental and quasi-experimental designs in this field (Teller 2008, 1-12). Nevertheless, it is important to understand that these types of studies have always been a small part of the USAID evaluation/research portfolio and many of the most recent ones can be found by scanning the Internet. A search result would include a randomized controlled trial design in a study in Bangladesh that was recently published in *Lancet* (USAID 2008b). Not only has USAID continued to use rigorous evaluation designs in a modest way over the years in agriculture, health and education, but it has also, on a decentralized basis, initiated efforts to apply these methods in new fields. Such new fields include micro-enterprise, through its Private Sector Development (PSD) Impact Assessment Initiative under the Accelerated Microenterprise Advancement Project (MicroLinks 2009), and in the field of democracy and governance, as noted above. While some development agencies, including the Inter-American Development Bank (IADB 2009) and the Millennium Challenge Corporation (MCC 2009) currently have much more extensive and well organized portfolios of rigorous impact evaluation underway, USAID is not without skills and experience in this field. USAID has a number of on-going studies on which it can draw, should it elect to examine more closely what is a very decentralized and unlinked set of activities.

## ***(2) Drawing a Bottom Line Across Many Evaluations***

Even when development assistance organizations have strong evidence about whether projects were effectively implemented or made a difference in people’s lives, some donor organizations have determined that their staff does not always have time to absorb that information. To address this problem at least two donor organizations have developed what might be called “at a glance” systems for scoring evaluation results using evidence concerning project performance.

- **SIDA.** Since January 2004, all SIDA efforts involving contributions above the value of 3.0 million SEK and with an agreement period above 24 months have been reviewed and rated with a SIDA Rating System (SiRS), which draws on monitoring and well as evaluative information. SiRS is a simple method that gathers information using a Logical Framework-based approach that follows and scores performance in terms of (a) immediate results (outputs) produced against plans, (b) overall

performance of the project/programme in relation to project purpose, and (c) the risks (project specific and external factors) that it may not achieve its purpose. Key dimensions include relevance, effectiveness, and sustainability. SiRS is an internal SIDA tool for promoting a focus on performance and results, and contributing to a more qualified and result-based dialogue with the project/programme partner. Ratings are done at least once a year and rating scores together with brief comments are registered in a special database. The system does not record specific results obtained (OECD 2004).

- **World Bank.** In the World Bank, when the need for summative information on performance was judged to be acute, the Bank developed a rating system for its Country Assistance Evaluations (CAEs) on country programs and rated them from “satisfactory” to “unsatisfactory” on a six-point scale. Because they communicate information rapidly to novices as well as evaluation practitioners, the World Bank reports that CAE scores have an observable impact on behavior at the country office level (IEG 2009a).

The value of systems of this sort is obvious. Their credibility depends not only on the strength of the underlying evidence but also on inter-rater reliability, which is the key to ensuring that specific ratings mean the same thing across countries and over time. Written explanations of both the SIDA and World Bank systems for producing evaluation bottom lines across many programs are lacking. They are both short on information regarding how questions of the quality of evidence feed into these rating systems and their procedures for ensuring inter-rater reliability. Nonetheless, these approaches represent interesting efforts and may warrant a comparison, in time and credibility terms, to USAID’s approach to collection; aggregating and certifying the quality of comparable data on standard performance indicators from all posts.

### **c. Adherence to Ethical Standards**

Ethical standards represent an important but largely passive element in current donor organization and evaluator calculations about how evaluation credibility can be improved. The only observable change over the years since 1974 -- when program evaluation standards including a set of ethical (propriety) standards were introduced by the American/Canadian Joint Committee on Standards for Educational Evaluation -- is the spread of their acceptance.

These standards have been incorporated in whole or in part into the DAC Principles for Evaluation of Development Assistance (OECD 2001) and the Norms and Standards issued by the UNEG. Norms set by these entities are subscribed to by nearly all bilateral and multilateral development assistance organizations. Parallel efforts in Canada produced a separate set of guidelines for ethical conduct for evaluators (Canadian Evaluation Society 2009). The reach of all of these standards also encompasses hundreds of evaluators by virtue of their membership in dozens of evaluation associations and networks that also subscribe to these standards.

### **d. Evaluation Reports: Guidance, Standards and Quality Assurance**

Evaluation reports play a critical role in transmitting the knowledge gained through an evaluation to its intended users. This section looks at three types of efforts that focus on improving the quality of evaluation reports. Steps taken to improve reports cannot compensate for poor evaluation design decisions or inadequate data collection and analysis, but they can improve user receptivity to evaluation recommendations by making clear the evidence, reasoning and judgments that produced them.

#### **(1) Evaluation Report Guidance**

Guidance on the preparation of evaluation reports is not new. USAID’s initial brief guidance on the content of evaluation reports was published in 1970, in the first edition of its Evaluation Handbook. USAID’s evaluation guidance is flexible on the structure of an evaluation report, but suggests in ADS 203.3.6.6 that at minimum evaluation reports should cover the evaluation scope, methods, findings, conclusions,

recommendations, and lessons learned for the design of future programs. Guidance on evaluation report preparation issued by other donor organizations varies with respect to coverage and on the degree to which they structure the contents of an evaluation report. Thus, for example:

- **CIDA's** evaluation report preparation guidance is reasonably typical of what development assistance organizations issue. CIDA does not require a standard evaluation report format; however, its Evaluation Guide describes a logical structure for a report and a detailed description of what each element would cover. It also provides a set of tips for writing evaluation reports and a checklist that teams can use when they are done (CIDA 2008).
- **DFID's** evaluation reporting guidance goes beyond CIDA's by characterizing reporting as a continuous requirement throughout the evaluation process. With respect to evaluation reports, DFID provides a two-page sample outline of what it expects evaluation reports to cover (DFID 2005).
- **UNDP's** evaluation report guidance also contains a sample outline. It includes an extended description of the quality criteria (utility, credibility, and relevance) on which evaluation reports submitted to UNDP are to be judged (UNDP 2009b).
- **AusAID's** evaluation guidance (AusAID 2005) and that of the **European Commission** (EuropeAid 2005) provide some of the most detailed and standardized sample outlines for preparing an evaluation report.

USAID has developed guidance on draft evaluation reports that goes beyond the minimum coverage requirement outlined in the ADS, but it has not yet been formally issued by the Agency nor is it referenced in ADS 203. This guidance document, *Constructing an Evaluation Report* (USAID 2006), was completed for USAID in 2006 under a contract that developed several new drafts for the USAID TIPS series. It provides a sample outline for an evaluation and guides the user through the preparation of each suggested report section. It includes report elements identified in ADS 203.3.6.6, and goes beyond to recommend that evaluations also include, among other things, information about the scale of the project/program being evaluated, in human and financial terms; the locations in which it was implemented; the program theory underlying the effort; annexes that include copies of the evaluation team's data collection instruments; and a description of study limitations in the methodology section.

## **(2) Evaluation Report Standards**

While most development assistance organizations subscribe to some set of evaluation standards to guide their evaluation process, it is rare for these organizations to have a separate set of formal standards that focuses exclusively on the contents of evaluation reports. In 2004, UNICEF issued just such a document, which may be the only one of its kind in the development assistance community (UNICEF 2004). UNICEF's Evaluation Report Standards includes 22 separate elements and for each it explains why the standard exists and what assessment process is to be used to determine whether any given evaluation meets that standard. One of the key steps UNICEF Evaluation Office has taken to help ensure that its Evaluation Report Standards are followed by evaluation teams is a requirement that these standards be explicitly identified in the Terms of Reference for every evaluation funded by the organization, whether in the field or through its headquarters. Direct observation of RFPs for evaluation issued by UNICEF indicates this is being done.

While USAID does not have a separate document that states its evaluation report standards, two sections of the ADS provide pertinent guidance in this regard that could, along the lines being pursued by UNICEF, be required references in USAID evaluation Statements of Work:

- **USAID ADS 203.3.6.6** guidance currently suggests the inclusion of an Executive Summary that accurately mirrors the report and, within the report, separate treatment of the evaluation scope and



methodology, findings, conclusions, recommendations for management, and lessons learned for the design of future programs. While ADS 203 explicitly requires that data quality be certified for performance monitoring data, it is silent on the data an evaluation collects. ADS 578, which deals with USAID information standards is more explicit in this regard, specifically:

- **USAID ADS 578.3.4.2** identifies types of documents that USAID routinely makes available to the public, including those disseminated through its Development Experience Clearinghouse (DEC), to which Agency information quality standards apply. These include documents that describe program and development activity performance and results, evaluation reports, press releases, employment information, and procurement and assistance opportunities. Documents in this group must comply with three Agency information quality standards: objectivity, utility, and integrity. In this definition objectivity has two dimensions: (a) substance, which states that the information included must itself be accurate (valid), reliable (collected using replicable methods), and unbiased; and (b) presentation, requiring that documents present information in an accurate, clear, complete, and unbiased manner.

### **(3) Quality Assurance**

Statements about the importance of evaluation quality and the need for managers to review evaluation products from a quality perspective before accepting them are found in the published evaluation policies of a number of development assistance organizations. Several of the agencies examined have gone beyond exhortations in this regard to develop tools and procedures for management reviews of evaluation quality. In some donor organizations these tools are applied at the level of an individual evaluation, as it is being reviewed prior to finalization. Other organizations use these kinds of tools for institutional level meta-evaluations that help them understand quality patterns over time.

#### **(a) Tools for Reviewing the Quality of Individual Evaluations**

There is no standard technique across agencies for reviewing the quality of evaluations as they are submitted. Several donor organizations, however, have developed techniques for doing this.

- **The European Commission** has developed a two-part evaluation quality rating system that it is using for all evaluations. The first part is to be used when reviewing the product of the inception phase of an EC evaluation. The second instrument, which is a matrix that considers eight quality constructs, each of which has several dimensions, is used to review a draft evaluation report. At that point in the evaluation process, every feature of an evaluation report is scored using a five point rating scale, as showing in a web-publication that describes this instrument (EuropeAid 2009). The results help guide the finalization of the report.
- **Agence Française de Développement** has a management review format for examining evaluation quality that includes a four point scale used to rate five aspects of an evaluation report on which its quality assurance process focuses (AFD 2009a).
- **CIDA**, while having its own evaluation policy, is also guided by government-wide evaluation policy, such as the evaluation quality review guidance released by the Centre of Excellence for Evaluation of the Treasury Board of Canada Secretariat. This guidance defines both a process and a set of criteria for reviewing every Canadian Government-financed evaluation and providing feedback to those who prepare them (CEE 2004).

Starting with the same set of concerns but building on domestic rather than international experience, U.S. Government evaluation offices such as the one at the Department of Energy, are also exploring new quality assurance approaches. One such approach adds to the range already explored by the development community by identifying questions that should be raised about evaluation quality early in the process, creating an opportunity to make improvements before, rather than after, data are collected. This identifies

specific quality assurance questions to ask when reviewing an evaluation Statement of Work, a contractor's detailed evaluation plan, and proposed evaluation instruments, as well as the draft evaluation report (DOE 2009).

USAID does not currently have any official procedures or tools for rating evaluation quality. Nevertheless, USAID staff participating in the Certificate Program in Evaluation use an evaluation review score sheet that is pegged to ADS 203.3.6.6 and USAID's draft TIPS on *Constructing An Evaluation Report*. Stylistically, this tool, which was developed around 2000, is similar to the one now found on the European Commission's website cited above.

### **(b) Institutional Level Meta-Evaluations**

Among the development agencies examined for trends in policy and practice, agencies within the U. N. Evaluation Group (UNEG) appear to have made significant progress this decade in establishing meta-evaluations as a quality assurance mechanism, as have some bilateral donor organizations including the New Zealand Agency for International Development (NZAID). As discussed below, USAID also has experience with meta-evaluation as a mechanism for monitoring evaluation quality, but it has not carried out any meta-evaluations for this purpose during the current decade.

The experiences of UNICEF, which developed its own Evaluation Report Standards and a linked meta-evaluation report rating system, and UNDP which rates evaluations against the U.N. Evaluation Standards issued by the UNEG, and NZAID provide useful illustrations of current quality assurance meta-evaluation practices.

- **UNICEF** -- In 2007, UNICEF conducted a meta-evaluation of evaluations carried out between 2002 and 2006, all but a fraction of which were undertaken by UNICEF country offices.<sup>4</sup> This was UNICEF's second quality review. The first was undertaken in 1995, covering evaluations completed in 1992 and 1993. In its 2007 meta-evaluation, UNICEF rated a sample of 313 evaluations, which represents 20 percent of all evaluations for the period, against its report standards. In total, the ratings for Country Office, Regional Office (RO) and HQ evaluations range from a *Very Good* to *Poor* rating from 2000-2006. Over half of the reports (58 percent) were in the *Satisfactory* range while 15 percent were considered *Very Good* and 26 percent considered *Poor*.

Looking at the trend of ratings since the year 2000, a remarkable increase is to be noted in evaluations with a *Very Good* rating from single percentages (2000-2003) to 59 percent in 2005. At the same time, the share of *Poor* rated reports shrank. The improvement in report quality can be attributed largely to the combination of the introduction of the evaluation report standards in 2004 and activities such as increased advocacy in relation to evaluation quality by the Evaluation Office, better hiring of M&E staff, and the issuance of the TOR technical note. However, a critical role remains with the Evaluation and Research Database's (ERD's) inherent quality review mechanism. It is clear that knowing that the reports will be rated causes Country Offices to be more attentive to quality. This is particularly important since despite improved evaluation policy, guidelines and technical notes, UNICEF reports state that Country Office evaluation reports continue to exhibit many of the deficiencies noted in evaluation reports in the early 1990s.

---

<sup>4</sup> United Nations Children's Fund, "Evaluation Report", UNICEF (2006), <[http://www.unicef.org/spanish/evaldatabase/index\\_40075.html](http://www.unicef.org/spanish/evaldatabase/index_40075.html)> (accessed January 29, 2009). The summary above is drawn from and paraphrases this UNICEF report.

- **UNDP** -- In 2007, UNDP conducted a pilot test of an evaluation scoring system based on Evaluation Standards adopted by the UNEG for the United Nations system<sup>5</sup>. The pilot test was carried out with 18 decentralized outcome evaluations, which consider broad results rather than outputs. These outcome evaluations also examine causes other than UNDP programs which may have contributed to observed outcomes. That assessment concluded that 34 percent of the reports were moderately to highly satisfactory, while 66 percent of the reports were less than satisfactory. Analysis of the ratings relative to each criterion assessed suggests that the evaluation reports are weakest in terms of evaluation design and methodology, followed by findings and conclusions. The pilot quality assessment also revealed that although poor evaluation design is a critical element in report quality, other external factors also have an influence. These factors include the quality of the project design, reporting and monitoring, the clarity and comprehensiveness of the terms of reference, and the resources (time and budget) made available to the evaluation. The pilot also provided valuable inputs to the revisions to guidelines and the handbook.

With a UNEG task force focusing on evaluation quality improvements system-wide, UNICEF and UNDP are not the only U.N. organizations to have introduced meta-evaluations as quality control mechanisms. Some U.N. organizations such as UNFPA are using the results of these studies to build quality improvement momentum.

- **UNFPA** for example, completed meta-evaluations of evaluations conducted in 2002-2004 and in 2006-2007, and issued reports showing that percentage rated satisfactory or better for those years was 41 percent and 37 percent respectively. UNFPA used this information to establish an organizational target for evaluation quality with 75 percent scoring in the satisfactory or better range by 2010. The organization has pre-scheduled the next meta-evaluation for 2010 for the specific purpose of seeing whether it is moving towards this target (UNFPA 2007b).

### Summary of UNDP 2007 Evaluation Report Ratings by Quality Criteria (n = 18)

Rating	Overall	Quality criteria						
		Report structure and presentation	Evaluation subject and context	Evaluation objectives and criteria	Evaluation design and methodology	Findings	Conclusions	Recommendations and lessons
Highly satisfactory	3 (17%)	5 (28%)	5 (28%)	3 (17%)	2 (11%)	3 (17%)	3 (17%)	2 (11%)
Satisfactory	1 (6%)	1 (6%)	1 (6%)	0 (0%)	1 (6%)	2 (11%)	1 (6%)	2 (11%)
Moderately satisfactory	2 (11%)	2 (11%)	3 (17%)	1 (6%)	1 (6%)	1 (6%)	1 (6%)	5 (28%)
Moderately unsatisfactory	2 (11%)	6 (33%)	5 (28%)	5 (28%)	3 (17%)	1 (6%)	5 (28%)	2 (11%)
Unsatisfactory	8 (44%)	2 (11%)	2 (11%)	4 (22%)	2 (11%)	5 (28%)	5 (28%)	1 (6%)
Highly unsatisfactory	2 (11%)	2 (11%)	2 (11%)	5 (28%)	9 (50%)	6 (33%)	3 (17%)	6 (33%)

Meta-evaluations as a quality control mechanism are also well institutionalized in some bilateral development assistance organizations. As the experience of the New Zealand Agency for International Development (NZAID) shows, these techniques can be used to improve evaluation solicitations as well as final products.

<sup>5</sup> United Nations Development Program, *Annual Report on Evaluation in UNDP in 2007* (New York: UN, 2008), <<http://www.undp.org/eo/documents/annual-report/2008/ENdp08-25.doc>>. The summary above is drawn directly from this UNDP report.

