

61547

**A Review of the Quality and
Coverage of A.I.D. Evaluations,
FY 1989 and FY 1990**

May 1992

Submitted to:

Nena Vreeland
CDIE/SDS
Agency for International Development



600 Water Street S.W., NBU 7-7
Washington, D.C. 20024

telephone: (202) 484-7170
telex: 4990821MANSY fax: (202) 488-0754

1558-015
(5/92)

TABLE OF CONTENTS

EXECUTIVE SUMMARY	i
CHAPTER ONE	
A. Background	1
B. Coverage and Methods	2
C. Evaluation Review Team	5
D. Structure of the Report	6
CHAPTER TWO	
A. Number of Evaluations Examined	7
B. Distribution of the Evaluations By A.I.D. Bureaus	7
C. Project Versus Non-Project Assistance	15
D. The Duration of Projects and Non-Project Assistance Efforts	15
E. Sectoral Affiliation of Projects Examined by the Evaluations	16
F. Special Interest Codes	17
G. Sources of Funds	20
H. Funding Levels of the Activities Evaluated	21
I. Share of the A.I.D. Portfolio Examined by FY 1989 and FY 1990 Evaluations	22
CHAPTER THREE	
A. Evaluation Scopes of Work	25
B. Descriptions of Evaluation Methods in Evaluation Reports	27
C. A.I.D. Evaluation Summaries	29
CHAPTER FOUR	
A. The Sponsorship of A.I.D. Evaluations	31
B. Evaluation Types and Timing	33
C. The Scope of A.I.D. Evaluations	35
D. The Purposes of A.I.D. Evaluations	35
CHAPTER FIVE	
A. The Size of A.I.D. Evaluation Teams	43
B. Team Structure and Composition	43
C. The Skills Provided By Evaluation Team Members	48
D. The Blending of Skills on A.I.D. Evaluation Teams	50
E. Gender Considerations and Team Composition	52
CHAPTER SIX	
A. Participation in the Development of Scopes of Work and Audience Identification	53
B. A.I.D. and Host Ministry Participation on Evaluation Teams	55

C.	Users Advisory Panels	58
D.	Team Planning Meetings	59
E.	A.I.D. and Host Country Participation in Oral Debriefings and the Review of Written Drafts of Evaluation Reports	60
F.	Formal Evaluation Reviews	61
G.	A Composite View of A.I.D. and Host Country Participation	62
 CHAPTER SEVEN		
A.	The Basis Evaluations Use to Judge Performance	64
B.	Evaluation Designs and Methods	67
 CHAPTER EIGHT		
A.	Findings and Conclusions About Basic Project Performance	78
B.	Findings and Conclusions about Cross-Cutting Issues	84
 CHAPTER NINE		
		92

LIST OF TABLES

Table 1-1	Evaluations and Other Types of Documents Screened for Inclusion in the Study Sample	3
Table 2-1.	Evaluations in the FY89 and FY90 That Examined a Relatively Large Number of A.I.D. Projects Simultaneously	8
Table 2-2.	Distribution of Evaluations of One or a Few Related Activities by Bureau	9
Table 2-3.	Percentage of Evaluations by Region Included in Successive Evaluation Reviews	10
Table 2-4.	Distribution of Asia Bureau Evaluations	11
Table 2-5.	Distribution of Near East Bureau Evaluations	11
Table 2-6.	Distribution of Latin America and Caribbean Bureau Evaluations	12
Table 2-7.	Distribution of Africa Bureau Evaluation	13
Table 2-8.	Distribution of R & D Bureau Evaluations	14
Table 2-9.	Degree to which Evaluations for Regional Bureaus Examined "Mission- owned" Projects	15
Table 2-10.	Length of Projects Examined by the Evaluations	16
Table 2-11.	Frequency with Which Aggregate Activity Codes Were Found in 141 Evaluations	17
Table 2-12.	Activity Codes with Which Projects Examined by 141 Evaluations Were Associated	18
Table 2-13.	Relationship of Evaluation Coverage to Functional Accounts	19
Table 2-14.	Frequency with Which Special Interests Appear in 141 Evaluations	20
Table 2-15.	Value of Projects and Non-Project Assistance Activities Covered by the Evaluations	21
Table 2-16.	Value of Projects and Non-Project Activities Covered by Evaluations	22