- **NZAID** as an agency has made a serious commitment to meta-evaluations as a quality assurance mechanism. Meta-evaluations have been undertaken annually since 2004. In reviewing completed evaluations, NZAID uses a rating sheet based on the DAC Evaluation Criteria and NZAID's own guiding principles for evaluation. What is unusual about the NZAID meta-evaluation is that it takes as hard a look at evaluation TORs as it does at evaluation reports, using a matrix to score TORs on a number of dimensions. Findings from the NZAID 2007 evaluation quality meta-evaluation showed that objectives and evaluation questions were not always sufficiently clear nor were TOR methodology sections. A number of TORs were scored as being weak on their description of evaluation reporting requirements and with respect to identifying key documents the evaluation team would need. These findings led to the incorporation of several recommendations for improving TORs in the 2007 NZAID report on evaluation quality (NZAID 2008).
- **DFID** recently agreed to initiate a quality review in 2009 that will assess DFID evaluations using recognized international standards such as those of the DAC and the ECG (ECGnet 2009a). "The study's scope will include Evaluation Department commissioned studies with further drill-down into decentralized evaluative material in two countries and two sectors in which DFID operates. While the main focus will be on evaluation studies, associated systems such as quality assurance and management response will also be considered (DFID 2009b)." This step is being taken at the urging of its Independent Advisory Committee on Development Impact (IACDI), which has made a number of substantial recommendations for improving the quality of DFID evaluations.

As noted above, during the preceding two decades, USAID undertook meta-evaluations for the purpose of understanding evaluation quality issues on a fairly routine basis. The third such review carried out by USAID covered evaluations completed in 1987–1988. This is the only such report that was formally published as an AID Evaluation Occasional Paper 19 and it appears to be the only one that is available in the USAID's Development Evaluation Clearinghouse (USAID 1989). USAID later completed one or two more quality assurance type meta-evaluations, one of which was carried out by MSI and covered the years 1989 to 1990.<sup>5</sup> An important lesson concerning this type of study, from MSI's perspective, was how critical it is at the start of such efforts to assess and calibrate inter-rater reliability, by having raters score and review ratings for the same evaluation(s) until they begin to score evaluations the same way on every dimension of the rating system being used.

### (3) POST-EVALUATION MONITORING EXPANDED

The third set of initiatives reflect a sustained organizational impulse to strengthen the link between conducting evaluations and program improvements, and focus on what happens after an evaluation report has been reviewed, revised and accepted. To a greater or lesser degree, evaluation policy statements of development assistance organizations prescribe three types of post-evaluation activities:

- the dissemination of evaluation reports, including oral briefings;
- decision meetings in which staff make choices about which evaluation recommendations to accept/reject, and prepare an action plan for implementing evaluation recommendations; and
- follow-up procedures that ascertain whether advice and recommendations from evaluation were acted upon.

---

<sup>5</sup> Meta-evaluations conducted by USAID's evaluation office (CDIE) in the 1980s and 1990s did not rate evaluation Statements of Work. Since conducting its last meta-evaluation for CDIE however, MSI has created a score sheet for reviewing evaluation Statements of Work. This score sheet was used in a Mission level meta-evaluation carried out for USAID/Uganda in 2007.

## **a. Dissemination of the Results of Evaluations**

In donor agency evaluation policies, the dissemination of evaluation reports subsumes both internal dissemination for management response and public dissemination. Since the same mechanisms are used for both purposes internally, the topic as a whole is addressed below. More emphasis, however, is placed on broad dissemination than on the kinds of immediate dissemination that link to agency processes for fostering management action based on evaluation results.

Policies and practices for the broad dissemination of evaluation results to staff and the public vary across development assistance donor organizations. Generally speaking, donor organizations link their broad dissemination policies to two motivations. The first, for a number of donors, is transparency and accountability. The second is fostering utilization beyond the sphere of immediate decisions about on-going programs, i.e., what the literature on evaluation utilization discussed above calls “enlightenment” use or “influence” versus “instrumental” use. There is no strong new trend in this area. What is observable instead is simply a gradual expansion of the number of donors that provide relatively easy public access to their evaluation reports.

Some donor organizations are very specific about their dissemination procedures. For example, the European Commission’s evaluation policy states that once received, evaluation reports are first distributed to senior management by the evaluation manager. Fifteen days later, unless senior management requests a delay, the evaluation manager automatically makes an evaluation report public, including an abstract, which the evaluation manager is responsible for creating. Looking beyond agencies and their websites, AusAID requires evaluation managers to prepare targeted evaluation dissemination plans that clarify what audiences most need to be aware of in evaluations results and the best ways to reach those audiences (ADB).

Other donor organizations, including the Inter-American Development Bank (IADB), include a discussion in their dissemination policy statements concerning document sensitivity. Thus while the IADB generally makes its evaluation reports public, its policy calls for the completion of a review by its Executive Directors prior to public dissemination. USAID’s ADS 230.3.6.7 calls for openly sharing evaluation reports with partners and other donors “unless there are unusual and compelling reasons not to do so,” but it does not outline specific procedures for making such decisions.

## **b. Electronic Libraries of Evaluation Reports**

Most donor organizations have electronic libraries, which are the most common mechanism for providing public access to evaluation reports. In its 2005 publication, *Norms for Evaluation in the UN System*, (section 2.7), the UNEG incorporated an expectation that all donors would create evaluation repositories.

Existing donor evaluation libraries vary in terms of their completeness and the ease with which evaluations stored in them can be accessed. The question seems to turn on whether donor organizations have systems in place that accurately count the number of evaluations conducted each year and ensure that copies of all evaluations reach their electronic libraries. UNICEF, for example, reports that 183 evaluations were completed in 2007. Similarly, in its 2008 Annual Report, the World Bank reports the exact numbers of evaluations it has completed or has underway. They state “158 impact evaluations are being conducted (under the Guidance of the World Bank’s Chief Economist) in comparison to 106 impact evaluations completed between 1999 and 207 (WB 2008a).” Other organizations publish approximate figures on the numbers of evaluations they conduct per year.

For the most part, USAID estimates rather than counts the number of evaluations completed each year. The Agency is using a combination of reports on evaluation plans submitted by decentralized units and occasional counts of the number of evaluations received by the DEC. Discussions with USAID staff over the last decade suggest that figures from evaluation plans may overstate the number actually conducted, while figures from the DEC understate evaluation coverage. This under-representation is despite the fact that submission of these reports is mandatory as outlined in ADS 203.3.6.8.

In terms of staff and external access to evaluation reports, some donor organizations, including USAID, post a link to evaluations on their main evaluation web pages. When this is not the case, then external users and staff have to be resourceful to locate them. Electronic libraries maintained by some donors, including USAID, include evaluations undertaken over several decades. In some other donor agencies, only recent evaluations are easy to access. Functional access to evaluations, i.e., getting to the sub-set of evaluations that might help you make a decision or plan a new program, also differs.

One of the issues is whether documents other than true evaluations are pulled out when searches for evaluations are conducted. The World Bank, for example, houses all of its impact evaluations in a single location that is searchable by country, sector, or evaluation methodology. USAID's DEC is also set up to be searched by country and sector, but searches using the key word "evaluation" results in other types of documents as well. Thus a search of the DEC for USAID evaluations completed between 2005 and 2008 that was undertaken as part of this study<sup>6</sup> yielded inaccurate results. Of a total of 1,303 documents, only 296 (23 percent) were true evaluations, the others were a mix of contractor final reports, GAO reports, etc. Discrepancies of this sort not only make it difficult to detect the number of evaluations an organization has on file, but may also serve as a deterrent to utilization by staff and other potential users who find the task of sorting true evaluations too time consuming.

Another relevant issue, but one which is not yet being widely discussed, is the mixed quality of what donor organizations release when they disseminate evaluation reports. As indicated above, most donor organizations that have undertaken meta-evaluations to determine the quality of their evaluations have yet to create a link between their evaluation quality assessments and their dissemination processes. This means that there are no quality scores to provide users with a better idea of how seriously they should treat the evaluation's findings and recommendations. UNFPA may be an exception in this regard. In a 2007 statement on compliance with internationally recognized evaluation standards, UNFPA indicated that it intended to design and populate an internal database for staff of quality-controlled evaluations which will be available to all staff (UNFPA 2007a).

### c. **Syntheses of Evaluation Findings by Year and/or Topic**

In addition to releasing individual evaluation reports to staff and the public, a number of development assistance donors view synthesizing evaluation results to make them more accessible for potential users as one of their dissemination responsibilities. A number of donors have prepared syntheses of evaluation results and issued them as separate publications:

- **UNDP** publishes a series called *Essentials* that periodically provides readers with a synthesis of evaluation findings on a specific topic or type of development program. Fourteen issues of *Essentials*, covering a wide range of topics, were published in English, French, and Spanish between 1999 and 2004 (UNDP 2009a).
- **CIDA** has a similar series titled *What are We Learning?* that has produced six evaluation syntheses (CIDA 2009).

In contrast to topical synthesis reports, some donor organizations have employed techniques to extract useful information on a country basis or for a certain period of time.

- For 2006-2007, **DFID** looked at evaluations for five African countries within a particular time period in order to synthesize the findings. What emerged were a set of conclusions about the country assistance programming process (DFID 2006).

---

<sup>6</sup> See Annex A

- The **Asian Development Bank (ADB)** has produced a mix of country and topical evaluation syntheses, which it posts on a page of its website (ADB 2009b).

USAID has also created a number of evaluation syntheses documents. This was a standard element of multi-country, topical ex-post evaluations (called “Impact Evaluations”) undertaken by CDIE from the 1980s to the early 2000s. In addition, USAID has occasionally prepared other evaluation synthesis documents when there appeared to be a need and a sufficient number of evaluations available to do so, e.g., on investments in agriculture (USAID 1996). The World Bank’s practice of producing occasional evaluation syntheses, e.g., summarizing the findings of transport sector evaluations over a decade, is similar (WB 2007).

A key difference between the practices of UNDP and CIDA, on the one hand, and USAID and the World Bank, on the other, is ease of access. Since UNDP and CIDA use a series as the mechanism for their dissemination of evaluation syntheses, they are much easier to locate than are similar products produced by other donors.

Also of note with respect to electronic libraries of evaluations and evaluation syntheses is that while some donor evaluation policy statements discuss the importance of making evaluation findings available in local or multiple languages to foster utilization, action on this front is limited (USAID 2008a). UNDP’s *Essentials* series is an exception in this regard.

#### **d. Follow-Up on the Link between Evaluation and Management Actions**

As research on the utilization of evaluation in government agencies has shown, the link between evaluations and action is largely a function of the organizational culture. Recent reports from the GAO (GAO 2003) on the role of “evaluation culture” and a Canadian Treasury Board Study on “effective evaluation (Canada 2004)” both highlight the crucial role that agency management plays in setting an organizational tone that is receptive to learning not only about what works (success stories) but also what does not work. Organizational structures for systematically processing all evaluations, rather than shelving those that deliver unsettling results, are a critical element in evaluation cultures that are likely to act on evaluation findings.

With respect to structures for processing evaluation results, evaluation policies of many donor assistance organizations identify post evaluation reviews and the development of action plan for implementing recommendations from an evaluation as a key stage in the evaluation process. A consensus has existed among bilateral donors on the importance of instituting and monitoring the use of evaluation follow-up mechanisms. This consensus arose from the publication of DAC *Principles for Evaluation of Development Assistance* in 1991, which was reconfirmed in the development of the 2005 *Norms for Evaluation in the UN System*. What is relatively new, however, is the seriousness with which some bilateral and multilateral donors are monitoring post-evaluation processes and outcomes.

- **SIDA**, for example, takes in to consideration the results of evaluations in policy-making, in processing ongoing development assistance activities and in decisions about new activities. A “management response system” is a formal part of SIDA’s evaluation process. Through this system departmental directors are made responsible for the follow-up of the implementation of action plans within their respective areas of responsibility. Every six months, SIDA follows up on the implementation of action plans made in response to evaluations it undertakes (Bandstein and Hedbloom 2008).
- The **World Bank** has also increased its attention to post-evaluation adoption and implementation of evaluation recommendations. Since 2002, the World Bank’s Independent Evaluation Group (IEG) has documented the percentage of evaluation recommendations from its studies that were adopted by Bank management. Data for the years 2002-2006 is published on the World Bank’s website, showing an average rate of adoption of evaluation recommendations of around 70 percent for the

period. This website also documents the impact of some of the actions taken based on evaluation recommendations (IEG 2009b).

- In June 2008 **DFID** issued its first annual accounting of the actions taken after evaluations were completed. The introduction of this type of report on an annual basis was one of the top priority recommendations to emerge from the work of DFID's Independent Advisory Committee on Development Impact (IACDI) (DFID 2008). In addition to citing a limited number of examples of actions taken following evaluations, this report highlighted specific measures and the general cultural change being introduced in DFID to ensure more effective post-evaluation reporting in the future, including the creation of a tracker system and a decision to designate specific directors to follow-up on each evaluation.
- **UNDP** in 2009 issued new and more detailed instructions for preparing management responses to evaluations and issued a new template for recording evaluation review meeting decisions about actions to be taken. Once completed, responsible operating units are required to enter these forms into the UNDP's Management Response Dashboard managed by Evaluation Resource Center (ERC) where they are monitored to determine the extent to which agreed upon actions have been implemented (UNDP 2009c). The status of actions by country and evaluation are available to senior managers on a real-time basis (UNDP 2008).

Other donors have well-defined procedures for making decisions on evaluation recommendations and following-up to determine whether they were implemented. DANIDA, IFAD, and the FAO, among others, have similar follow-up initiatives underway. In 2006, a Canadian peer review of an evaluation in UNICEF congratulated that organization on its new system for tracking the degree to which evaluation recommendations were followed, and suggested that UNICEF expand this effort by checking at intervals, even over several years, to determine the long-term impact of evaluations (CIDA 2006). This is essentially what the World Bank did in its 2004 report on Influential Evaluations (WB 2009).

USAID, in ADS 203.3.6.7, recommends that missions should follow-up after an evaluation has been completed. These steps are neither mandatory nor require missions to inform Washington of the accepted recommendations or plans for taking action. [Until the 1990s, USAID required this kind of post-evaluation reporting (USAID 1995).] This form in many ways resembles the form UNDP has introduced as part of its Management Response Dashboard. USAID, which never had an implementation tracking system similar to the World Bank and UNDP, has set up its requirement for this type of reporting. USAID states in its 1997 TIPS on the role of evaluation in USAID that project summaries (form 1330-5) are no longer required (USAID 1999).

#### **e. Improving the Linkage Between Evaluation and Performance-Based Budgeting**

Making sensible linkages between what works and what doesn't in development assistance and forward budget decisions is a recurrent theme. It is also an arena in which evaluation has been expected to be of more help than has generally been the case. Given the current fiscal climate improving this linkage could once again rise to the top of the agenda.

The passage of the U.S. Government Performance and Results Act in 1993 was considered a serious effort in this regard. The act had a built-in plan for performance budgeting pilot tests that were expected to draw heavily on knowledge gained from performance monitoring and evaluation. In 1999, however, the GAO provided Congress with a report on the difficulties involved in performance based budgeting and OMB's decision to postpone efforts to undertake these pilots. Among other things, the GAO reported that no agencies had volunteered (GAO 1999). PART, developed by OMB, had much the same intent but also overlapped GPRA initiatives. While somewhat useful as a performance-to-budget tool between agencies and OMB, it was less successful as an evaluation-to-budget bridge with Congress (GAO 2005).



While some efforts have been made since then, largely through PEPFAR and the new Foreign Assistance Framework, confidence about what is being measured is uneven. Likewise, there are uneven views regarding the appropriateness of using information from standard indicator exercises as the basis for significant programming decisions. Other donor agencies are not much further along. AusAID is the one standout, at least on paper, and appears to be a fairly strong advocate of the concept of linking evaluation and budget decision making.

Lack of progress in this arena in the U.S. and in most development assistance donor organizations does not, however, mean that others are standing still. Impressive strides toward linking evaluation and budget process have been made in a number of Latin American countries (WB 2008c). Chile is setting the pace and developing government-wide evaluation to budget mechanisms (McKay 2007). This and other examples of “managing for development results” efforts that integrate evaluation were presented at a 2008 performance budgeting conference in Mexico (WB 2008b). The conference included a session focused on the use of Logical Frameworks and other development planning tools as aids in making the evaluation to budget linkage (Robinson). This experience is a valid basis for donor organization evaluators and budget to improving the linkage within their own organizations.

#### **f. Documenting the Linkage between Evaluation and “Next Generation” Programs**

While the immediate or instrumental use of evaluation findings is the type of utilization to which the greatest attention is given in donor organization, many in the field think that the real payoff comes downstream, as people reflect on evaluations. While no oral history of such experiences has been written, some in the development community can cite case examples of which they are aware.

The literature on evaluation has long encompassed the idea that evaluations can be used on a deferred basis and in ways well beyond what is termed “instrumental” use. This possibility was initially referred in evaluation literature as “enlightenment” use, coined by Carol Weiss (Weiss 1979, 426-431), and more recently as “influence,” based on Karen Kirkhardt’s model describing the way in which evaluation findings permeate communities and gradually shift paradigms. What is missing from the evaluation literature, however, is a body of research that demonstrates that this actually occurs. Some case studies have been published over the years but they, like the World Bank’s volume on Influential Evaluations, tend to provide an *ad hoc* look at mid-term results of utilization. The alternative is based on empirical evidence to help understand how “enlightenment” or “influence” actually work, and what can be done to make those kinds of uses happen more systematically. Simply stated, research in this field had not gone as far as the kind of long-term wide-band use of evaluations that Canada recommended for consideration in its peer review of UNICEF evaluation practices.

### **C. NEW THINKING ABOUT THE ROLE AND PLACEMENT OF EVALUATION OFFICES**

In addition to documenting trends in the way evaluations are conducted, this study examined shifts in the way evaluation offices define their role and their location, as donor operations emphasize independence from their organization’s operations.

#### **(1) REDEFINING THE ROLE OF THE CENTRAL EVALUATION OFFICE**

Evaluation policy is an important responsibility of almost all central evaluation offices in donor agencies. Most evaluation offices, at least nominally, are considered to be the managers of their organization’s evaluation systems. Beyond that, these offices vary with respect to the way they define their role. One key way in which they differ is in how they manage organizational evaluation systems. They also differ on direct engagement in the conduct of evaluations.

### a. Managing the Evaluation System

Most evaluation offices play some role in (a) determining evaluation staff qualifications, (b) meeting training needs, and (c) issuing guidance that helps evaluation staff to implement agency evaluation policy. In addition, a number of central evaluation offices have become much more active managers of evaluation processes in their agencies. As previously discussed, more assertive evaluation management practices include scoring the quality of agency evaluation reports, recording evaluation recommendations to their agencies, and monitoring and reporting on the implementation of these recommendations. Uniformly, central evaluation offices link their more assertive practices to strengthen the “if-then” chain that connects evaluations, through utilization to improved aid effectiveness. Managing for results in this sense, and transparently keeping score, is a new trend in these offices.

### b. Conducting Evaluations

When evaluation was new in donor agencies, some central evaluation offices were the only units capable of conducting evaluations. In other organizations, where evaluation competency existed in a number of units, central evaluation offices sometimes undertook evaluations as a means of demonstrating new methods, supporting learning about cross-cutting issues or for other narrowly-defined reasons. As their agencies decentralized operations, some central evaluation offices handed off the operational responsibility for project level evaluations, and either cut back or ceased to directly engage in evaluation studies. Other offices continue to look to their central evaluation offices to conduct a portion of their evaluations. Other variations, found in one or more donor agencies, involve the creation of centers of excellence in evaluation within the organization, such as the World Bank’s Development Impact Evaluation Unit (DIME).

## (2) EXPERIMENTING WITH THE RIGHT DEGREE OF INDEPENDENCE

The evaluation literature and standards for evaluation issued by the DAC, UNEG, and national evaluation associations tend to view questions about evaluation quality and the evaluator independence as being interrelated. Independence of an individual evaluator, or for evaluation offices, is viewed to mean that evaluators and the products they produce are impartial, objective, unburdened, and balanced (Mathison 2005, 199). Threats to the evaluation independence arise from multiple sources, as the text box on page 31 indicates. Every donor organization to some degree addresses the evaluation independence through its policy related practices. On the issue of independence, most donor organizations associate themselves with the evaluation standards of some external entity, e.g., the DAC, the UNEG, the ECG or a national evaluation association. USAID, on the other hand, does not explicitly make such a linkage in its evaluation policy.

With respect to evaluation independence in structural terms, two trends spanning the past twenty years are important, in part because they move in somewhat different directions and the implications of those differences are not yet completely clear. The first trend involves the decentralization of some evaluation functions (Manning 2008). As noted previously, the decentralization of project-level monitoring and evaluation to field staff or to host governments moves the responsibility for the quality of a large portion of evaluations in a donor’s portfolio to individuals who generally have limited evaluation training. USAID, which began decentralizing evaluation responsibilities to the field in 1968 (USAID 1970) and to Foreign Service Nationals a decade later, has provided evaluation training to a number of its FSN Mission Evaluation Officers. Other donors, with more recent experiences of decentralization and the sharing of evaluation responsibilities with host governments, are facing similar challenges to build evaluation capacity.

The second trend is the elevation of the central evaluation units in the name of greater independence. In its 1991 publication of *Principles for Evaluating Development Assistance*, the DAC stated that “impartiality and independence will best be achieved by separating the evaluation function from line management responsible for planning and managing development assistance.” The UNEG’s 2005 product, *Norms for Evaluation in the UN System*, parallels the DAC position. A 1996 Multilateral Development Task Force report follows this line

of thought but also recommends that heads of multilateral bank evaluation units should be allowed to release evaluation reports to bank management and the Executive Board of Directors without outside clearance (ECGnet 2009).

### Threats to Evaluation Independence

- **External Pressure from Individuals in Positions of Authority.** Suggestions from the institutional manager of a contracted evaluation or from others in that institution to an evaluation team about what it is acceptable to say in an evaluation report may constitute a threat to independence. On the other hand, suggestions focusing on factual errors, gaps and even on whether “bad news” was presented in a diplomatic way do not constitute threats to evaluation independence. Communications that have the effect of forcing evaluators to remove evaluation findings, or conclusions and recommendations that are directly traceable to those findings may be considered external pressure, regardless of whether those communications include explicit or implicit penalties, e.g., non-payment, blacklisting, or salary or promotion impact if the team is internal to the organization. This would not necessarily be the case if the material an evaluation team was asked to remove was beyond the agreed upon scope of the evaluation. Refusal of individuals within a chain of command inside an organization to move an evaluation up to higher levels or refusal to disseminate an evaluation despite policies that require publication and/or dissemination also represent instances of threats to evaluation independence.
- **Self Censorship and/or Implicit Social Pressure.** The independence of evaluation products can be compromised by evaluators themselves, even in the absence of any external pressure. Wanting to be liked, not wanting to offend colleagues or upset them, or withholding truths that the evaluator thinks the audience would not be able to cope with suggest a range of evaluator motivations that could compromise the independence of an evaluation, even where no direct conflict of interest or issue of future monetary gain exists. In 2003, CNN's top war correspondent, Christiane Amanpour, leveled charges related to this type of threat to independence against the U.S. press, including her own network.
- **Conflict of Interest/Future Monetary Gain.** While a good deal of the literature on evaluation suggests that external evaluators are inherently more independent than internal evaluators, others challenge this claim. The basis of this challenge is prospects being hired to do other work, including evaluations, rather than conflict of interest, which is often guarded against by provisions in contract that preclude external evaluators from bidding on future phase of a project or future projects in the same sector in the same location.
- **Non-Disclosure.** Evaluation independence may be threatened when members of an organization deny access or hide program documents or sources from an evaluation team. Non-disclosure of pertinent information threatens evaluation independence by preventing an evaluation team to present an objective and balanced report. Weaknesses in the monitoring systems associated with projects, programs and policies that are being evaluated may indirectly threaten evaluation independence of the same nature. Monitoring data that does not meet data quality standards (validity, reliability, integrity) but is presented to an evaluation team as if it does, falls under this rubric as it could just as easily mislead an evaluation team as could other type of non-disclosure.
- **Bias.** Any predisposition towards a project or program that interferes with objectively reporting on performance may constitute a bias that can threaten evaluation independence. This is why decisions about how to construct an evaluation team, and the role of experts on such teams, can be problematic. Too little expertise and an evaluation team may flounder; too much and it may ratify a faulty design. Deeper problems of this nature, i.e., cultural biases of development assistance program evaluators and those involved in projects they evaluate, were the focus on Nobel Laureate Gunnar Myrdal's still valid 1969 challenge to social scientists in his short volume, *Objectivity in Social Science Research*: “We are deceiving ourselves if we naively believe.... that we do not tend to aim opportunistically for conclusions that fit prejudices markedly similar to those of other people in our society”

In practice, most donor evaluation structures reflect these principles (Foresti 2007).

- **SIDA** has a three-part evaluation system. SIDA departments carry out evaluations within their purview, generally for directly related management purposes. SIDA's Evaluation Office, which reports directly to the head of the organization, carries out broad thematic and strategic evaluations. In addition, SIDA programs are also subject to evaluation by a separate organization, the Swedish Agency for Development Evaluation (SADEV) (SADEV 2009).
- The **World Bank** has a central evaluation office populated by evaluation staff who conducts a relatively large number of evaluations each year. The Independent Evaluation Group (IEG), formerly known as the Operations Evaluation Department (OED), reports directly to World Bank's Board of Executive Directors, not to the Bank's management. Similar changes were made at the Asian Development Bank (**ADB**) in 2008 (ADB 2009a). At both of these multilateral institutions, these independent offices conduct some but not all evaluations. Country and technical offices conduct and report to the Bank's management on what are called self-evaluations of every project.
- **USAID**, from the early 1980s to the early 1990s, housed its evaluation responsibilities and the conduct of thematic evaluations, typically on an ex-post basis for the purpose of ascertaining their long-term effects, in a conceptually separate Center within its Policy Bureau. This structure, the Center for Development Information and Evaluation (CDIE), had some level of optical if not bureaucratic independence. Though in practice, most evaluations were carried out by field Missions and central technical offices rather than by this unit.

In 1996, the ECG went beyond principles of independence generally accepted in the development community by calling for greater structural independence for multilateral development bank evaluation units. In other words, establishing a direct reporting relationship to the Boards of these organizations. The ECG also recommended having Boards, rather than management, make decisions with respect to hiring and terminating the heads of multilateral bank evaluation units (ECGnet 2009a). In parallel, the ECG issued a template for assessing the independence of multilateral development bank evaluation units (ECGnet 2009b). This step resulted in changes in the organizational structure and reporting relationships at the World Bank and in the Asian Development Bank.

While organizational independence at the level recommended for the multilateral development banks ensure access to the Boards of these entities, it does not necessarily guarantee that they will function independently. At the IMF, for example, where the evaluation unit has been "fully independent from the IMF's management and staff, [and] operates at arms length from the Executive Board" since its creation in 2001 (IMF 2009), questions about the functional independence of this unit, or lack of it, became a public issue in 2006 (Bretton Woods 2009).

Moreover, moving too far in the direction of structural independence, in the eyes of the authors of the DAC Principles, may be counterproductive:

The organizational aspects must address three requirements: developing a policy and a set of guidelines for evaluation; ensuring impartiality and independence; linking evaluation findings to future activities. Often, certain types of organizations will tend to strengthen one of the above requirements at the expense of others, e.g., ensuring independence of the process may weaken the potential for providing a strong linkage between evaluation findings and decision-making. An optimal solution should be sought to balance all of these requirements (DAC 1998).

Consistent with this view, one DAC review of the evaluation function in France's *Agence Française de Développement* (AFD) concluded that the evaluation unit was too isolated from the rest of the institution, which resulted in its being brought into the research unit of the organization's Strategy Branch in 2006 (AFD

2009b). The DAC recommendation and the AFD's ensuing action reconfirms the basic conclusion of the DAC Principles on the issue of structural independence, i.e., "whatever approach is chosen, the organizational arrangements and procedures should facilitate the linking of evaluation findings to programming and policy making."

With regard to these choices, DFID is currently in the process of revisiting issues of evaluation independence. Similar to the structure of other agencies, DFID has a central evaluation office (EVD) composed of about 20 people who oversee, but do not themselves participate in, external evaluations with a country or thematic focus. In addition, operating units undertake project-level evaluations which, as in some other agencies, are called self-evaluations even if conducted by external evaluators. In 2006, a DAC peer review suggested an effort to revitalize the evaluation function in DFID. In parallel, a survey undertaken by Britain's National Audit Organization reported that while 80 percent of DFID staff viewed evaluation independence to be extremely important only 25 percent viewed evaluation in DFID to be entirely or very independent (Picciotto 2009). That survey also indicated that decentralized evaluation needed to be enhanced. An Independent Advisory Committee on Development Impact (IACDI) was soon established. Since 2007 evaluation in DFID has been overseen by IACDI, reporting directly to the International Development Secretary.

To help formulate permanent arrangements for evaluation in DFID, IACDI commissioned an external review of evaluation independence in DFID. This study, like the DAC principles, started with the proposition that "optimum independence is not maximum independence (Picciotto 2009, 5)." However, because this study looked primarily at the EDV, rather than more holistically at the dual aspects of DFID's evaluation system, it used the ECG's template for assessing evaluation independence as a conceptual framework. The result, not unexpectedly given the imperatives built into that instrument, was a decision by IACDI to "match good practice," as defined by the EDC for multilateral development banks. Additional decisions were to make the head of EDV directly accountable to the Permanent Secretary of DFID, reduce the number of external evaluations, and increase the degree to which EDV staff themselves conduct evaluations, while at the same time elevating the post to more senior grade and status to "give it greater visibility and clout." What these decisions did not do, by the IACDI's own admission, was effectively incorporate the task of improving the quality of decentralized evaluations and their utilization into the EDV's mandate (Picciotto 2009, 2). Defining how decentralized evaluation would be improved was left for the attention of a new evaluation policy. The proposed policy indicates only that encouragement and possibly some training will be given to program staff to make projects more evaluable, and that EVD will periodically sample and report on the quality of decentralized evaluations (DFID 2009b).

USAID is also reconsidering how its evaluation function should be structured for reasons quite different from DFID or the *Agence Française de Développement*. In addition to the caution about swinging too far in any one direction, the experiences of these two organizations, and of other donor agencies as described elsewhere in this paper, have much to offer to this process. On one hand there are tools that could be used to get a clear reading on USAID's current status on evaluation quality and utilization, and on the other hand, a more precise understanding of the degree to which evaluation independence is an issue in USAID.

## **5. OPPORTUNITIES FOR IMPROVING EVALUATION AND ITS UTILIZATION IN USAID**

This section is intended to be useful for USAID's new central evaluation office by presenting a number of opportunities for improving evaluations and their utilization in USAID that seem to emerge from the trends examined by this report. The options presented recognize that it would not be practical or even desirable to invest in all of them at the same time.

## A. UTILIZATION OF EVALUATION RESULTS

USAID has a range of options it could pursue to improve the utilization, and thus the development impact of evaluation in the Agency.

### (1) STRENGTHEN THE EVALUATION CULTURE IN USAID

The term “evaluation culture” is used by the GAO and others who observe and comment on the integration of evaluation into the management processes of government agencies. A strong evaluation culture has a commitment to learning, even if that means acknowledging what is not working, is strong at all levels of an organization. Problems arise when incentives within agencies are not consistent with a learning model and when staff commitment to evaluation is uneven across levels of the organizational hierarchy. While USAID top management commitment to evaluation has been strong on paper over the decade, other aspects of USAID’s organizational culture have not necessarily echoed that commitment. USAID’s stress on *Success Stories*, for example, sends out a conflicting message. Further, anecdotal evidence from USAID staff suggests that top management’s commitment to evaluation has been stronger than has been the commitment of mid-level managers. The persistence of low numbers of evaluations since 2000, compared to prior decades, correlates with this anecdotal information. USAID has a number of options it can pursue that could help strengthen the Agency’s evaluation culture:

- **Invest in a Situation Analysis**

The GAO’s approach to analyzing evaluation culture across a number of organizations (GAO-03-454) offers USAID a basis for conducting its own review, either on an in-house basis or with assistance from either the GAO or another external resource. A systematic effort would look broadly across the Agency, gathering impressionistic as well as factual data. It would help USAID determine, from an empirical base, where and how to focus efforts to strengthen USAID’s evaluation culture.

- **Diversify USAID’s Evaluation Training Focus**

Regardless of what a situation analysis reveals about USAID’s evaluation culture, the Agency might want to explore options for diversifying its evaluation training portfolio to better reach mid-level managers with new evaluation information. Currently evaluation training is embedded in a course for new staff or involves a significant commitment of time. Other types of courses, of shorter duration, would help mid-level managers and key technical advisors to Missions stay up-to-date on development in USAID and in the evaluation field.

- **Consider the Evaluation Culture Implications of Options for the Locus of USAID’s new Evaluation Office**

USAID’s new Evaluation Office has a significant role to play in terms of improving the Agency’s evaluation culture. In this regard, the stature of the Office within the organization is likely to matter. In earlier eras, USAID’s Evaluation Office drew on its status within the Agency’s policy bureau to assert the importance of evaluation in the Agency’s management process. Although the Office’s influence also depended on the personal leadership skills of the Office’s Directors and their access to the Administrator, it was not the only avenue used to promote a positive evaluation culture. Active participation by this Office in mission director and program office conferences were equally important and perhaps had even greater influence on the extent to which evaluations were integrated into mission and technical office management processes.

As shown in the previous section, some donor organizations are elevating their evaluation offices, thus making them accountable to agency administrators or boards of directors. Others are cautioning that moves of this sort isolate the Evaluation Office and reduce its influence on evaluation practice in the field. The two issues involved in such decisions are influence and independence. Influence has to do

with the authority of the Evaluation Office to not only set but also oversee, support, monitor and criticize the adequacy of evaluation policy and practice. Independence has more to do with the ability of an Evaluation Office to set an evaluation agenda, monitor and report on evaluation practices and conduct evaluations it believes are needed without interference and voice their results no matter what they show.

For USAID today, the importance of independence in deciding on a final home for the new Evaluation Office depends on whether the Office plans to conduct evaluations or not. If the new Evaluation Office does not plan to conduct evaluations, or if it does and there is no history in USAID of efforts to suppress the findings of evaluations conducted centrally, then independence may not be a valid criteria for making a decision on office location.