C

Table 2-17.	Active Projects Covered by FY89-FY90 Evaluations as a Percent of All Projects Included in Bureau Portfolios	23
Table 2-18.	Value of Projects Covered by FY89-90 Evaluations as a Percent of Bureau Portfolios	24
Table 3-1.	Frequency with which Evaluations Had a Full or Partial Scope of Work . .	26
Table 3-2.	Share of Evaluations that Include Scopes of Work Across Several Evaluation Reviews	27
Table 3-3.	Degree to Which Evaluations Described the Data Collection and Analysis Methods That Were Used	28
Table 3-4.	Frequency with which A.I.D. Evaluations Discuss Evaluation Methods . . .	28
Table 3-5.	Frequency with which Evaluation Reports were Accompanied by Evaluation Summaries	30
Table 3-6.	Share of the Evaluations that are Accompanied by A.I.D. Evaluation Summaries	30
Table 4-1.	Organizational Sponsors of Evaluations Examined Through This Study . . .	32
Table 4-2.	Relationship Between Project Financing and Evaluation Sponsorship	32
Table 4-3.	Distribution of FY 89 and FY 90 Evaluations by Type	34
Table 4-4.	Percentage of Evaluations by Evaluation Type Across Several Evaluation Reviews	34
Table 4-5.	Types of Evaluation by Bureau	36
Table 4-6.	Scope of Evaluations Included in the Review	37
Table 4-7.	Scope Of Evaluations by Evaluation Type	38
Table 4-8.	Purpose for which Evaluations Were Carried Out	39
Table 4-9.	Frequency with which Illustrative Purposes Other than Management/Implementation Reasons Were Cited as at Least One Reason for Conducting Evaluations	39
Table 4-10.	Evaluation Implications Beyond the Project Level	41
Table 4-11.	Distribution of Evaluations with Implications Beyond the Project Level . . .	41
Table 5-1.	Distribution of Evaluation Teams by Size	43
Table 5-2.	Nature of U.S. Participation on Evaluation Teams	44
Table 5-3.	A.I.D. Participation by Bureau	45
Table 5-4.	Host Country Participation on Evaluation Teams	45
Table 5-5.	Frequency with Which Bureau Evaluations Include Host Country Personnel as Team Members	46
Table 5-6.	Host Country Participation as Evaluation Team Members	47
Table 5-7.	Distribution of Skills Across 181 Evaluation Teams Where Team Skills Were Identified	49
Table 5-8.	Complexity of Evaluation Teams	50
Table 6-1.	Frequency with Which Evaluations Identified Specific Audiences	54
Table 6-2.	Host Country Participation in Interim Evaluations of "Mission-Owned" Projects	57
Table 6-3.	Evaluations for Which a Team Planning Meeting (TPM) was Part of the Evaluation Process	60



Table 6-4.	Evaluation Participants Involvement in Reviewing Evaluation Findings . . .	61
Table 6-5.	Frequency with Which Evaluations Indicate that an Evaluation Review Meeting was Held by A.I.D.	62
Table 7-1.	Frequency with Which A.I.D. Evaluation Use Performance Indicators from Project Logical Frameworks as the Basis for Judging Success and Failure	66
Table 7-2.	Distribution of Evaluations According to Their Basic Approach or Design	69
Table 7-3.	Frequency with Which Various Sources of Evidence Were Used by Evaluations	71
Table 7-4.	Frequency with Which Evaluations Gathered Data Directly from Project or Program Beneficiaries	72
Table 7-5.	Interviewed Approaches Used by Evaluation Teams	74
Table 8-1.	Degree to Which Evaluations Report That Projects are Achieving Their Outputs and Realizing their Purposes	80
Table 8-2.	Frequency with Which Evaluation Identified Unanticipated Effects of Project Activities	81
Table 8-3.	Frequency with Which Evaluations Judged that Overall Projects Were Achieving Some or All of the Objectives	83
Table 8-4.	The Effect of Project Status/Evaluation Timing on Evaluation Conclusions Concerning Achievements	84
Table 8-5.	Frequency With Which Bureau Evaluations Explicitly Examined Sustainability Issues	85
Table 8-6.	Frequency with Which Evaluations Conclude That Some or All Project Benefits Would be Sustained	86
Table 8-7.	Frequency with Which Evaluations Reported That Some or All Projects Are Sustainable	87
Table 8-8.	Frequency With Which Evaluations Identify Other Sources of Project Funding	87
Table 8-9.	Frequency with Which Bureau Projects Reportedly Utilize User Fees During the Life of a Project	88
Table 8-10.	Degree to Which Evaluations Addressed Environmental Concerns	89
Table 8-11.	Share of Project Benefits Which Evaluations Report Are Received by Women	91

LIST OF FIGURES

Figure 5-1.	The Relationship Between Skill Complexity and Adherence to A.I.D.'S Evaluation Skills Requirement	51
-------------	---	----

ANNEXES

ANNEX A	A-1
ANNEX B	B-1

E

EXECUTIVE SUMMARY

This Management Systems International (MSI) study reviewed 286 evaluations completed by the Agency for International Development (A.I.D.) and submitted to its central evaluation office in the Center for Development Information and Evaluation (CDIE) during FY89 and FY90. It examined the basic characteristics of these evaluations, their compliance with A.I.D. evaluation requirements, the composition of evaluation teams, the methods they used to collect data and their conclusions about project performance.