## **(2) UPDATE USAID'S DETAILED KNOWLEDGE OF ITS UTILIZATION OF EVALUATIONS**

Nearly three decades have passed since USAID last formally examined its patterns of evaluation use. While theory and a modest amount of empirical data generally point to stakeholder ownership, evaluation quality and follow-up as key to evaluation use, the relative importance of these factors in USAID today is not clear. USAID's new Evaluation Office stands to benefit from any effort it makes to gather detailed information about the extent of evaluation use throughout the Agency, including factors that support and discourage evaluation use. Such information will help the new office make informed decisions about the appropriate levels of investment in improving determinants of evaluation use. An effort of this sort should not hold up Evaluation Office action in areas where it already knows needs exist. Rather gathering data on these questions could take place over an extended period and could be executed in one or more approaches, including:

- An in-house survey of technical office staff and M&E officers. These types of surveys have been done periodically in the past, including earlier in this decade by USAID's previous evaluation office. This might also be a project of the informal network of seasoned M&E officers in Washington that has emerged over the past several years.
- Off-agenda roundtables at regional Program Officer conferences using a common set of questions across continents.
- Key informant interviews, using a common format, with Mission Directors, Deputy Directors and field based Technical Office Directors as they rotate through Washington.
- An external study carried out by a specialist in evaluation utilization, of the type undertaken for USAID decades earlier.

## **(3) SPOTLIGHT GOOD EXAMPLES OF EVALUATION USE**

While USAID's ADS encourages the use of evaluation findings, documented examples are rare. Individual USAID staff may be aware of instances where the immediate use of evaluation findings led to improved aid effectiveness, or of technical areas in which the cumulative impact of a number of evaluations led, over time, to change in policies or technical approaches. Absent mechanisms for communicating these experiences, and encouraging staff to view them as models, they have limited impact on the organizational culture. Periodic publication of examples of effective evaluation use could help in this regard. Existing approaches for identifying and publicizing *Success Stories* offer a model. Further, mechanisms for periodically disseminating good examples of evaluation use already exist, e.g. Front Lines and the evaluation section of USAID's website.

#### **(4) MAKE THE FINDINGS AND LESSONS OF EVALUATIONS MORE ACCESSIBLE**

Most USAID staff would say that they do not have the time to read all of the evaluations that are added to the Development Experience Clearinghouse (DEC), even in their own technical field. Annual digests of evaluation findings by technical field and sub-field, published separately, rather than in a single large compendium could improve this situation. Digests of this type could include a short essay on implications as well as abstracts of the findings of prior year's evaluations in a field. Rather than trying to prepare these digests itself, USAID's new Evaluation Office might collaborate with and co-fund efforts of relevant technical offices. Another option would be to provide a searchable database of evaluation abstracts on its website, modeled on the DAC evaluation abstract database ([http://www.dac-evaluations-cad.org/dac/abstracts\\_e.htm](http://www.dac-evaluations-cad.org/dac/abstracts_e.htm)). Should USAID not want to introduce a new database for this purpose, it could, as a third option, improve the accessibility of evaluations by including abstracts or free standing Executive Summaries for every new evaluation it adds to the DEC. In addition, some types of information, including critiques of USAID evaluations, might be more forthcoming if posted on an internal website.

The completeness of any USAID effort aimed at summarizing and synthesizing the contents of evaluation reports depends on staff and implementing partner adherence to USAID's mandatory requirement in ADS 203.3.6.8 for forwarding evaluations to the Development Assistance Clearinghouse (DEC). Anecdotal information suggests that compliance with this requirement is incomplete. The reasons for this also suggest options for improvement.

- Missions and offices vary in terms of who they make responsible for compliance with this mandatory requirement, e.g., the Mission M&E Officers, technical team leaders, contract/grantee activity managers, or implementing partners. Encouraging all operating units to handle this responsibility in the same way might help to improve compliance.
- Some evaluations may be too sensitive to submit to the DEC. USAID has a procedure for handling these situations that involves making such documents available only internally. Missions may not all be aware of this and the new Evaluation Office could act to improve understanding of this provisions for sensitive evaluations.

#### **(5) ADD UTILIZATION TO M&E JOB DESCRIPTIONS**

USAID job descriptions help guide the actions of individual staff and their supervisors, since they form the backdrop against which annual plans and performance reviews are written. Fostering evaluation utilization across the Agency can be encouraged by making evaluation utilization a core element of every M&E officer's job description. In practice, this would lead to decentralized efforts to define what these officers could do each year to improve evaluation utilization in their operating unit. Given the decentralized nature of the organization, this could be one of the most powerful steps USAID could take to strengthen action in support of this aspect the evaluation mandate set forth in ADS 203. This might involve collaboration between USAID's new Evaluation Office and the Agency's Human Resources staff to draft a new sample M&E officer's job description that includes strengthening monitoring and evaluation utilization as a core component.

#### **(6) EXPAND THE EMPHASIS ON UTILIZATION IN MONITORING AND EVALUATION TRAINING MODULES**

USAID teaches monitoring and evaluation skills in a number of Agency courses, including courses offered through technical offices (e.g., HIV/AIDs) that include a monitoring and evaluation segment. These courses provide a natural venue for promoting improved evaluation utilization. A review of what is currently being taught about evaluation utilization in all of these courses could suggest where improvements could be made. A willingness on the part of the Evaluation Office to collaborate on improvement efforts, e.g., helping courses find case examples, or reviewing new materials, could go a long way towards ensuring that such recommendations are implemented.



## **B. STAKEHOLDER “OWNERSHIP” OF THE EVALUATION PROCESS**

Evaluation theory and a modest amount of empirical data indicate that, among key program or project stakeholders, a strong sense of ownership of the evaluation process and its findings and recommendations tends to enhance utilization. In a development context, stakeholders generally mean key decision-makers, i.e., organizations that fund or implement programs or projects, or will be expected to sustain them after donor funding ends. These stakeholders are the natural audience and users of evaluation results. Their sense of ownership is thought to rise when they play a substantial role in initiating an evaluation and framing the questions evaluators will address. An evaluation need not be participatory in the fullest sense, i.e., involve stakeholders in data collection, analysis, framing recommendations, to establish a strong sense of ownership among stakeholders. To the degree that USAID determines that lack of ownership is constraining evaluation use in USAID, several options for intervening to improve use may warrant consideration:

### **(1) IMPROVE PRACTICAL GUIDANCE ON APPROACHES THAT STRENGTHEN “OWNERSHIP”**

USAID guidance on evaluation in ADS 203.3.6 encourages stakeholder involvement in the evaluation process. It supports a decision-driven rationale for undertaking evaluations; states that stakeholders should be consulted to help prioritize evaluation questions; encourages clarity in an evaluation purpose statement with respect to the intended audience and how they will use the information an evaluation provides; encourages customer and partner participation in the evaluation process; and fosters the dissemination of evaluation reports to a wide audience. Supporting information on approaches and techniques for implementing this guidance is only provided in one of these areas; participatory evaluation, for which a USAID TIPS was created. Making improvements in the range of tools available to USAID staff for strengthening stakeholder “ownership” of evaluation processes and outcomes need not be expensive or time consuming.

- **Stakeholder Consultations**

The European Union has produced practical guidance on how to conduct evaluation stakeholder consultations that includes, among other things, suggestions on using the Delphi technique to help evaluation funding agencies and other stakeholders reach consensus. USAID could draw on this resource to create its own supporting publication or it might simply provide USAID staff with a link to the EU’s version at:

[http://ec.europa.eu/regional\\_policy/sources/docgener/evaluation/evalsed/sourcebooks/method\\_techniques/planning\\_structuring/stakeholders/description\\_en.htm](http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/planning_structuring/stakeholders/description_en.htm)

- **Joint Evaluations Guidance and Examples**

USAID’s ADS specifically identifies joint evaluations as one of the types of evaluation’s it approves. The Development Assistance Committee (DAC) of the OECD takes the position that joint evaluations, in which developing country representatives and other key stakeholders are equal partners, are important for expanding the ownership of the evaluation process and the utilization of its results in developing countries. The DAC recently issued a new “how to” guide for joint venture evaluation managers. USAID could draw on this resource to create its own supporting publication or it might simply provide USAID staff with a link to the DAC’s version. Further, USAID has participated in a number of joint evaluations, the best of which could be posted on its website in conjunction with other resources such as the new DAC guide at:

<http://www.oecd.org/dataoecd/58/36/37656526.pdf>

- **Participatory Evaluation Guidance and Examples**

USAID already has a TIPS on participatory evaluation, which was produced a number of years ago, which it may already be planning to update. In addition, USAID has undertaken a number of participatory evaluations over the years, the best of which could be posted to its website. Finally, USAID could improve the range of materials available on conducting participatory evaluations by simply posting links to other materials on this well documented approach to enhancing both “ownership” and evaluation capacity, particularly at the community level. A starter list of additional resources is available at:

<http://www.nwhf.org/areas/cbpr2006/documents/ParticipatoryEvaluationReadingList.doc>

**(2) DIRECTLY ENGAGE MULTIPLE SENIOR MANAGERS IN SETTING EVALUATION AGENDAS**

Over the past three decades, USAID has made efforts to engage its Administrators in the Agency’s evaluation process by working with them to develop annual evaluation agendas at the Agency level. Success with this process varied for two reasons: first, the level of interest of each successive Agency Administrator and second, the fact that some of the evaluations took so long to complete that the Administrator who put them on the agenda had left by the time they were published. While USAID may want to continue this effort, moving it down a level or two in the organizational structure could be more productive. There are two types of senior managers in the Agency that are best suited to identify evaluations capable of yielding important answers or having a broad impact, and who would be on post long enough to appreciate their results. The first are Mission Directors and Deputies, who often have questions about the relative effectiveness and impact of various aspects of their portfolios, but no independent mechanism for conducting these kinds of evaluations. The second group includes Technical Office Directors and Division Chiefs with sub-sector portfolios. Encouraging Mission and Technical Office managers to define, implement and complete even one serious evaluation per year, aimed at improving aid effectiveness, could go a long way towards fostering senior management “ownership” of evaluations and strengthening an Agency’s overall evaluation culture.

**(3) POST EVALUATION INTERVIEWS AND FEEDBACK**

As a support service to the Agency, the evaluation function has users both inside and beyond USAID. In the spirit of ADS 200.3.2.2., and to support a program of continuing improvement in evaluation, USAID’s new Evaluation Office might want to consider creating a customer feedback mechanism, e.g., a short survey form that could be sent to the heads of units that undertook evaluations in the preceding year asking them to provide information on their own satisfaction with the evaluation products they received. These individuals could also be asked to name one or two external stakeholders for each evaluation, who could be contacted concerning their satisfaction with USAID’s evaluation process and products. Posting an annual tally of the results of customer satisfaction surveys of this sort, would not only help M&E officers throughout the Agency understand how the work they support is perceived, but it could also facilitate a broad and open process through which future self-improvement agendas could be developed and implemented by M&E offices through USAID, not simply by the new Evaluation Office or only on a top-down basis. A USAID TIPS on Customer Service Assessments already exists and might serve as an in-house process of this sort.

## **C. STRENGTHENING EVALUATION QUALITY AND EVIDENCE**

Opportunities for learning from evaluation trends in other organizations and improving evaluation quality are numerous and are examined here in clusters.

**(1) USAID STAFF EVALUATION COMPETENCIES**

USAID is already pursuing a range of options for improving staff evaluation competencies. It has made significant progress over the past eight years in expanding evaluation training opportunities for its staff, and is

currently developing an on-line evaluation course to add to the Agency's offerings in this field. USAID would benefit from expanding access to educational opportunities and other resources to Implementing Partner M&E Officers. Additional steps the new Evaluation Office may want to take to support these efforts include:

- Periodic monitoring that ensures that all Agency M&E officers have taken Agency evaluation courses. USAID might also consider covering the cost of USAID direct hire M&E officers' membership in an evaluation association or the cost of non-USAID evaluation courses, particularly those that go beyond what USAID itself teaches.
- Guidance for Missions regarding appropriate minimum RFP/RFA M&E qualifications for Implementing Partner M&E officers, and
- An information campaign aimed at making USAID Implementing Partners field staff aware of USAID evaluation resources that are available to them. This includes the new on-line training course once it is made available, and trainings available abroad through national and regional evaluation networks, other donors and local institutions. USAID might also consider allowing Implementing Partners to cover the costs of locally available M&E training for host national field staff, as these individuals often move from one USAID project to another over a number of years.

## (2) **EVALUATION DESIGN AND STRONG EVIDENCE**

As the review of trends in the previous section indicates, recent critiques of donor agencies evaluations, which focus on the strength of the evidence they provide concerning program and project impact, generally find them to be inadequate. USAID, as indicated in the previous section, already undertakes rigorous impact evaluations through some of its technical offices, i.e., health and education have a longer history in this regard than newer initiatives such as micro-enterprise and democracy and governance programs. Further, its newly revised ADS 203 highlights the importance of rigor in USAID evaluations. Beyond this, USAID's new Evaluation Office has a number of options for fostering evaluations that provide stronger evidence of impact.

- **Acknowledge the Validity of Discussions About Strengthening Evaluation Evidence**

While USAID may not need to be ready to change its evaluation policies to explicitly address evaluation rigor issues, enough people are aware of public critiques of development agencies on this issue to make it likely that some statement from the new Evaluation Office would be helpful. Even if the statement only indicates that USAID is looking into where and how frequently more rigorous evaluations would benefit the Agency.

- **Spot Existing Rigorous Evaluation Initiatives**

USAID's experience conducting rigorous impact evaluations, and new impact evaluation initiatives underway in various technical offices, are not familiar to all USAID staff or even all USAID M&E officers. The new Evaluation Office is well positioned to publicize these efforts as examples of the type of rigorous work its ADS encourages.

- **Expand USAID Guidance Related to Rigorous Evaluation Approaches**

Guidance is lacking on when rigorous evaluations are most beneficial to USAID. Developing a consensus in the Agency on this question, and publishing it, would help staff throughout the Agency understand when to consider more rigorous approaches. This kind of a document could clarify whether USAID only needs these kinds of studies when testing novel hypotheses, or also when

moving a model that worked in one country to other locations. The incorporation of rigorous evaluations in all pilot projects could also be addressed in this type of guidance.

In addition, USAID could supplement new guidance on when to use more rigorous methods with “how to” materials on field experiments and quasi-experimental designs, as well as a number of quasi-experimental econometric approaches that do not necessarily require new fieldwork. Resources that already exist on these topics could be made available on USAID’s website, or new TIPS on these topics could be developed. Alternatively a number of different rigorous techniques might be covered in a single but somewhat longer compendium.

- **Directly Invest in Rigorous Evaluations**

There are several ways in which USAID’s new Evaluation Office could invest in more rigorous evaluations. One option would be to help technical offices and missions identify areas where they believe they would benefit from rigorous evaluations. The next step would be to help to carry them out, either through direct financial support or access to expertise through the Evaluation Office, i.e., a centrally-funded rigorous evaluation support contract, could help in this regard. Another option would be to establish a sub-unit inside the new Evaluation Office or under the Agency’s Chief Economist that conducts certain types of rigorous studies, e.g., quasi-experimental econometric analyses that other offices in USAID might not have the expertise to carry out. On this point, the World Bank chose the Chief Economist to oversee most of its econometric evaluations.

Before committing the new Evaluation Office to conducting evaluations, particularly rigorous ones, USAID needs to consider the trade-offs this might involve. Prior to 1980, USAID’s Evaluation Office did not conduct evaluations. When it started conducting evaluations in the 1980s, those efforts drew staff attention away from other aspects of the Office’s evaluation system management function. In the early 1980s, when few evaluations made an effort to determine how successful projects were beyond the output level, this made sense as a demonstration effort. Today, questions about whether this continues to make sense need to be addressed as decisions about the structure and function of the new Evaluation Office are made.

### **(3) EVALUATION PLANNING AND STATEMENTS OF WORK (SOWS)**

While USAID has not undertaken a special study that examines the relationship between evaluation planning and evaluation Statements of Work (SOWs), anecdotal evidence from evaluation contractors and USAID staff suggests that SOWs do not adhere to ADS guidance with respect to limiting evaluation SOWs to a small number of questions. SOWs also indicate a lack of attention to the relationship between the number of questions asked, study duration and study budgets. This negatively affects the evaluation’s quality.

USAID has options for making improvements in this regard, including updated guidance to M&E officials, Program Officers, and Contract Officers on how to guide the development and assess the adequacy of evaluation SOWs. Simple tools, such as an SOW review checklist that follows the ADS requirements could be helpful. Likewise, a periodic meta-review of evaluation SOWs by the new Evaluation Office and the dissemination of information on the percentage of SOWs that complied with ADS guidance, or scored well on a checklist, could be helpful. Further guidance to these officers concerning an appropriate balance in an SOW’s evaluation schedule would be of great use. A schedule should balance preparation time (i.e., review of existing monitoring data, instrument development and testing, preparation of an inception report); field data collection; data analysis; and report preparation. Another useful option would be a short course on evaluation SOWs for Contract Officers. Providing USAID staff with examples of good evaluation SOWs would go a long way in this regard.

#### **(4) EVALUATION REPORTS**

Trends in other donor agencies are demonstrating the merits of doing more than simply defining what should be included in evaluation reports. While the UNDP has moved furthest in the direction of rating evaluations against its evaluation report criteria, other donors are also looking more closely at evaluation report quality as an indicator of the likelihood that reports will be used. Options exist for moving in this direction in USAID. In the evaluation course it provides for staff, a checklist is already used to score evaluation reports, similar to what the UNDP is doing. Participants are encouraged to use this tool when they return to their jobs. The new Evaluation Office could move further in this direction by recommending the use of a tool of this sort to Missions and technical offices, to be used to score drafts and provide feedback to evaluation authors. In line with UNDP, USAID might also want to consider scoring evaluation reports that are sent to the DEC each year. Regardless of whether this was done on a comprehensive or sample basis, it would provide the new Evaluation Office with a basis for issuing annual “feedback” reports on evaluation report quality. It could also be used in setting and monitoring Agency targets for improvement, in line with what a number of other donors are already doing.

#### **D. IMPROVING EVALUATION FOLLOW-UP**

As the trend review in the previous section demonstrated, donors increasingly view effective post-evaluation action as critical for ensuring that evaluations have a discernable impact on aid effectiveness. While USAID’s ADS 203.3.6.7 calls for post-evaluation reviews and follow-up action, USAID does not currently monitor how comprehensively this aspect of its evaluation guidance is being implemented or how action on evaluation recommendations is affecting development programs and projects.

##### **(1) IMPROVE PRACTICAL GUIDANCE FOR CONDUCTING POST-EVALUATION DECISION-MAKING REVIEWS**

Anecdotal information from USAID staff indicates that Missions and headquarters offices vary widely in terms of their adherence to expectations about steps to be taken after evaluations are completed. This is particularly true with regard to evaluations aimed at ensuring evaluation utilization, by:

- determining whether USAID accepts and supports each evaluation finding, conclusion and recommendation;
- identifying actions that need to be taken to implement recommendations that are accepted, or which USAID on its own decides are necessary;
- assigning responsibility for carrying out those actions;
- establishing a timeline for their completion; and
- following up on whether those actions are fully implemented in a timely manner.

When USAID staff members are asked about how their Missions and offices carry out these steps, some describe well-established processes, while others report that post-evaluation reviews are carried out on an *ad hoc* basis or not at all. More complete guidance on how post-evaluation reviews are to be conducted and tools for recording the outcomes of such reviews could help Missions and other offices to be more systematic in this regard. Questions have been raised in evaluation courses for USAID staff, that discuss this process and that could be answered in a simple how to guide. Such questions include: (a) Who should call for and chair evaluation review meetings, e.g., the Mission Director or Deputy, or the Program Office, or technical team leaders?; (b) How soon after receipt of the final version of an evaluation report should they be held?; (c) How should disagreements be resolved regarding which recommendations to accept?; (d) What format should be used to record decisions?; and (e) Who is responsible for following-up on whether agreed-upon actions are taken and, if they are not taken, to whom should such failures to act be reported?

## **(2) MONITORING POST-EVALUATION DECISIONS AND COMMITMENTS TO ACT ON EVALUATION RECOMMENDATIONS**

For many years prior to USAID's reengineering exercise in the early 1990s, the Evaluation Office in Washington knew whether post-evaluation reviews were being undertaken, and had enough information on hand to calculate what percentage of evaluation recommendations USAID adopted, should it have wanted to monitor evaluation practice in that way. USAID Form AID 1330-5 (10-87), which Missions and technical offices included when they submitted an evaluation report to the Evaluation Office, documented the operating units post-evaluation review decisions. An example of this report is provided in Annex B. USAID dropped its requirement for this post-evaluation review report in the early 1990s, during its reengineering exercise.

As a result, USAID's new Evaluation Office has no systematic way of monitoring whether operating units are carrying out post-evaluation reviews, or with what frequency they accept and commit to acting on evaluation recommendations. Anecdotal evidence from participants in evaluation classes provided for USAID staff, indicates that operating units vary considerably with respect to whether they routinely hold post-evaluation review meetings or not, and those that do vary in terms of the formality of the process and the level at which such review meetings are held.

One option for encouraging more consistent adherence to USAID's guidance on post evaluation reviews would be to reintroduce USAID Form AID 1330-5 (10-87) or a variation thereof. Doing so would help to standardize the post-evaluation review process in the field; it would also provide the new Evaluation Office with information that could be used to calculate, as the World Bank is now doing, the percentage of evaluation recommendations the Agency is accepting and committing to act upon. This information not only serves to demonstrate the value of evaluations to users, but can also be used to probe why that percentage is low, if initial results show that to be the case. The World Bank's investigation of why its adoption rate was low led that organization to discover that lack of specificity and practicality in evaluation recommendations was an issue. Through subsequent efforts, the Bank has been able to correct that problem and improve its adoption rate for evaluation recommendations, which it views as a critical first step towards ensuring that evaluations lead to improvements in aid effectiveness.

## **(3) MONITORING IMPLEMENTATION OF AGREED UPON POST-EVALUATION ACTIONS AND THEIR EFFECTS**

Monitoring follow-through on decisions to act on evaluation recommendations is a natural extension of efforts that DFID and UNDP are making to ensure that their evaluations have an impact on development. Adequate information on what decisions are made in a post-evaluation review meeting is a pre-requisite for post-evaluation action monitoring systems that track the percentage of accepted recommendations that are actually implemented. USAID does not currently have information on whether post-evaluation reviews are being held, or on what post-evaluation action commitments its operating units have made. Due to this, follow-up monitoring on the degree to which actions are implemented is not an option in the near term. Should USAID want to move in this direction over time, UNDP's Management Response Dashboard through which the status on post-evaluation action on recommendations is tracked, might be worth examining. It can be found through the following address:

[http://cfapp1-docs-public.undp.org/eo/evaldocs1/eo\\_doc\\_979011621.ppt](http://cfapp1-docs-public.undp.org/eo/evaldocs1/eo_doc_979011621.ppt)

Another option that USAID might want to explore further is documenting what happens once evaluation recommendations are implemented. The World Bank's 2004 publication, *Influential Evaluations*, represents a ground-breaking example of what can be learned from following up on post-evaluation actions. In some of the case studies in this volume, comparisons of the cost of evaluations to the downstream benefits realized when evaluation recommendations were implemented were impressive. Others demonstrated how evaluations improved the development impact of projects and programs. Feedback of this sort may be

particularly valuable in organizations where the priority accorded to evaluation by mid-level managers is lower than senior Agency managers believe it should be.

## ANNEX A.

### PATTERNS IN RECENT USAID EVALUATIONS IN THE DEVELOPMENT EXPERIENCE CLEARINGHOUSE<sup>7</sup>

2005-2008 Universe		Frequencies and Percentages			
Total documents in the DEC coded "evaluation" (January 2005 – October 2008)		1,303			
<b>Evaluations and Other Types of Documents (N = 1,303)</b>					
<b>Verified as "evaluations"</b>		<b>296</b>	23%		
Called "evaluations"				234	79%
Called "assessments"				62	21%
<b>Other types of documents<sup>8</sup></b>		<b>1,007</b>	77%		
<b>Basic Evaluation Characteristics (N = 296)</b>					
Year	2005			116	39%
	2006			82	28%
	2007			73	25%
	2008 (through 3 <sup>rd</sup> quarter)			25	8%
Initiating Bureau	Africa			75	25%
	GH			55	19%
	Asia			52	18%
	E&E			32	11%
	DCHA			30	10%
	LAC			25	8%
	ME			14	5%
	EGAT			11	4%
	PPC			2	1%
Topic	Health			112	38%
	Democracy/Governance			37	13%
	Economic Growth <sup>9</sup>			32	11%
	Education			30	10%
	Agriculture			23	8%
	Environment			13	4%
	Other or multi-sector			49	16%
Type	Formative			137	46%
	Summative			126	43%
	Both			33	11%
Scope	Single Project, Single Country			220	74%
	Single Project, Multiple Countries			43	15%
	Multiple Projects in One Country			21	7%
	Multiple Project & Countries			10	3%
	Meta-evaluations			2	< 1%

<sup>7</sup> Evaluation holdings in the DEC are widely believed to be incomplete.

<sup>8</sup> Frequently "other" means: pre-project assessments (703); other project documents (127) or GAO reports (77), and other evaluation documents, e.g., handbooks, best practices documents, etc. (41).

<sup>9</sup> Financial sector, private sector competitiveness, economic opportunity, trade, etc.



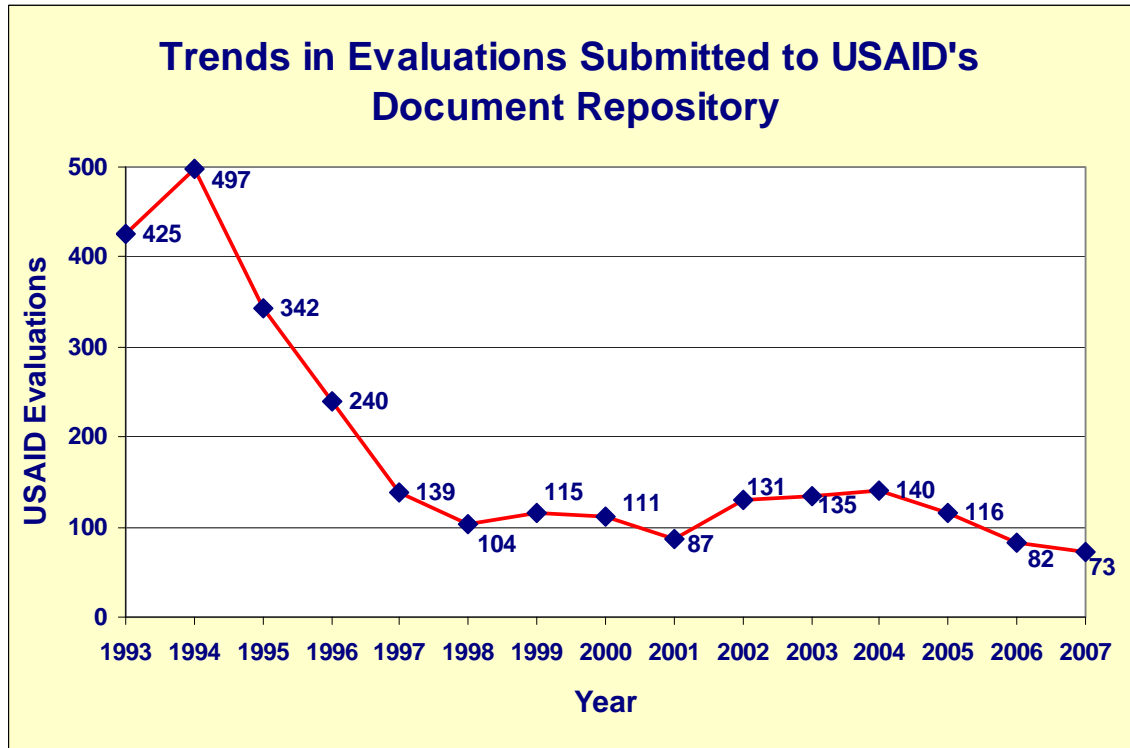


Figure I. USAID's "Evaluations in the DEC" Trend Line Updated<sup>10</sup>

#### Purposes of Recent Evaluations in the DEC (2005-2008)<sup>11</sup>

The most common purposes for 296 evaluations were:

- Determine project effectiveness, i.e., in achieving Outputs, Intermediate Results (42 percent)
- Facilitate the design of future projects (42 percent), of which roughly half were expected to contribute to the design of follow-on to the project that was evaluated.
- Determine whether there is a need to modify/redesign an existing activity (31 percent)
- Determine degree of attainment of higher level results, e.g., Intermediate Results, Strategic Objectives, Goals (30 percent)

<sup>10</sup> Sources:

**1993-1997:** Katrina Greene, "Narrative Summary of FY97 Evaluations," SAID/R&Rs, 7 January 1999.

**1998:** Cindy Clapp-Wincek & Richard Blue, Evaluation of Recent USAID Evaluation Experience, USAID/PPC/CDIE Working Paper 320, 2001.

**1999-2004:** Janet Kerley USAID/CDIE.

**2005-2008:** MSI. [Note: 2008 data is not included in the chart as the year has not come to a close making the data incomplete.]

<sup>11</sup> Percentages should not be added, since some evaluations had multiple purposes

### Other Characteristics of Recent Evaluations in the DEC (2005-2008)

- All 296 evaluations reported that some data was collected from individual interviews.
- 114 (39 percent) used a survey to collect data.
- 156 evaluations (53 percent) reported using group interviews to collect data, of which 21 conducted community interviews/town meetings and 93 reported using focus group techniques.
- 98 evaluations (33 percent of 296) reported that they used structured observation techniques to gather data about individuals or groups.
- 66 evaluations (22 percent of 296) were participatory at some level, involving beneficiaries and/or stakeholders in roles in the evaluation beyond that of informant. Of these 12 involved beneficiaries, 34 involved stake holders and 20 involved both. Looking at those that involved beneficiaries or stakeholders, but not both, their roles were somewhat different.

	Worked on SOW	Data Collection & Analysis Role	Worked Recommendations
Beneficiaries (12)	0	9	5
Stakeholders (34)	5	28	23

- 27 evaluations (9 percent) of the evaluations reported data on a comparison group as well as the project target group, of these nine included both pre and post intervention data for the comparison group. One was an experimental design involving randomized assignment.
  - The UJYALO program was a three year community peace promotion project, funded by USAID/Nepal. A quasi-experimental design was used: the project intervention was carried out in 13 districts; other districts from which baseline data had been collected served as comparison districts. Post intervention data was collected for 742 households in treatment districts and 209 comparison districts. Basis for project’s selection of treatment districts not discussed in the evaluation report.
  - An evaluation in Kenya examined an on-site training program for reproductive health care supervisors that used a true experimental design with pre- and post-test measures and randomly assigned 60 health facilities into intervention and control groups.

MHageboeck and MFrumin, MSI, 3/6/09

ANNEX B.<sup>12</sup>

SAMPLE USAID EVALUATION SUMMARY (FORM AID 1330-5 (10/87))

13N 96116  
A.I.D. EVALUATION SUMMARY - PART I PD-ABL-667

1. BEFORE FILING OUT THIS FORM, READ THE ATTACHED INSTRUCTIONS.  
2. USE LETTER QUALITY TYPE, NOT "DOT MATRIX" TYPE.

IDENTIFICATION DATA					
A. Reporting A.I.D. Unit: USAID/Indonesia		B. Was Evaluation Scheduled in Current FY Annual Evaluation Plan? Yes		C. Evaluation Timing INTERIM	
Mission or AID/W Office <u>Agro-Enterprise and Environment (AEE), USAID/Indonesia</u> (ES# <u>  </u> )		Yes <input type="checkbox"/> Slipped <input checked="" type="checkbox"/> Ad Hoc <input type="checkbox"/> Evaluation Plan Submission Date: FY <u>94</u> Q <u>2</u>		Interim <input checked="" type="checkbox"/> Final <input type="checkbox"/> Ex Post <input type="checkbox"/> Other <input type="checkbox"/>	
D. ACTIVITY or Activities Evaluated (list the following information for project(s) or program(s) evaluated; if not applicable, list title and date of the evaluation report.)					
Project No.	Project/Program Title	First PROAG or Equivalent (FY)	Most Recent PACD (Mo/Yr)	Planned LOP Cost (000)	Amount Obligated to Date (000)
497-0382	Natural Resources Management Project	July 27, 1990	July 31, 1997	US\$20,500,000	US\$11,468,069
ACTIONS					
E. Action Decisions Approved By Mission or AID/W Office Director			Name of Officer Responsible for Action		Date Action to be Completed
<p>Action(s) Required</p> <p>a. Revise the project's <sup>logical</sup> framework to reflect recommended changes in the role of the Policy Secretariat in expanding a national network of policy analysts and supporting policy task forces that have been organized within the Ministry of Forestry (MOFr). These changes have been necessitated by the difficulties encountered in attempting to establish a self-sustaining group of policy analysts within the framework of the Secretariat. As a result, some of the project resources during the remainder of the project will be redirected to encompass a wider range of potential analysts. While strengthening the ongoing network of policy analysts for cross-sectoral issues in nature resource management, part of the resources to foster the development of a self-sustaining policy center within the Secretariat will be used to strengthen the policy task-forces that have been organized around project activities by the MOFr. Concurrently, the broader network of policy analysts that has been organized will be expanded by: 1) providing greater opportunities for on-the-job training for research assistants in carrying out the Secretariat's studies agenda; 2) developing more effective means for network participants to gather and communicate with each other; and 3) initiating on-going exchanges with policy groups in the U.S.</p> <p>b. Expand the project's pilot activities in industrial waste minimization by providing supplemental funding for the project's Cleaner Industrial Production (CIP) Initiative. Expanded activities will be focused on: 1) identifying cost effective measures which provide effective opportunities for private sector investments in pollution abatement strategies in a number of industries; 2) training Indonesian counterparts in the methodologies of planning and implementing environmental auditing procedures; and 3) strengthening government policies to accelerate the adoption of pollution abatement technologies by promoting greater environmental awareness through organized workshops, study tours, and training courses. Scope of the activity will be dependent upon availability of funds.</p> <p>c. Assess the impacts of the project's policy recommendations on women. Measurable impacts on women will be identified and incorporated into an on-going analysis of project activities that is intended to promote a broader discussion of the consequences of pursuing alternative resource policy options among government agencies, the private sector, NGOs, local communities, and other segments of society.</p>			<p>- Project Officer - Agus Widianto, Project Manager</p> <p>- Alfred Nakatsuma, Project Officer - Adhiyana, Project Manager</p>		<p>August 1995</p> <p>August 1995</p> <p>September 1995</p>
APPROVALS					
F. Date of Mission Or AID/W Office Review of Evaluation: June - July 1994			(Month)	(Day)	(Year)
G. Approvals of Evaluation Summary And Action Decisions:					
Name (Typed)	Project/Program Officer	Representative of Borrower/Grantee	Evaluation Officer	Mission or AID/W Office Director	
Signature	Agus Widianto, Acting	Prof. Dr. Herman Haeruman	Patricia Chaplin	Vivikka Mollidrem	
Date	7/20/95	7/21/95	7/20/95	7/24/95	

<sup>12</sup> The documents in Annex B are originals that have been replicated for use in this paper.