The findings of this review are summarized below:

- Evaluations carried out in FY89 and FY90, on an annualized basis, examined 8% percent of the 1910 A.I.D. projects and programs which A.I.D.'s Budget Office reports were active during FY90. These evaluations covered activities which account, on an annualized basis, for approximately \$3 billion (8%) of the \$38 billion dollar life-of-project value of the projects and programs which were active as of September 1990, the last month of A.I.D.'s fiscal year.
- Of the 268 FY89 and FY90 evaluations included in the database, roughly a fourth dealt with projects in Asia and the Near East. Another fourth examined with projects in Africa and a fourth dealt with projects in Latin America and the Caribbean. The final quarter of the evaluations examined projects funded by A.I.D.'s central bureaus.
 - The 51 evaluations covering Asia Bureau projects represented 18% of that Bureau's active project portfolio.
 - The Near East Bureau with 17 evaluations, the Latin American and Caribbean Bureau (LAC) with 68 evaluations; and the Africa Bureau, with 67 evaluations, each covered 12% of their portfolios.
 - With respect to Central bureaus, the Research and Development (R&D) Bureau covered 19% of its active projects, with 39 evaluations; the Food and Humanitarian Assistance (FHA) Bureau examined 7% of its active projects, through 5 evaluations; and the Private Enterprise Bureau, with 5 evaluations, examined 21% of its active project portfolio.
- Ten (4%) of the evaluations examined non-project assistance efforts carried out by A.I.D. Within the report, these ten evaluations were treated in the same manner as project evaluations.
- Of A.I.D.'s evaluations, 224 (83%) examined individual projects. Another 38 (14%) examined multiple projects, while the remainder examined other aspects of A.I.D.'s work. On average, the projects examined by evaluations were carried out over a six-year period.

- Consistent with the findings of past evaluation reviews, the majority of the FY89 and FY90 evaluations were interim evaluations. Of the 268 evaluations in the database, 159 (59%) were interim evaluations and 68 (25%) were final evaluations. The remainder of the evaluations included a few ex-post evaluations and several reviews of "lessons learned" from a large number of evaluations.
- The scope of interim evaluations, as compared to final and ex-post evaluations, was broader than expected. 21% of all interim evaluations confined their scope to an examination of management and implementation issues. The majority (70%) examined a fuller range of questions as did final and ex-post evaluations. The evaluation review also found that of the evaluations that reached sectoral or multi-sectoral conclusions, a substantial number were interim evaluations.
- Contract personnel were involved in more evaluations (61% of the evaluations in the database) than were host country personnel (33%) or A.I.D. staff (26%).
 - On interim evaluation teams, where A.I.D.'s guidance encourages host ministry and A.I.D. staff participation, host ministry personnel were on 14% of all interim evaluation teams and A.I.D. staff were members of 30% of these teams.
- 33% of A.I.D.'s evaluations failed to provide any information regarding the skills or expertise of evaluation team members.
- The number of evaluations that contained scopes of work and methodology sections was higher in FY89 and FY90 than it had been in earlier years: 74% of the evaluations included scopes of work and 89% presented at least a partial description of the evaluations methods. On the other hand, the frequency with which evaluations were accompanied by A.I.D. Evaluation Summaries declined, with 49% of A.I.D.'s evaluations being accompanied by this required summary, as compared to 68% in FY85-86.
- The majority (87%) of A.I.D.'s evaluations used single-point-in-time, or "snapshot," evaluation design. Among these evaluations there is a great deal of variety with respect to the types of information collected and the methods used to acquire it. Both high- and low-quality methods and approaches were found in these single-point-in-time evaluations.
- Virtually all (94%) of A.I.D.'s evaluations used indicators drawn from a project's context as a basis for judging performance. Half of the evaluations (52%) explicitly stated that these performance indicators were drawn from project Logical Frameworks developed at the time projects were designed. The remaining evaluations did not identify the source of the indicators they used.

G

- A large percentage of A.I.D. evaluations appear to rely more heavily upon the "expert judgement" of evaluation team members than on rigorous evaluation designs and structured data collection techniques.
- With respect to project performance:
 - Roughly 90% of the evaluations were judged to be at least somewhat successful in an overall sense.
 - 80% of all interim and final evaluations reported that at least some project outputs were being achieved and roughly 60% of both interim and final evaluations reported that projects were achieving their purposes to some degree;
 - On the broad question of whether project benefits would be sustained, 43% of the evaluations examined this issue. This percentage compares favorably to the FY87-88 review where 36% addressed this issue.
- 34% of the evaluations that examined bi-lateral mission projects addressed the question of whether there was a good fit between a project and the missions's overall country development strategy statement (CDSS).
- Environmental impact and the gender-specific results of evaluations are rarely examined.

Based on the results of this review, MSI concluded that the A.I.D. evaluation system is serving its intended purpose of providing management-useful information to large numbers of mid-level staff who design and administer A.I.D. projects. MSI also identified improvements that could be made, including a number of steps that A.I.D. can take to enhance the coverage and quality of its evaluation work.