## ABSTRACT

### H. Evaluation Abstract (*Do not exceed the space provided*)

The Natural Resources Management (NRM) project has two objectives: a) to identify key policies and practices which constrain sustainable economic growth; and b) to design and implement improved policies and practices to address those constraints. The seven year project started in 1990, and has entered its fourth year of implementation. Significant progress has been made, but many problems have been encountered. To enable project personnel to review project progress, document significant achievements, and resolve problems, a mid-term evaluation was needed. Indicators in the logical framework were used to measure achievements. Results of the evaluation will be used as a basis for developing a strategy to implement "second-phase" project activities.

Overall achievements of the project have been notable. An effective technical assistance team is in place. Eight of twelve policy studies have been completed and the results widely disseminated. Follow-on steps have been taken by the Ministry of Forestry as a result of policy studies, and the Ministry has established task-forces around project activities, including logging waste minimization and traditional forest area management. Two draft management plans for protected areas, developed with full participation with local communities, are nearing completion. Eighteen students have been sent to the U.S. for MSc degree training with provisions for some returning to Indonesia to conduct their thesis research; three ongoing PhD students have been supported to complete their programs; 522 person days of short-term international training has been provided for 299 participants, and over 4,000 person days of in-country training has been carried out. Pilot activities in industrial waste minimization have been conducted in factories, and new technologies have been adopted by some of the plants. Despite the project achievements described above, however, the evaluation team observed that project design was too ambitious — trying to cover various complicated issues with the involvement of a wide range of key players with limited resources or project staff. In general, the team recommended that in the second phase of project implementation, the project staff should focus on priority issues achievable by the project, including developing appropriate policies for reducing logging wastes, and incorporating local communities into resource management planning activities.

The evaluation team was selected through a buy-in with the Environmental and Natural Resources Policy and Training (EPAT) project. The team carried out the tasks by reviewing project documents, visiting project sites, and interviewing project-related staff in Jakarta, as well as in Bogor, Pontianak, Bukit Bakau, Bukit Raya National Park, Manado and Bunaken National Park, and visiting factories in Bandung, Jakarta and Surabaya.

#### Lessons learned:

- Need to consolidate and disseminate results of policy studies and field experiences to encourage discussions of critical resource issues, exchange information/experience, and coordinate activities.
- Need to continue developing constructive policy options by integrating field-tested pilot site activities with policy analysis at the national level.
- Need to develop management plans for protected areas through extensive consultations with stakeholders to achieve consensus on the purpose of the protected area, and by developing mechanisms for community participation in planning, implementing and monitoring management activities.

## COSTS

### 1. Evaluation Costs

Name	1. Evaluation Team Affiliation	Contract Number OR TDY Person Days	Contract Cost OR TDY Cost (U.S. \$)	Source of Funds
<b>1. Evaluation Team</b>			<b>\$114,321</b>	<b>NRM Project</b>
John Clark (Team Leader)	University of Miami	45 person-days		
Peter Burbridge	Independent Consultant	42 person-days		
M. Soerjani	University of Indonesia	42 person-days		
H. Soeparwadi	Bandung Inst. of Tech.	36 person-days		
Krisnawati Suryanata	University of California, Berkeley	36 person-days		
Richard Tobin	EPAT	15 person-days		
<b>2. Mission/Office Professional Staff</b>			<b>\$7,600</b>	<b>Operating Expense (OE)</b>
Dennis Cengel		21 person-days		
Ketut Djati		21 person-days		
Agus Widianto		10 person-days		
<b>3. Borrower/Grantee Professional Staff</b>			<b>\$3,350</b>	<b>NRM counterpart budget</b>
Indah Dianti (BAPPENAS)		7 person-days		
Ichwanuddin (BAPPENAS)		4 person-days		
Widarya Noer (MOFr)		5 person-days		
Afrizal (MOFr)		5 person-days		
<b>2. Mission/Office Professional Staff</b> Person-Days (Estimate) <u>51 person-days</u>		<b>3. Borrower/Grantee Professional Staff</b> Person-Days (Estimate) <u>21 person-days</u>		

## A.I.D. EVALUATION SUMMARY - PART II

<b>SUMMARY</b>				
<p>J. Summary of Evaluation Findings, Conclusions and Recommendations (Try not exceed three (3) pages provided) Address the following items:</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 50%; border: none; vertical-align: top;"> <ul style="list-style-type: none"> <li>• Purpose of Evaluation and methodology used</li> <li>• Purpose of activity(ies) evaluated</li> <li>• Findings and conclusions (relate to questions)</li> </ul> </td> <td style="width: 50%; border: none; vertical-align: top;"> <ul style="list-style-type: none"> <li>• Principal Recommendations</li> <li>• Lessons Learned</li> </ul> </td> </tr> </table>			<ul style="list-style-type: none"> <li>• Purpose of Evaluation and methodology used</li> <li>• Purpose of activity(ies) evaluated</li> <li>• Findings and conclusions (relate to questions)</li> </ul>	<ul style="list-style-type: none"> <li>• Principal Recommendations</li> <li>• Lessons Learned</li> </ul>
<ul style="list-style-type: none"> <li>• Purpose of Evaluation and methodology used</li> <li>• Purpose of activity(ies) evaluated</li> <li>• Findings and conclusions (relate to questions)</li> </ul>	<ul style="list-style-type: none"> <li>• Principal Recommendations</li> <li>• Lessons Learned</li> </ul>			
<p>Mission or Office: Office of Agro-Enterprise and Environment (AEE)</p>	<p>Date This Summary Prepared: May 9, 1995</p>	<p>Title and Date of Full Evaluation Report: Mid-term Evaluation of USAID/Indonesia's NRM Project April 1995</p>		
<p>As the Natural Resources Management (NRM) project entered its fourth of seven years of implementation, a mid-term evaluation was conducted to evaluate progress towards achieving specified outputs in the logical framework (logframe) and to recommend Improvements in project design and/or implementation to facilitate accomplishment of project outputs.</p> <p>NRM project purpose stated in the logframe is: "To strengthen selected Indonesian institutions so that they can: a) Identify natural resources management policies and practices which constrain economic growth; and b) design and implement improved policies and practices to address these constraints". The evaluation team assessed eight (8) key Issues: 1) achievements of project purpose and outputs; 2) project design; 3) policy analysis; 4) technical assistance performed and future needs; 5) the cost-effectiveness of a management information system; 6) the role of community groups; NGOs, and private sector interactions; 7) women in development, and 8) two other concerns related to incorporation of environmental considerations into tourism development and implementation of an environmental action plan related to environmental assessments of the NRM project.</p> <p>The evaluation team reviewed available project documents and discussed and clarified the team's tasks with the USAID NRM Project Committee. The team then split into several sub-teams to cover policy analysis in Jakarta, field activities in Bukit Bakau Bukit Raya in West/Central Kalimantan, and Bunaken National Park in North Sulawesi. Two meetings were held to present progress and findings of the team. These were held in the middle of the evaluation and at a final presentation. Presentations were made to GO1 and USAID staff.</p> <p>The following are issues assessed by the evaluation team, specific project accomplishments that were recognized, and constraints and recommendations that were made:</p>				
<p><b>I. POLICY SECRETARIAT</b></p> <p><b>1.1. Policy Secretariat.</b> Eight of twelve planned macroeconomic studies have been completed and they all will soon have been presented to GOI officials and representatives of other groups and agencies. A series of seminars and workshops have been held to promote discussion of the issues that were raised in the studies and support dissemination of the results. One study has achieved results well beyond the analytical phase of the endeavor, focusing on the dynamic Input-output modeling. Constraints that were identified included the need for more time and resources to support the complex process of catalyzing policy reforms. Recommendations are that: a) the project's NRM Policy Secretariat should continue to respond to critical demand-driven requests from BAPPENAS; b) the number of policy studies should be reduced from 15 to 12 by deleting the Marine Sector study (insufficient project resources), and combining studies on integrated development and community development; and c) the results and recommendations</p>				

of studies that have been completed should be consolidated into summary documents highlighting prescriptive policies.

**1.2. Support Unit.** The team assessed the Secretariat's tasks in supporting the Policy Working Group (PWG) in developing policy options for natural resources management. Support is provided by the contractor using a management information system designed and implemented to provide a link between information and project management, document impacts of project activities, and monitor impacts of project activities in achieving its purposes and goals. Constraints identified included lack of fulltime counterpart staffing within the Policy Secretariat. The evaluation team recommended that the project strengthen ongoing networking activities among natural resource policy analysts by reallocating resources to the policy task forces that have been organized around project activities by the Ministry of Forestry, enhancing training opportunities for research assistants in implementing the Secretariat's studies agenda, facilitating meetings of the policy network's members, expanding exchanges with policy groups in the United States, and initiating discussions of alternative means of providing policy advice to affected government ministries after the completion of the project.

**1.3. Training.** Training achievements under the NRM project include: 30 research assistants under in-service training in the Policy Secretariat; 18 students sent to the U.S. for graduate degree training; 3 ongoing PhD students in the U.S. supported to complete their programs; 19 participants sent to Asia, and 23 to the U.S. under the project's International short-term training program; and over 4,000 person days of in-country training provided. The team noted the opportunity to better integrate research conducted under the academic training component with project activities. The team recommended that the TA team, in consultation with the PWG and Project Coordinating Committee (PCC), expand the information currently provided to degree participants regarding project activities, and provided a potential list of research topics for consideration by degree participants.

**1.4. Dissemination of results.** NRM project short- and long-term advisors have presented their findings in open forum workshops. Supplementary publications have been prepared and disseminated, including: project brochures; abstract series of reports; newsletters; and policy briefs. The evaluation team recommended that issuance of policy briefs should be based on the audience that is being targeted and its format consistent with the purposes of reaching that audience.

## **II. BUKIT BAKA – BUKIT RAYA AND BUNAKEN FIELD SITES**

**II.1. National Park Management.** A draft management plan was completed by the PHPA (Perlindungan Hutan dan Palestarian Alam/Forest Protection and Nature Conservation) office assisted by project advisors, and discussed with the provincial Ministry of Forestry (MOFr) in West Kalimantan. Constraints include the lack of boundary delineation and zonation of the park. The team recommended that ARD (Associates in Rural Development) hire a park planner to be team leader and to revise and complete the draft management plan, including production of maps to establish management zones.

**II.2. Management of natural production forests.** Accomplishments include: short-term consultancies working with PT. Sari Buma Kusuma (a collaborating concessionaire) in road construction and damage control logging; biodiversity conservation research in logging areas, including production of tree identification manual by Harvard University consultant; recommendations accepted by the MOFr to reduce the annual allowable cut from 31 to 22 million cubic meters/year, and to increase forest royalty charges; and support for development of specific forestry policies by studies and field activities on reduction of logging waste, regulation of harvesting plans, and guidelines for establishing biodiversity areas in natural production forests. The team recommended that the NRM Project strengthens ongoing donor collaboration even more, particularly with respect to applying the forest management planning documents prepared as part of the project to the permanent forest area management units being developed by the MOFr with the assistance of ODA.

**II.3. Community Development and Participatory Management.** The NRM Project has achieved measured progress in increasing the economic and social welfare of local communities in the Bukit Baka-Bukit Raya area. The project has provided water supply systems, started vegetable gardens and nursery

programs, supported Institutionalization of traditional laws and regulations by strengthening local community advisory councils, and developed a concept of and prepared initial plans for a Traditional Forest Area (TFA). A particular problem exists with respect to strengthening local institutional capability to carry out program activities after the project. The evaluation team recommended that the Institutional Development and Agroforestry Advisors help promote participation of local NGOs in three areas by: a) consolidating the strength of Village Advisory Councils through training programs and active participation in park planning and management activities; b) extending agroforestry techniques; and c) monitoring the progress of village-managed water supply systems.

**II.4. Project Coordination.** Implementation of project activities at the Bukit Bake - Bukit Raya site has been difficult because of its remoteness, and its status of a Class Two security zone. There have also been problems associated with the inappropriate choice of long-term technical advisors fielded at the site, and limited MOFr counterpart participation. The evaluation team is aware of significant progress made in managing/replacing technical advisors and coordinating project activities. However, the lack of counterpart support in the field has been a significant issue in coordinating project activities. Recommendations include continuing the Forum Komunikasi (Coordinating Forum) started under the project, and using this mechanism to address the need for the MOFr to allocate travel funds for counterpart staff to travel to field sites on a regular basis.

**II.5. Training.** As of April 1994, 748 persons in West and Central Kalimantan had received over 2,750 person-days training. Training programs included use of GPS, database management, and techniques for implementing biodiversity surveys.

**II.6 Forestry Research and Training Station.** Under the parallel Sustainable Forest Management Project, supported by the international Tropical Timber Organization (ITTO) and the MOFr, the Kaburai Biodiversity Research and Training station will be constructed at the site. USAID has provided a mini-hydroelectric power and potable water supply system (completed in May 1993), and plans to procure some research equipment for the station. Construction of the station has been delayed as a result of contracting problems. Accordingly, the evaluation team recommended that USAID provisions of project equipment be contingent upon completion of the station.

### III. COMPETITIVE AWARDS SYSTEM (CAS)

A short-term consultant developed a conceptual framework for implementing the project's Competitive Awards System (CAS) for Applied Field Research in early 1993. Meetings and seminars were held to introduce and develop the concept. An agreement was reached that, in 1994, the CAS committee would solicit proposals from West and Central Kalimantan. As a result of delays in the timetable within the Ministry of Forestry, however, the team recommended that the efforts to initiate the awards should be intensified, and a smooth transition be made to help the MOFr maintain the system after completion of the NRM Project.

### IV. BUNAKEN PILOT SITE

**IV.1. National Park Management and Community Participation.** The draft (in Bahasa Indonesia) of the twenty-five year Master Plan, five-year Development Plan, and one-year Operational Plan for the park have been completed and submitted to the MOFr/PHPA, Bogor for review. The accomplishments of the TA team and GOI counterparts in obtaining a consensus on the draft plan were thought by the team to be impressive. The consensus integrates dissimilar demands of various resource users into what the team called a cohesive and ecologically sound management plan. The project team is currently assessing the prospects for expanding the role of Manado-based NGOs as advocates for local community interests.

The evaluation team recommended: **first**, that the project should assist PHPA in developing specific guidelines and policies for marine protected areas by addressing the following issues: a) improving PHPA's Guidelines for National Park Management Plans by including marine areas; b) developing mechanisms for establishing a generic participatory planning process; and c) developing mechanisms for implementing the participatory management plan that allocates responsibilities for managing the park's



resources, provides measures for the equitable distribution of benefits, and induces collective compliance with the park's regulations. **Second**, the team recommended that a short-term consultant be recruited to develop a strategic plan for sustainable use of mangrove forest resources in the park. A **third** recommendation suggested that PHPA designate Bunaken National Park as a pilot site for testing the effectiveness of the Swadana (self-supporting financing) mechanism for supplementing park funding, while invoking the principle of Swakelola (self-management). A **fourth** recommendation was to continue testing the effectiveness of the Forum Komunikasi (Communication Forum) to sort out competing interests of the various groups of users of the park's resources. A **fifth** recommendation was to incorporate environmental considerations into plans for developing ecotourism, by: a) establishing an understanding and a consensus opinion on the "acceptable levels of change" of the park's resources; b) cooperating with the certifying organization for diving instructors to train and certify dive guides in marine conservation issues; and c) coordinating compliance with "acceptable levels" of visitation.

**IV.2. Training.** As of July 1994, 194 participants had received training, much of it designed to support project implementation activities. The evaluation team applauded the training that was provided to field assistants, particularly in participatory rural appraisal techniques, and environmental education.

**IV.3. Donor Collaboration.** Ongoing collaboration with CIDA (Canadian International Development Agency) should be expanded to embrace the development of an Integrated coastal zone management plan to control buffer zone activities adjacent to the park's boundary.

## V. KEY ISSUES IN THE ENVIRONMENTAL ASSESSMENT

An environmental assessment was conducted for the NRM Project in late 1992. It produced nearly 50 recommendations designed to improve the project's implementation and to mitigate potentially adverse environmental impacts. The mid-term evaluation team, with USAID/I concurrence, concentrated attention on seven priority issues as part of the assessment team's recommendations. These included: 1) the possibility of NRM Project assistance to PT. Kurnia Kapuas Plywood (PT. KKP); 2) the possible exclusion of PT. Sari Bumi Kusuma (SBKI) from the mandatory industrial timber plantation for transmigration (HTI-Trans) program; 3) the promotion of community participation at Bukit Baka - Bukit Raya National Park; 4) studies on the legal status of park residents; 5) the assessment environmental implications of the Bina Desa (Concessionaire Community Development) program, 6) the promotion of local management of natural resources by facilitating development of local skills; and 7) the promotion of a community-based fire control program.

The team's recommendations were to: 1) collaborate with the ITTO (International Tropical Timber Organization)-supported Sustainable Forest Management Project (SFMP) in activities related to community development, including communities receiving assistance from PT. KKP; 2) provide constructive inputs for policy changes in excluding PT. SBKI from a HTI-Trans program at Bukit Baka - Bukit Raya; 3) include in the management plan for Bukit Baka - Bukit Raya the influence of outside forces on park management, and determine specific mechanisms to promote active participation of local communities; 4) conduct studies to clarify legal status of local residents; 5) document policy changes NRM Project has contributed for improving guidelines for the Bina Desa program, and decide whether NRM Project should be seriously involved in improving policies in that program; 6) expand the role of local NGOs in project implementation; and 7) include the fire control program in a broader agenda for community development, not as a fragmented activity.

## VI. INDUSTRIAL WASTE MINIMIZATION PILOT ACTIVITIES

The rapid growth of Indonesia's industrial sector has resulted in serious pollution problems, including air and water pollution. Recognizing these problems, USAID/I signed a two-year cooperative agreement with the World Environmental Center (WEC) in June 1992 to undertake industrial waste minimization pilot activities. The pilot activities had two primary objectives: a) to strengthen the capability of the private sector to minimize the discharge of pollutants while maintaining or enhancing economic efficiency through improved technology and operating procedures; and b) to promote the formulation and enforcement of

improved policies for waste minimization and pollution prevention.

The two most important lessons learned from the implementation of the pilot program were that: 1) the potential impact of minimizing pollution while enhancing industrial profitability in Indonesia is great; and 2) there is a strong demand for assistance in this area from both the public and private sectors of the economy. The team interviewed a sample of participating companies and noted that these companies obtained significant benefits from their participation in the project, including water consumption reduction, more efficient production processes, on-site recycling of raw materials, and reduction of waste. The evaluation also found that waste reduction assessments generated enthusiasm, and increased interest in industrial waste minimization in Indonesia. An industrial waste minimization seminar conducted in seven major cities revealed that a great majority of the participants believe that waste reduction assessments and clean technology information would be very useful to their companies.

While the evaluation team noted some deficiencies in the pilot program, including a lack of emphasis linking pilot activities with policy formulation, USAID/G01 have been encouraged by the positive results of the activities, and have agreed to amend the NRM Project to include an expanded industrial pollution reduction component. The emphases of the expansion will be on continuing to identify cost-effective pollution reduction strategies, in training host-country counterparts in environmental auditing procedures, and linking the project's pilot activities to remedial policy measures.

## VII. CROSS-CUTTING ISSUES

**V11.1. Women in Development.** Considerable efforts are being made by the project to monitor the number of women who participate in and receive benefits from project activities. However, these numbers are not able to reflect the quality of these benefits. Accordingly, the team recommended that the project: a) incorporate increasing women participation as a cross-cutting theme in project implementation; b) review results of policy studies to determine gender implications of recommendations have been made, c) increase the emphasis on recruiting women staff for field and policy analysis positions; and d) increase the number of women participants in training programs.

**VII.2. Collaboration with Other Donors.** While communications with other donors have been reasonably good, the team recommended that USAID should seek ways to improve such communication even more by supporting the Consultative Group on Indonesia Forestry by: inviting other donor representatives to visit NRM Project sites, have other donors review terms of reference for policy studies to avoid redundancy, and continue inviting donor representatives to NRM Project presentations.

## ATTACHMENTS

### K. ATTACHMENTS

*Attachments (List attachments submitted with this Evaluation Summary; always attach copy of full evaluation report, even if one was submitted earlier; attach studies, surveys, etc, from "on-going" evaluation, if relevant to the evaluation report.)*

## COMMENTS

### L. Comments By Mission, AID/W Office and Borrower/Grantee On Full Report

1. The major recommendations of the evaluation, which are discussed in the PES, have all been addressed and are being incorporated as part of on-going project activities.
2. There was a considerable lapse of time before the midterm evaluation was completed, and approved by the Mission. An acceptable final report was not provided until April 1995, almost nine months after

the evaluation team had left Indonesia. The principal reason for the delay was the intransigence of the team leader, who seemed to be unwilling to provide a reasonably well-structured draft of the team's findings, and was, indeed, reluctant to respond to Mission comments regarding unsubstantiated statements and factual errors that were incorporated in the original draft. In order to complete the evaluation, the policy team leader of the Environmental and Natural Resources Policy and Training Project (EPAT), and the Mission's natural resource policy advisor, together conducted additional In-country discussions, gathered supplementary information, and, rewrote the report. These additional activities required an extra level of effort of three-to-four person months.

## BIBLIOGRAPHY

- The American Presidency Project. 1970. "162 - Memorandum Calling for an Evaluation of Current Federal Programs." *The American Presidency Project*. <<http://www.presidency.ucsb.edu/ws/index.php?pid=2516>> (accessed January 15, 2009).
- Amira, Rafika. 2007. "Certification for Evaluations in the Government of Canada: The Way Ahead." (paper presented at the Danish Evaluation Society Conference in Kolding, Denmark, September 14, 2007). <[http://www.danskevalueringselskab.dk/PowerPoint/Ses3\\_RafikaAmira.ppt#278](http://www.danskevalueringselskab.dk/PowerPoint/Ses3_RafikaAmira.ppt#278)>.
- Asian Development Bank. 2009a. "Attributes of Independence." *ADB*. <<http://www.adb.org/Evaluation/about/attributes.asp>> (accessed January 12, 2009).
- Asian Development Bank. 2009b. "Syntheses of Evaluation Findings." *Evaluation*. <<http://www.adb.org/Evaluation/reports.asp?s=1&type=12>> (accessed February 2, 2009).
- Asian Development Bank. No Date. *Guidelines for Disseminating Findings and Recommendations of the Operations Evaluation Department*. Manila: ADB. <<http://www.adb.org/Documents/Guidelines/Guidelines-Disseminating-Findings-Recommendations/Guidelines-Disseminating-Findings-Recommendations.pdf>>.
- Australia. Australian Agency for International Development. 2005. *Undertaking Activity Evaluations*. Canberra: AusAID. <<http://www.ausaid.gov.au/ausguide/pdf/ausguideline5.2.pdf>>.
- Bandstein, Sara and Erik Hedblom 2008. "Management Response Systems in AID Organisations; Case Studies of SIDA, IFAD and EUROPEAID. <http://www.wu.ac.at/inst/fsnu/vienna/papers/bandstein-hedblom.pdf> (accessed September 10, 2009)
- Bayley, John S.. 2008. "Maximizing the Use of Evaluation Findings." *Asian Development Bank*. <<http://www.adb.org/Documents/OED/Occasional-Papers/Evaluation-Findings.pdf>> (accessed January 9, 2009).
- Bollen, Kenneth; Pamela Paxton and Rumi Morishima. 2005. "Assessing International Evaluations: An Example From USAID's Democracy and Governance Program." *American Journal of Evaluation* 26.2.
- Bretton Woods Project. 2009. "Independence of IMF's evaluation arm questioned: IEO sets broad work plan." *The Bretton Woods Project*. <<http://www.brettonwoodsproject.org/art-542342>> (accessed February 4, 2009).
- Canada. Canadian International Development Agency. 2006. *Peer Review of Evaluation Function at United Nations Children's Fund (UNICEF)*. Ottawa: CIDA 2006. <[www.uneval.org/documentdownload?doc\\_id=98&file\\_id=205](http://www.uneval.org/documentdownload?doc_id=98&file_id=205)>.
- Canada. Canadian International Development Agency. 2008. "What We're Learning." <<http://www.acdi-cida.gc.ca/CIDAWEB/acdicida.nsf/En/EMA-218131539-PFQ>> (accessed January 28, 2009).
- Canada. Treasury Board of Canada Secretariat. 2004. *Report on Effective Evaluation Practices*. Ottawa: Treasury. <[http://www.tbs-sct.gc.ca/eval/tools\\_outils/eep-pee/eep-pee\\_e.asp](http://www.tbs-sct.gc.ca/eval/tools_outils/eep-pee/eep-pee_e.asp)>.
- Canadian Evaluation Society. 2009. "CES Guidelines for Ethical Conduct." *Canadian Evaluation Society*. <<http://evaluationcanada.ca/site.cgi?s=5&ss=4&lang=en>> (accessed January 28, 2009).
- Caracelli, Valerie J.; and Hallie Preskill. 2000. *The Expanding Scope of Evaluation Use: New Directions for Evaluation*, No. 88. San Francisco: Jossey-Bass.

- Centre of Excellence for Evaluation. 2004. "Guide for the Review of Evaluation Reports." *Treasury Board of Canada Secretariat*. <[http://www.tbs-sct.gc.ca/eval/tools\\_outils/4001752\\_e.asp](http://www.tbs-sct.gc.ca/eval/tools_outils/4001752_e.asp)> (accessed February 1, 2009).
- The Council for Excellence in Government. 2009. "What Constitutes Strong Evidence of a Program's Effectiveness." *CEG*. <[http://excelgov.org/admin/FormManager/filesuploading/OMB\\_memo\\_on\\_strong\\_evidence.pdf](http://excelgov.org/admin/FormManager/filesuploading/OMB_memo_on_strong_evidence.pdf)> (accessed January 8, 2009).
- DAC Evaluation Network. 2006. *DAC Evaluation Quality Standards: (for test phase application)*. Paris: OECD. <<http://www.oecd.org/dataoecd/30/62/36596604.pdf>>.
- DAC Evaluation Resource Centre. 2006. *Guidance for Managing Joint Evaluations*. Paris: OECD. <<http://www.oecd.org/dataoecd/58/36/37656526.pdf>>.
- The Empowerment Evaluation Institute. 2009. "Empowerment Evaluation." *Fetterman & Associates and Stanford University*. <<http://www.stanford.edu/~davidf/institute.html>> (accessed February 3, 2009).
- Estrella, Marisol; and John Gaventa. 1998. "Who Counts Reality? Participatory Monitoring and Evaluation: A Literature Review." *IDS Working Paper 70*. Brighton: Institute of Development Studies. <[http://siteresources.worldbank.org/INTPCENG/1143331-1116505657479/20553240/IDSwp70\\_PME1.pdf](http://siteresources.worldbank.org/INTPCENG/1143331-1116505657479/20553240/IDSwp70_PME1.pdf)>.
- European Evaluation Society. 2007. "EES Statement: The importance of a methodologically adverse approach to impact evaluation - specifically with respect to the development aid and development interventions." *EES*. <<http://www.europeanevaluation.org/download/?noGzip=1&id=1969403>> (accessed January 4, 2009).
- European Union. EuropeAid. 2005. "Content of the Final Report." *EuropeAid*. <[http://ec.europa.eu/europeaid/evaluation/methodology/guidelines/gui\\_ctt\\_flr\\_en.htm](http://ec.europa.eu/europeaid/evaluation/methodology/guidelines/gui_ctt_flr_en.htm)> (accessed January 31, 2009).
- European Union. EuropeAID. 2009. "Untitled Document." *European Commission*. <[http://ec.europa.eu/europeaid/evaluation/methodology/examples/mod\\_qal\\_pro\\_en.doc](http://ec.europa.eu/europeaid/evaluation/methodology/examples/mod_qal_pro_en.doc)> (accessed January 2, 2009).
- Evaluation Cooperation Group. 2009a. "Good Practice Standards – Approved." *ECGnet*. <[https://wpqp1.adb.org/QuickPlace/ecg/Main.nsf/h\\_B084A3976FF5F808482571D90027AD16/1E8F8A367033183248257463002F0726/](https://wpqp1.adb.org/QuickPlace/ecg/Main.nsf/h_B084A3976FF5F808482571D90027AD16/1E8F8A367033183248257463002F0726/)> (accessed February 3, 2009).
- Evaluation Cooperation Group. 2009b. "Good Practice Standards for Evaluation of MDB Supported Public Sector Operations." *ECGnet*. <[https://wpqp1.adb.org/QuickPlace/ecg/Main.nsf/h\\_C5732B0904D1CCED48257463002F9480/E1A428AA6F69E53048257447002FDDBF/?OpenDocument](https://wpqp1.adb.org/QuickPlace/ecg/Main.nsf/h_C5732B0904D1CCED48257463002F9480/E1A428AA6F69E53048257447002FDDBF/?OpenDocument)>.
- Evaluation Cooperation Group. 2009c. "Letter on the Importance of Independent Evaluation." *ECGnet*. <[https://wpqp1.adb.org/QuickPlace/ecg/Main.nsf/h\\_01C91A617A00448948257463002FE916/7CE08AA2FA6D684B482573CB0001BBAE/?OpenDocument](https://wpqp1.adb.org/QuickPlace/ecg/Main.nsf/h_01C91A617A00448948257463002FE916/7CE08AA2FA6D684B482573CB0001BBAE/?OpenDocument)> (accessed January 7, 2009).
- Evaluation Cooperation Group. 2009d. "Template for Assessing the Independence of Evaluation Organizations." *ECGnet*. <[https://wpqp1.adb.org/QuickPlace/ecg/Main.nsf/h\\_01C91A617A00448948257463002FE916/1FC1011483FC85644825729300357D73/?OpenDocument](https://wpqp1.adb.org/QuickPlace/ecg/Main.nsf/h_01C91A617A00448948257463002FE916/1FC1011483FC85644825729300357D73/?OpenDocument)> (accessed January 7, 2009).

- Feinstein, Osvaldo; and Gregory K. Ingram. 2003. "Lessons Learned from World Bank Experiences in Joint Evaluation," (paper presented at the Development Assistance Working Party on Aid Evaluation workshop "Partners in Development Evaluation: Learning and Accountability", Paris, France, March 25-26, 2003). <<http://www.oecd.org/dataoecd/secure/15/13/31736431.pdf>>.
- Foresti, Marta; et al. 2007. *A Comparative Study of Evaluation Policies and Practices in Development Agencies*. Paris: French Agency for Development. <<http://www.afd.fr/jahia/webdav/site/afd/users/admirecherche/public/ODI.pdf>>.
- French Agency for Development. 2009a. "Evaluation and Capitalisation Unit." *AFD*. <<http://www.afd.fr/jahia/Jahia/site/afd/lang/en/Evaluation-Capitalisation>> (accessed February 16, 2009).
- French Agency for Development. 2009b. "Grille d'appréciation des rapports d'évaluations décentralisées," *Agence Française de Développement*. <<http://www.afd.fr/jahia/Jahia/site/afd/lang/fr/pid/12604>> (accessed February 1, 2009).
- Guba, Egon G.; and Yvonna S. Lincoln. 2001. "Guidelines and Checklist for Constructivist (a.k.a. Fourth Generation) Evaluation." The Evaluation Center, Western Michigan University. <<http://www.wmich.edu/evalctr/checklists/constructivisteval.pdf>> (accessed February 4, 2009).
- Hogan, R. Lance. 2007. "The Historical Development of Program Evaluation: Exploring the Past and Present," *Online Journal of Workforce Education and Development* 2.4. <[http://wed.siu.edu/Journal/VolIIInum4/Article\\_4.pdf](http://wed.siu.edu/Journal/VolIIInum4/Article_4.pdf)> (accessed February 5, 2009).
- Huse, Irene; and James C. McDavid. 2006. *Literature Review: Professionalization of Evaluators*. Victoria, BC: University of Victoria. <[http://www.evaluationcanada.ca/txt/2\\_literature\\_summary\\_e.pdf](http://www.evaluationcanada.ca/txt/2_literature_summary_e.pdf)>.
- Independent Evaluation Group. 2008. *Annual Review of Development Effectiveness 2008*. Washington DC: World Bank. <<http://siteresources.worldbank.org/EXT2008ANNREVDEVEFFE/Resources/appC.pdf>>.
- Independent Evaluation Group. 2009a. "CAE Methodology." *World Bank*. <<http://web.worldbank.org/WBSITE/EXTERNAL/EXTOED/EXTCOUASSEVA/0,contentMDK:21107046~menuPK:3098030~pagePK:64168445~piPK:64168309~theSitePK:2832047,00.html>> (accessed February 1, 2009).
- Independent Evaluation Group. 2009b. "Tracking Management Actions on IEG's Recommendations," World Bank. <<http://web.worldbank.org/WBSITE/EXTERNAL/EXTOED/EXTANNREPOPEEVAL/0,contentMDK:21115889~currentSitePK:1324361~pagePK:64829573~piPK:64829550~theSitePK:4425661,00.html>> (accessed January 15, 2009).
- Independent Evaluation Office. 2009. "Our Guarantees of Independence." *International Monetary Fund*. <<http://www.imf.org/external/np/ieo/ind.htm>> (accessed January 4, 2009).
- Inter-American Development Bank. 2009. "Impact Evaluations." *Office of Evaluation and Oversight*. <<http://www.iadb.org/ove/DefaultNoCache.aspx?Action=WUCPublications@ImpactEvaluations>> (accessed January 30, 2009).
- International Development Evaluation Association. 2009. "Internationalization of Evaluation.," *IDEAS*, <<http://unpan1.un.org/intradoc/groups/public/documents/APCITY/UNPAN017424.pdf>> (accessed January 15, 2009).