CHAPTER ONE

OBJECTIVES AND SCOPE

A. Background

The Agency for International Development's (A.I.D.) evaluation system, which has been in place for over twenty years, was designed to function on a decentralized basis, providing project managers as well as program planners and policy makers with management-useful information concerning on-going as well as completed projects. As a complement to this largely decentralized system, A.I.D.'s central evaluation office in A.I.D.'s Center for Development Information and Evaluation (CDIE) has, for over a decade, been responsible for developing evaluation guidance, culling lessons from A.I.D.'s evaluations, and defining the need for, as well as demonstrating approaches to, innovative evaluations that meet the needs of A.I.D.'s senior management team.¹

Since 1982, CDIE has carried out bi-annual reviews of reports of evaluations carried out by A.I.D.'s decentralized management units (e.g., its overseas missions), which these units are required to provide to CDIE. The purpose of these reviews has been to provide CDIE with an understanding of the scope of A.I.D.'s decentralized evaluation work as well as insights into the coverage, quality and findings of those evaluations.

The present report examines evaluations completed during Fiscal Years 1989 and 1990. The scope of work for this review, like those for prior evaluation reviews, called for basic descriptive statistics on the evaluations completed during the period. In addition, it asked for and the MSI team undertook:

- An assessment of the extent to which A.I.D.'s development project portfolio had been covered by evaluations during the period.
- An assessment of the degree to which evaluations carried out during the period focused on strategic, program and impact issues as distinct from issues of project management and implementation. Differences in the evaluation models and processes used in these two clusters of evaluations were of particular interest.
- An examination of the way in which the issues of gender, environmental concerns and the sustainability of A.I.D.-financed activities were handled in the evaluations of projects and programs during the period.

The full scope of work for this study is included as Annex A of this report.

¹ This study coincided with the announcement by A.I.D.'s Administrator of his intention to strengthen the role of evaluation in A.I.D. as part of an overall management improvement initiative. Subsequently, CDIE's mandate and resources were strengthened, particularly with respect to monitoring indicators of program performance.

B. Coverage and Methods

The FY89-FY90 evaluation review examined a database of 268 single- and multi-project evaluations which collectively addressed more than 300 projects. These 268 evaluations are themselves a subset of the 381 documents MSI received from A.I.D. Table 1-1 displays the frequency with which the full set of 381 documents contained evaluations versus other types of documents. As Table 1-1 indicates, 293 of these documents were evaluations, 77 were portions of evaluation reports and 11 documents did not qualify as evaluations. In reviewing these documents, the MSI team also noted that, at times, elements of an evaluation package, e.g., the evaluation report and the A.I.D. Evaluation Summary, have different CDIE card catalogue numbers. This made the task of selecting a final database more tedious than it might have been.

The 28 evaluations which were not included in the final database of 268 single and multi-project evaluations examined by this review included evaluations which were written-up in languages other than English; evaluations carried out before the study's FY89-90 time frame; and a group of evaluations of specialized A.I.D. activities, i.e., food aid programs (PL480); housing investment guarantees (HIGs) and commodity import programs (CIPs). These specialized programs are, perhaps, best examined through the kind of focused synthesis that A.I.D. recently carried out for a large number of food aid programs.

All evaluations which were included in the final database were evaluations for which a full report was available. The presence of an A.I.D. Evaluation Summary, a form used to transmit evaluations to CDIE as well as provide mission or office comments on an evaluation and a discussion of follow-up actions, was considered a desirable, but not necessary element of an evaluation package, for purposes of selecting the final set of evaluations.

The data for this study came from two sources. The first and most extensively used source was the set of FY89 and FY90 evaluations. The second data source was A.I.D.'s computerized database of information on project funding levels as well as the activity and special interest codes A.I.D. uses to describe project characteristics.

In order to link the evaluation review database created by this study with A.I.D.'s financial database, at least one seven-digit A.I.D. project number had to be associated with each evaluation. To the degree possible, the evaluation team sought to include only evaluations for which A.I.D. project numbers were available in the review². In the end, two types of exceptions to this rule were allowed. Sixteen broad multi-project evaluations, i.e., evaluations that examined five or more projects, were included in the database without accompanying financial data. In addition financial data turned out to be missing for two other evaluations.

² Project numbers were not consistently provided in evaluation reports. This was particularly true for evaluations that did not include an A.I.D. Evaluation Summary, which included a section that calls for the seven-digit project numbers of the projects covered by an evaluation. Extensive research was needed to locate correct project numbers for between 10% and 20% of the evaluations included in the database. The process of locating correct project numbers was further complicated by the fact that some evaluations displayed the seven-digit project number of a centrally funded project that paid for the evaluation. At times, these seven-digit numbers were assumed, during initial coding, to be the project numbers of the projects examined by the evaluation. Efforts to correct these initial coding errors and to locate and substitute correct project numbers proved to be very time consuming.

Table 1-1 Evaluations and Other Types of Documents Screened for Inclusion in the Study Sample

Type of Document	Number of Documents
<u>Included in the Sample</u>	
Complete Evaluations in English	268
<u>Excluded from the Sample</u>	
Evaluations in the Languages	7
Evaluations Completed Prior to the Period of Time Covered by the Study	4
"Self-Evaluation" Carried out by Implementation Contractors	7
Project Amendments	1
Design Documents	1
Discussion Papers	2
Evaluation Annexes Only	2
Evaluation Executive Summaries Only	4
Evaluation Summaries (PESs) Only	71
Evaluation of CIPs	2
Evaluation of PL480 Activities	7
Evaluation of HIGs	5
Total of All Documents Examined	381

The scoring instrument used to extract data from the 268 evaluations included in this study was developed based on a examination of the scoring instruments used in previous evaluation reviews and discussions with CDIE. This scoring instrument is presented in Annex B. It is divided into several discrete parts that focus on:

- The evaluation's scope, timing and sponsor;
- Evaluation teams -- their size, skills, host country participation, gender mix and A.I.D.'s direct involvement;

- The degree to which an evaluation was a participatory exercise in which potential evaluation users were involved;
- The stated purpose of an evaluation;
- The methods used to carry out the evaluation;
- The degree to which the evaluation's recommendations focused beyond the immediate concerns of a project, i.e., on lessons for similar projects or program level issues;
- Whether the project was viewed as having succeeded and whether there had been unplanned consequences;
- Methods for sustaining A.I.D.-financed activities reported upon by evaluations;
- Environmental impacts discussed therein, and
- Gender issues on which evaluations commented.

The MSI evaluation review team assigned to this study was trained in the use of this instrument in a two-day workshop at Florida State University. Raters learned to rate aspects of evaluations in a parallel manner through calibration exercises that improved their ability to work interchangeably on this project. Inter-rater reliability checks were also incorporated into the rating process used for the full set of 268 evaluations. Those few items where inter-rater reliability proved to be less than satisfactory, i.e., wherever fewer than four of the six raters agreed on a rating, were identified and dropped from the analysis.

Data from the two sources described above was analyzed using a standard data processing program (Statistical Package for the Social Sciences--SPSS--a commercially available and widely used data analysis program). To facilitate the full analysis of data collected during this study, including cross-tabulations that display responses on two variables simultaneously, information on several rating form variables had to be compressed into a more readily usable form. For example, a number of questions included in the basic study instrument allowed those who scored the evaluations to check "yes" on several multiple choice answers offered in connection with a particular variable. In order to run cross-tabulations these answers were transformed into exclusive choices. Most often this resulted in the creation of an analysis category entitled "both" or "combination of responses," as can be seen on tables provided throughout this report.

C. Evaluation Review Team

The evaluation review was carried out in Washington, D.C. and Tallahassee, Florida by an evaluation team made up of MSI staff and associates, including:

- **Molly Hageboeck**, Team Leader; a Senior Consultant and Director at MSI with over 20 years experience in the evaluation of economic development programs.
- **Monteze Snyder**, Assistant Professor and Director, International Public Management and Policy Center, Florida State University; 13 years experience in the design and evaluation of development programs.
- **Joseph James Gagnier**, owner and Director of Survey Design and Analysis; an MSI associate with extensive experience in statistical analysis.
- **Peter Doan**, Assistant Professor, Department of Urban and Regional Planning, Florida State University; experience in the design and evaluation of A.I.D. projects; teaches a course in project design and evaluation for developing countries.
- **Mark Renzi**, MSI Program Associate; staff experience with A.I.D.; evaluation and management consulting experience in developing countries.
- **Julie Koenen-Grant**, MSI Program Associate; Deputy Project Manager for MSI's contract with A.I.D. on Implementing Policy Change; Master's Degree candidate in International Management.
- **Al Bavone**, Florida State University; Doctoral candidate in Development Administration; U.S. domestic and international evaluation experience.
- **Joanne Snair**, Florida State University; Doctoral candidate in Public Administration; U.S. domestic evaluation synthesis experience as well as experience with benefit-cost and other types of program evaluations for service delivery projects.

Roles of the evaluation team during the evaluation reviews were as follows. Ms. Hageboeck and Dr. Snyder developed the evaluation rating instrument. Evaluation documents were rated by Dr. Doan, Mr. Renzi, Ms. Koenen-Grant, Mr. Bavone, Ms. Snair and Dr. Snyder. The data analysis plan for this review was developed and implemented by Ms. Hageboeck and Dr. Snyder. Mr. Gagnier prepared and processed the database. Ms. Hageboeck and Dr. Snyder wrote the report.

The evaluation review team wishes to acknowledge the assistance of Mr. James Painter, Mr. Peter Thiel and Mr. Frank Lin of A.I.D.'s Budget Office, without whose support the financial aspects of projects examined by A.I.D.'s FY89 and FY90 evaluations could not have been analyzed.

D. Structure of the Report

The remainder of this report on MSI's review of A.I.D.'s FY89 and FY90 evaluations is divided into seven chapters that present the study's findings. A final chapter offers MSI's conclusions and its recommendations to A.I.D. The coverage of these chapters is outline below:

- **Chapter Two** presents an analysis of the evaluations in the database, identifying the bureaus with which they were associated, the funding levels of the projects that were examined, etc.
- **Chapter Three** examines the degree to which evaluations were complete, i.e., their conformance with A.I.D. evaluation requirements.
- **Chapter Four** discusses the timing, coverage and purposes of the evaluations.
- **Chapter Five** examines the composition of the teams that carry out the evaluations.
- **Chapter Six** looks at the question of the degree to which A.I.D. and host country personnel participated in the evaluation process.
- **Chapter Seven** examines the conceptual frameworks, evaluation designs and methods used in the evaluations.
- **Chapter Eight** looks at the types of findings that emerge from the evaluations concerning project and program performance, sustainability and other cross-cutting issues.
- **Chapter Nine** presents MSI's conclusions and recommendations.