- Japan. 2001. Japan International Cooperation Agency, *Participatory Evaluation and International Cooperation*. Tokyo: JICA.  
<<http://www.jica.go.jp/english/publications/reports/study/topical/participatory/pdf/par.pdf>>.
- Kirkhart, Karen E.. 2000. "Reconceptualizing Evaluation Use: An Integrated Theory of Influence." *New Directions for Evaluation* 88.  
<<http://www3.interscience.wiley.com/journal/109752045/abstract?CRETRY=1&SRETRY=0>>.
- Liverani, Andrea; and Hans E. Lundgren. 2007. "Evaluation Systems in Development Aid Agencies." *Evaluation* 13.2. <<http://evi.sagepub.com/cgi/content/abstract/13/2/241>>.
- Manning, Richard. 2008. *Development Co-operation Report 2007*. Paris: OECD Publishing.  
<<http://fiordiliji.sourceoecd.org/pdf/dac/432008011e-02.pdf>>.
- Mathison, Sandra. 2005. *Encyclopedia of Evaluation*. Thousand Oaks: Sage.
- McKay, Keith. 2007. "How to Build M&E Systems to Support Better Government." *World Bank Independent Evaluation Group*. <[http://www.worldbank.org/ieg/eecd/docs/How\\_to\\_build\\_ME\\_6.pdf](http://www.worldbank.org/ieg/eecd/docs/How_to_build_ME_6.pdf)>.
- MicroLinks. 2009. "Impact Assessments." *MicroLinks*.  
<[http://www.microlinks.org/ev\\_en.php?ID=13199\\_201&ID2=DO\\_TOPIC](http://www.microlinks.org/ev_en.php?ID=13199_201&ID2=DO_TOPIC)> (accessed January 29, 2009).
- Millennium Challenge Corporation. 2009. "Impact Evaluations at MCC." *Millennium Challenge Corporation*.  
<<http://www.mcc.gov/programs/impacetevaluation/index.php>> (accessed January 31, 2009).
- National Research Council. 2008. *Improving Democracy Assistance: Building Knowledge Through Evaluations and Research*. Washington: National Academies Press.
- New Zealand. NZAID. 2008. *Annual Report on Quality*. Wellington: NZAID.  
<<http://www.nzaid.govt.nz/what-we-do/docs/eva0700rep-report.pdf>>.
- OECD Development Co-operation Directorate. 2008. "Accra Agenda for Action" (paper presented at the 3<sup>rd</sup> High Level Forum on Aid Effectiveness, Accra, Ghana, September 2-4, 2008).  
<<http://siteresources.worldbank.org/ACCRAEXT/Resources/4700790-1217425866038/AAA-4-SEPTEMBER-FINAL-16h00.pdf>>.
- Organization for Economic Co-operation and Development. 2001. *Principles for Evaluation of Development Assistance*. Paris: OECD. <<http://www.oecd.org/dataoecd/21/32/41029845.pdf>>.
- Organization for Economic Co-operation and Development. 2004. *Combined Members Views on Managing for Development Results*. Paris: OEDC.  
<<http://www.mfdr.org/documents/OECD%20DAC%20-%20Combined%20Views.doc>>.
- Office of Evaluation and Oversight. 2009. "Evaluation Capacity Development." *LADB*.  
<<http://www.iadb.org/ove/Default.aspx?Action=WUCHtmlContent@evaluationCapacityDevelopment>> (accessed January 21, 2009).
- Office of Evaluation and Strategic Planning. 2009. "Participation Within UNDP." UNDP.  
<<http://www.undp.org/eo/documents/whop3.htm>> (accessed February 16, 2009).
- Patton, Michael Q.. 2008. *Utilization-Focused Evaluation*. Thousand Oaks: Sage.
- Picciotto, Robert. 2003. "International Trends and Development Evaluation: The Need for New Ideas." *American Journal of Evaluation* 24.2: 227-234.

- Picciotto, Robert . 2009. "Evaluation Independence at DFID." *Independent Advisory Committee for Development Impact*. <<http://iacdi.independent.gov.uk/wp-content/uploads/iacdi-evaluation-independence-at-dfid-final.pdf>> (accessed January 7, 2009).
- <[http://eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?\\_nfpb=true&\\_ERICExtSearch\\_SearchValue\\_0=EJ680065&ERICExtSearch\\_SearchType\\_0=no&accno=EJ680065](http://eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=EJ680065&ERICExtSearch_SearchType_0=no&accno=EJ680065)>.
- Robinson, Mark. *No Date*. "Performance Information for Performance Budgeting." *International Monetary Fund*. <<http://siteresources.worldbank.org/INTMEXICO/Resources/2-2MarcRobinson.pdf>>.
- Russo, Luca; et al. 2005. *Joint Evaluation of Effectiveness and Impact of the Enabling Development Policy of the World Food Programme (WFP): Vol. 1 Synthesis Report*. Bonn: Federal Ministry for Economic Cooperation and Development. <<http://documents.wfp.org/stellent/groups/public/documents/reports/wfp066300.pdf>>.
- Schaumburg-Müller, Henrik. 1996. *Evaluating Capacity Building, Donor Support and Experiences*. Copenhagen: Danish International Development Agency. <<http://www.oecd.org/dataoecd/20/52/16546669.pdf>>.
- Segone, Marco. 2006., "New Trends in Development Evaluation." Geneva: UNICEF and IPEN. <[http://www.unicef.org/ceecis/New\\_trends\\_Dev\\_Evaluation.pdf](http://www.unicef.org/ceecis/New_trends_Dev_Evaluation.pdf)>.
- Segone, Marco. 2009. "Creating and Developing Evaluation Organizations: Lessons Learned from Africa, Asia, America, Australasia and Europe." UNICEF. <[http://www.unicef.org/ceecis/0702-Segone\\_IOCE\\_book.pdf](http://www.unicef.org/ceecis/0702-Segone_IOCE_book.pdf)> (accessed February 17, 2009).
- Shadish, William R., Jr.; Thomas D. Cook and Laura C. Leviton. 1991. *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, CA: Sage Publications.
- Simonson, Michael . 2005. "Quantitative Research Returns: Why did it Leave?" *Quarterly Review of Distance Education* 6.3. <[http://www.infoagepub.com/products/journals/qrde/articles/6-3\\_txt.pdf](http://www.infoagepub.com/products/journals/qrde/articles/6-3_txt.pdf)>.
- Stiglitz, Joseph E. 2002. "Participation and Development: Perspectives from the Comprehensive Development Paradigm." *Review of Development Economics* 6 (2). <<http://siteresources.worldbank.org/INTPCENG/1143240-1116506251485/20553233/stiglitzpart.pdf>>.
- Swedish Agency for Development Evaluation. 2009. "About the Agency." *SADEV*. <<http://www.sadev.se/Bazment/12.aspx>> (accessed January 12, 2009).
- Teller Charles H. 2008. "Are We Learning About What Really Works? Lost Opportunities and Constraints to Producing Rigorous Evaluation Designs of Health Project Impact." *Evaluation Cooperation Group*. <[http://www.worldbank.org/ieg/nonie/docs/WP9\\_Teller.pdf](http://www.worldbank.org/ieg/nonie/docs/WP9_Teller.pdf)>.
- United Kingdom. Department for International Development. 2005. *Guidance on Evaluation and Review for DFID staff*. London: DFID. <<http://www.dfid.gov.uk/aboutdfid/performance/files/guidance-evaluation.pdf>>.
- United Kingdom. Department for International Development. 2006. *DFID Country Program Evaluations; Synthesis of 2005/2006 Evaluations*. London: DFID. <<http://www.dfid.gov.uk/aboutdfid/performance/files/ev671.pdf>>.
- United Kingdom. Department for International Development. 2008. *Report on DFID's Responses to Evaluation Reports in 2006/7*. London: DFID. <<http://iacdi.independent.gov.uk/wp-content/uploads/report-on-follow-up-to-evaluation-reports-in-2006-07.doc>>.
- United Kingdom. Department for International Development. 2009a. "Building the evidence to reduce poverty: The UK's policy on independent evaluation for international development." *DFID*. <<http://www.dfid.gov.uk/consultations/draft-evaluation-policy2.pdf>> (accessed February 28, 2009).



- United Kingdom. Department for International Development. 2009b. Draft For Consultation - Building the evidence to reduce poverty: The UK's policy on independent evaluation for international development. London: DFID. <<http://mande.co.uk/blog/wp-content/uploads/2008/12/dfid-evaluation-policy-consultation-draft-2.pdf>>.
- United Kingdom. The Independent Advisory Committee on Development Impact. 2009. "Meeting Minutes." London: DFID. <<http://iacdi.independent.gov.uk/wp-content/uploads/minutes-iacdi-jan-09.doc>>.
- United Nations Children's Fund. 2004. *UNICEF Evaluation Report Standards*. New York: UNICEF, 2004. <[http://www.unicef.org/evaluation/files/UNICEF\\_Evaluation\\_Report\\_Standards.pdf](http://www.unicef.org/evaluation/files/UNICEF_Evaluation_Report_Standards.pdf)>.
- United Nations Development Programme. 2007. "Regional Workshop Report on the UNDP Evaluation Policy: Latin America and the Caribbean (January 17- 19 2007; Antigua, Guatemala)." *UNDP*. <[http://www.undp.org/eo/documents/workshop/regional\\_ws/wsr\\_LAC.pdf](http://www.undp.org/eo/documents/workshop/regional_ws/wsr_LAC.pdf)> (accessed December 28, 2008).
- United Nations Development Program. 2008. "Evaluation Documents." *Evaluation Resource Center*. <[http://cfapp1-docs-public.undp.org/eo/evaldocs1/eo\\_doc\\_975063034](http://cfapp1-docs-public.undp.org/eo/evaldocs1/eo_doc_975063034)> (accessed February 6, 2009).
- United Nations Development Program. 2009a. "Essentials." *Evaluation*. <<http://www.undp.org/eo/essentials.htm>> (accessed February 3, 2009).
- United Nations Development Program. 2009b. *Evaluation Report*. New York: UNDP. <[http://www.undp.org/eo/documents/erc/Evaluation\\_Report.doc](http://www.undp.org/eo/documents/erc/Evaluation_Report.doc)>.
- United Nations Development Program. 2009c. "UNDP Management Response Template." *UNDP*. <[http://www.undp.org/eo/documents/erc/UNDPManagement\\_Response\\_template.doc](http://www.undp.org/eo/documents/erc/UNDPManagement_Response_template.doc)> (accessed January 5, 2009).
- United Nations Evaluation Group. 2009a. "Core Competencies for Evaluators of the UN System." *UNEG*. <[http://www.uneval.org/papersandpubs/documentdetail.jsp?doc\\_id=81](http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=81)> (accessed on January 10, 2009).
- United Nations Evaluation Group. 2009b. "UN Reform and Evaluation." *UNEG*. <[http://www.unevaluation.org/currentwork/currentworklist.jsp?currentworkid=102&doc\\_cat\\_source\\_id=2&doc\\_source\\_id=102](http://www.unevaluation.org/currentwork/currentworklist.jsp?currentworkid=102&doc_cat_source_id=2&doc_source_id=102)> (accessed on January 10, 2009).
- United Nations Population Fund. 2007a. *UNFPA Compliance with Internationally Recognized Evaluation Quality Standards*. New York: UN. <[http://www.unfpa.org/exbrd/2007/annualsession/evaluation\\_progress\\_report.doc](http://www.unfpa.org/exbrd/2007/annualsession/evaluation_progress_report.doc)>.
- United Nations Population Fund. 2007b. "UNFPA Executive Board Documents." *UNFPA*. <[http://www.unfpa.org/exbrd/2007/2007\\_annual.htm](http://www.unfpa.org/exbrd/2007/2007_annual.htm)> (accessed February 5, 2009). (under Section 17)
- United States. Department of Energy. 2009. *EERE Quality Assurance Guidelines for General Program Evaluation Studies*. Washington DC: DOE. <[http://www1.eere.energy.gov/ba/pba/pdfs/eval\\_guide\\_appendix\\_7.pdf](http://www1.eere.energy.gov/ba/pba/pdfs/eval_guide_appendix_7.pdf)>.
- United States. Government Accountability Office. 1999. *Performance Budgeting: Initial Agency Experiences Provide a Foundation to Assess Future Directions*. Washington: GAO. <<http://archive.gao.gov/f0902b/162387.pdf>>.
- United States. Government Accountability Office. 2003. *Program Evaluation: An Evaluation Culture and Collaborative Partnerships Help Build Agency Capacity*. Washington DC: GAO. <<http://www.gao.gov/new.items/d03454.pdf>>.

- United States. Government Accountability Office. 2005. *Performance Budgeting: PART Focuses Attention on Program Performance, but More Can Be Done to Engage Congress*. Washington: GAO. <<http://www.gao.gov/new.items/d0628.pdf>>.
- United States. U.S. Agency for International Development. 1970. *Evaluation Handbook*. Washington: USAID. <[http://pdf.usaid.gov/pdf\\_docs/PDABR615.pdf](http://pdf.usaid.gov/pdf_docs/PDABR615.pdf)>
- United States. U.S. Agency for International Development. 1989. *Review of the Quality of AID Evaluations FY 1987 and FY 1988*. Washington DC: USAID. <[http://pdf.usaid.gov/pdf\\_docs/PNABC321.pdf](http://pdf.usaid.gov/pdf_docs/PNABC321.pdf)>.
- United States. U.S. Agency for International Development. 1995. *AID Evaluation Summary*. Washington: USAID. <[http://pdf.usaid.gov/pdf\\_docs/PDABM718.pdf](http://pdf.usaid.gov/pdf_docs/PDABM718.pdf)>.
- United States. Agency for International Development. 1996. *Investments in Agriculture: A Synthesis of the Evaluation Literature*. Washington DC: USAID. <[http://www.usaid.gov/pubs/usaaid\\_eval/pdf\\_docs/pnaby205.pdf](http://www.usaid.gov/pubs/usaaid_eval/pdf_docs/pnaby205.pdf)>.
- United States. U.S. Agency for International Development. 1999. *Performance Monitoring and Evaluation*. Washington: USAID. <[http://pdf.usaid.gov/pdf\\_docs/pnaby239.pdf](http://pdf.usaid.gov/pdf_docs/pnaby239.pdf)>.
- United States. U.S. Agency for International Development. 2006. *Constructing an Evaluation Report* (Washington DC: USAID. <[http://pdf.usaid.gov/pdf\\_docs/PNADI500.pdf](http://pdf.usaid.gov/pdf_docs/PNADI500.pdf)>.
- United States. U.S. Agency for International Development. 2008a. "ADS Chapter 203; Assessing and Learning," *Programming Policy*. <<http://www.usaid.gov/policy/ads/200>> (accessed February 6, 2009).
- United States. U.S. Agency for International Development. 2008b. "Home-Based Care Significantly Reduces Neonatal Mortality in Developing Countries." *USAID*. <<http://www.usaid.gov/press/releases/2008/pr080618.html>> (accessed January 28, 2009).
- Weiss, Carol H.. 1979. "The Many Meanings of Research Utilization." *Public Administration Review* 39.5: 426-431, <[http://www.g-  
rap.org/docs/Mixed/Weiss%201979%20The%20Many%20Meanings%20of%20Research%20Utilisations.  
pdf](http://www.g-<br/>rap.org/docs/Mixed/Weiss%201979%20The%20Many%20Meanings%20of%20Research%20Utilisations.<br/>pdf)>.
- World Bank. 2004. *Influential Evaluations: Evaluations that Improved Performance and Impacts of Development Programs*. Washington: World Bank. <[http://lnweb90.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/67433EC6C181C22385256E7F0073BA1C/\\$file/influential\\_evaluations\\_eed.pdf](http://lnweb90.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/67433EC6C181C22385256E7F0073BA1C/$file/influential_evaluations_eed.pdf)>.
- World Bank. 2007. *A Decade of Action in Transport: An Evaluation of World Bank Assistance to the Transport Sector, 1995-2005*. Washington DC: World Bank. <[http://siteresources.worldbank.org/EXTTRANS/Resources/transport\\_eval.pdf](http://siteresources.worldbank.org/EXTTRANS/Resources/transport_eval.pdf)>.
- World Bank. 2008a. *Annual Report 2008*. Washington, D.C.: World Bank. <[http://siteresources.worldbank.org/EXTANNREP2K8/Resources/YR00\\_Year\\_in\\_Review\\_English.pdf](http://siteresources.worldbank.org/EXTANNREP2K8/Resources/YR00_Year_in_Review_English.pdf)>.
- World Bank. 2008b. "Mexico: Agenda, Papers, and Bios." *World Bank Countries*. <<http://web.worldbank.org/WBSITE/EXTERNAL/COUNTRIES/LACEXT/MEXICOEXTN/0,,contentMDK:21789925~pagePK:141137~piPK:141127~theSitePK:338397,00.html>> (accessed January 13, 2009).

- World Bank. 2008c. *Performance-Informed Budgeting in Latin America: Experiences and Opportunities*. Washington: World Bank. <<http://siteresources.worldbank.org/MEXICOEXTN/Resources/0-3NickManningFinal.pdf>>.
- World Bank. 2009. *Influential Evaluations.: Evaluations that Improved Performance and Impacts of Development Programs*.
- The World Bank Group. 2009a. "Building Government Monitoring and Evaluation (M&E) Systems." *Independent Evaluation Group*. <<http://www.worldbank.org/ieg/eed/index.html>> (accessed January 20, 2009).
- The World Bank Group. 2009b. "Evaluations by Country." *World Bank*. <<http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTTSPMA/0,,contentMDK:20209970~menuPK:435976~pagePK:148956~piPK:216618~theSitePK:384329,00.html>> (accessed January 12, 2009).
- Yin, Robert K.; and Carol H. Weiss. 1988. *Preliminary Study of the Utilization of AID's Evaluation Reports*. Washington DC: USAID. <[http://dec.usaid.gov/index.cfm?p=search.getCitation&CFID=9801169&CFTOKEN=36491238&id=sDD19F9D6-D566-FC5C-D98314451C9F0EBB&rec\\_no=52769](http://dec.usaid.gov/index.cfm?p=search.getCitation&CFID=9801169&CFTOKEN=36491238&id=sDD19F9D6-D566-FC5C-D98314451C9F0EBB&rec_no=52769)>.