CHAPTER TWO

DESCRIPTION OF THE EVALUATION DATABASE

This chapter provides basic information on the set of evaluations that were examined in the course of MSI's review.

A. Number of Evaluations Examined

The database for this study consisted of 268 which examined over 300 A.I.D. projects. For the most part, evaluations included in this database were completed and delivered to CDIE, together with their required evaluation summaries, during FY89 and FY90. Sixteen percent of the evaluations included in the database were completed prior to the beginning of FY89, although in many of these cases, A.I.D. Evaluation Summaries were not signed until after the FY89 fiscal year began.

For all but 18 of the evaluations, the review team obtained financial information on the projects those evaluations examined. The 18 evaluations for which financial data were not obtained include an evaluation of a USAID/Honduras private enterprise project, an evaluation of a USAID/Egypt university linkage project and 16 evaluations that examined more than four projects simultaneously. Half of the evaluations in this latter group were broad efforts to synthesize the "lessons learned" from previous evaluations and related studies. The sixteen evaluations that make up this broad multi-project evaluation group are identified in Table 2-1. In lieu of project numbers, these 16 evaluations were assigned numeric codes that described the breadth of their coverage, i.e., 777 meant multiple projects (more than 4) in a single country; 888 indicated continent or bureauwide evaluations; and 999 connoted world-wide evaluations and evaluation syntheses.

B. Distribution of the Evaluations By A.I.D. Bureaus

Of the 250 evaluations for which MSI was able to develop financial profile information, 80% examined efforts developed and managed by A.I.D.'s regional bureaus and field missions. Table 2-2 displays the distribution of those evaluations by bureau. In this table, and throughout this report, bureau names and portfolios accord with A.I.D.'s 1991 reorganization.³

³ Under the reorganization, which took effect on October 1, 1991, A.I.D. established five geographic bureaus and three central bureaus. The five geographic bureaus cover Africa; Latin America and the Caribbean (LAC); Asia; the Near East, and Eastern Europe. All geographic regions with the exception of Eastern Europe, the newest region, submitted evaluations which are covered by this review. Among A.I.D.'s central bureaus, Private Enterprise has retained its pre-reorganization name. The Science and Technology Bureau has been renamed Research and Development (R&D), with no change in its portfolio. The Food and Voluntary Assistance Bureau has been renamed the Bureau for Food and Humanitarian Assistance (FHA), and the Office of Disaster Assistance, which was formerly a free standing unit, has been incorporated into the new FHA bureau.

Table 2-1. Evaluations in the FY89 and FY90 That Examined a Relatively Large Number of A.I.D. Projects Simultaneously

Coverage Code	Evaluation Document Number	Evaluation Title
Multiple Projects in a Single Country		
777	PDAAZ085	Philippine Economic Reform Program, Past Assistance and Future Recommendations
777	PDABB058	Salvadorian Foundation for Economic and Social Development
777	PDABA337	USAID/Kenya Private Sector Program
777	PNAAX211	The Sustainability of U.S. supported Health, Population and Nutrition Programs in Honduras, 1942-1986
777	PDAAAY457	Overall Program Review of USAID in Pakistan, 1982-1987
777	PDABD830	USAID/Guatemala, Forty Years on the Altiplano
777	PNABE652	OFDA, Ten Years of Disaster Preparedness Assistance
777	PNABF535	USAID/Dominican Republic, Returned Participants in the Agricultural Sector
777	PDAAZ022	USAID/Honduras, Honduras Rural Roads
777	PDAAX210	The Effectiveness and Impact of Policy-Based Cash Transfer Programs, The Case of Jamaica, 1981-1987
777	PNAAX220	The Effectiveness and Impact of Policy-Based Cash Transfer Programs, The Case of Costa Rica
Multiple Projects in a Single Geographic Region		
888	PDAAZ829	Agricultural Crop Diversification Export Promotion in Latin America
888	PDABC294	The Impact of Rural Credit Programs in Africa
Multiple Projects on a World-wide Basis		
999	PNAAX232	A.I.D.'s Experience with Democratic Initiatives
999	PNAAX227	A.I.D.'s Microenterprise "Stocktaking"
999	PNAAX230	A.I.D.'s Experience with Farming Systems Research and Extension

While projects and non-project assistance managed through A.I.D.'s regional bureaus and by its Central Bureau for Research and Development (R&D) were well represented by the evaluations in the database, only a small fraction of the work carried out by the Bureau for Food and Humanitarian Assistance (FHA) and its Private Enterprise Bureau (PRE) is addressed by the evaluations the review team examined.

The figures shown in Table 2-2 slightly understate the number of evaluations for several of the bureaus. The broad multi-project evaluations listed on Table 2-1, plus the USAID/Honduras and USAID/Egypt evaluations for which financial data were not obtained, when allocated on a bureau basis raise bureau totals as follows: the total for LAC rises to 76; Asia becomes 53; Africa becomes 69, Near East becomes 17, and FHA becomes 6.

MSI used these higher bureau totals to compare, on a percentage basis, the distribution of FY89 and FY90 evaluations to the distribution found in reviews carried out for FY85-86 and for FY87-88.⁴ As Table 2-3 indicates, these percentages have remained relatively stable for the LAC and Africa Bureaus over the years. The share of central bureau evaluations rose significantly, while the share for Asia and the Near East declined.⁵

Table 2-2. Distribution of Evaluations of One or a Few Related Activities by Bureau

Bureau	Number of Evaluations for Which Financial Data was Obtained	Percent
Asia	51	20%
Near East	16	6%
Latin America/Caribbean (LAC)	67	27%
Africa	67	27%
Research and Development (R&D)	39	16%
Food and Humanitarian Assistance (FHA)	5	2%
Private Sector (PRE)	5	2%
Total	250	100%

⁴ John Kean, et.al., Synthesis of A.I.D. Evaluation Reports: FY1985 and 1986, A.I.D. Evaluation Occasional Paper No. 16., and Hopstock, Paul et.al., Review of the Quality of A.I.D. Evaluations: FY1987 and 1988, A.I.D. Evaluation Occasional Paper No. 19, Washington, D.C., Agency for International Development, 1988 and 1989, respectively.

⁵ Had MSI included all of the FY89 and FY90 housing investment guarantee and food aid evaluations available in CDIE the central bureau share of FY89-90 evaluations would have been slightly higher, i.e., one or two percentage points.

**Table 2-3. Percentage of Evaluations by Region
Included in Successive Evaluation Reviews**

Bureaus	FY 85-86	FY 87-88	FY 89-90
Asia and the Near East	39%	28%	26%
Latin America and the Caribbean (LAC)	28%	34%	28%
Africa	29%	25%	26%
Central Bureaus	4%	13%	20%
Total	100%	100%	100%

In the remainder of this report, most tables that present data on a bureau basis show the 16 broad multi-project evaluations as a separate row rather than on a bureau basis. The USAID/Honduras and USAID/Egypt project evaluations for which the review team did not locate financial information were, however, integrated back into their bureaus in subsequent tables in this chapter. With the addition of these 18 projects, tables in subsequent chapters generally total to 268, the size of the full database examined by MSI review team.

Further details on the share of evaluations contributed by countries within each region and by the offices in A.I.D.'s R&D Bureau are provided in the following set of bureau-level tables:

- Table 2-4, which focuses on Asia, where 17 evaluations of USAID/Indonesia projects constitute 33% of the region's total;
- Table 2-5, which covers the Near East, and shows that 10 evaluations from USAID/Egypt dominate the evaluation reports submitted from this region; i.e., they represent 59% of the bureau's total.
- Table 2-6, profiles the Latin America/Caribbean region. As this table indicates, the share of evaluations contributed by Honduras, while large, i.e., 18%, is not as dominant on a percentage basis, as are those of Egypt in the Near East region;

Table 2-4. Distribution of Asia Bureau Evaluations

Asia Bureau	Number of Evaluations	Percent of Regional Total
Afghanistan	3	6%
Bangladesh	6	12%
India	5	9%
Indonesia	17	33%
Nepal	3	6%
Pakistan	6	12%
Philippines	6	12%
Sri Lanka	2	4%
Thailand	1	2%
South Pacific Regional	1	2%
Asia Regional	1	2%
Total	51	100%

Table 2-5. Distribution of Near East Bureau Evaluations

Near East Bureau	Number of Evaluations	Percent
Egypt	10	59%
Jordan	3	17%
Morocco	1	6%
Tunisia	1	6%
Yemen	1	6%
Near East Regional	1	6%
Total	17	100%

Table 2-6. Distribution of Latin America and Caribbean Bureau Evaluations

Latin America Caribbean Bureau	Number of Evaluations	Percent of Regional Total
Belize	2	3.0%
Bolivia	6	9.0%
Costa Rica	1	1.5%
Dominican Republic	4	6.0%
Ecuador	6	9.0%
El Salvador	5	7.5%
Guatemala	7	10.0%
Haiti	1	1.5%
Honduras	13	18.0%
Jamaica	6	9.0%
Panama	1	1.5%
Peru	1	1.5%
Caribbean Regional	6	9.0%
Central American Regional	6	9.0%
ROCAP	1	1.5%
Latin America/Caribbean Regional	2	3.0%
Total	68	100%

- Table 2-7, on the following page, covers Africa and shows a more evenly distributed pattern of evaluation than was found elsewhere. There are, however, two exceptions: 15 evaluations of regional projects for Africa, or 23% of the bureau's total, and 8 evaluations of USAID/Somalia projects, which represent 17% of Africa's total; and
- Table 2-8 focuses on the R&D Bureau. It indicates that while R&D/Population contributed the largest share of evaluations (33%), other offices, notably R&D/Agriculture and R&D/Nutrition, contributed quite a few evaluations to the bureau's total.

Table 2-7. Distribution of Africa Bureau Evaluation

Africa Bureau	Number of Evaluations	Percent of Regional Total
Benin	1	1.5%
Botswana	3	4.0%
Burkina Faso	2	3.0%
Cameroon	2	3.0%
Djibouti	1	1.5%
Gambia	1	1.5%
Ghana	1	1.5%
Kenya	2	3.0%
Lesotho	1	1.5%
Liberia	2	3.0%
Madagascar	1	1.5%
Malawi	3	4.0%
Mali	2	3.0%
Mauritania	2	3.0%
Niger	2	3.0%
Nigeria	1	1.5%
Senegal	1	1.5%
Somalia	8	17.0%
South Africa	1	1.5%
Sudan	1	1.5%
Swaziland	6	10.0%
Togo	1	1.5%
Zambia	1	1.5%
Zaire	3	4.0%
Southern Africa Regional	3	4.0%
Africa Regional	15	23.0%
Total	67	100%

Table 2-8. Distribution of R & D Bureau Evaluations

Research and Development Bureau	Number of Evaluations	Percent of Bureau Total
Agriculture	8	21%
Education	2	5%
Energy	3	8%
Forestry, Environment and Natural Resources	1	3%
Health	4	10%
Nutrition	6	15%
Population	13	33%
Rural and Institutional Development	2	5%
Total	39	100%

As these tables indicate, several USAID missions, as well as one office in the R&D Bureau, were found to have submitted an unusually large number of evaluations in FY89 and FY90. Prior evaluation reviews have also noted instances where missions have turned in relatively large numbers of evaluations in a given two-year period.

As to the evaluations that covered projects developed and managed by other central bureaus:

- 2 of the 5 evaluations of PRE Bureau projects focused on projects managed by that Bureau's Investment Office. PRE's Offices for Housing and Urban Affairs, its Emerging Markets Office and its Office of Small, Micro and Informal Enterprises each administered one of the projects that was covered by an evaluation in the database.
- In the FHA Bureau, all five of the evaluations focused on projects managed by the Office for Private and Voluntary Cooperation.

While the evaluations identified on a regional basis in Tables 2-4 through 2-7 present a broad picture of the level of evaluation activity in each region, they do not draw a sharp distinction between evaluations of bilateral projects (or "mission-owned" projects) and evaluations of regional projects. Table 2-9 makes this distinction, which is utilized in subsequent chapters of this report to highlight the degree to which host country personnel are involved in those evaluations where their participation is most logical. As Table 2-9 suggests, the share of regional projects evaluated in the LAC and Africa Bureaus is quite a bit higher than is the case for the Asia or Near East Bureaus.

**Table 2-9. Degree to which Evaluations for Regional Bureaus
Examined "Mission-owned" Projects**

Bureau	Evaluations of "Mission owned" Projects	Evaluations of Regional Office and Bureau Level Projects	Total Evaluations
Asia	49	2	51
Near East	16	1	17
Latin America and the Caribbean	52	16	68
Africa	19	18	67

C. Project Versus Non-Project Assistance

The evaluations examined in the course of this review overwhelmingly focused on projects as distinct from non-project assistance efforts. Of the 251 evaluations for which such information is available, 241 (96%) examined projects while 10 (4%) evaluated non-project assistance activities. The share of non-project assistance evaluations in the FY89 and FY90 database was somewhat higher than was found in the FY87-88 review. Of the ten non-project activities examined by evaluations in the database, 8 were from countries in the Africa region; one came from USAID/Indonesia, and the final evaluation of a non-project activity came from the Caribbean regional office. Given the small number of non-project assistance evaluations in the FY89 and FY90 database, they were not given special treatment in this review. In subsequent chapters these evaluations are treated in the same manner as project evaluations.

D. The Duration of Projects and Non-Project Assistance Efforts

Information about the beginning and ending dates of the projects and non-project assistance efforts presented in the evaluation reports indicated that, on average, A.I.D.-funded activities lasted 5.8 years. The range for activity length was found to be from less than one year to 16 years. Table 2-10 provides a frequency distribution on the length of projects examined through evaluations included in this review. For a total of 31 evaluations, information on project length was not available.

Table 2-10. Length of Projects Examined by the Evaluations

Length of the Longest Project Evaluated	Number of Evaluations	Percent of Evaluations
Less Than 1 Year	17	7%
1 Year	11	5%
2 Years	7	3%
3 Years	16	7%
4 Years	25	11%
5 Years	35	15%
6 Years	33	14%
7 Years	28	12%
8 Years	19	8%
9 Years	17	7%
10 Years or More	29	11%
Evaluations for which data was available	237	100%

E. Sectoral Affiliation of Projects Examined by the Evaluations

In recent years, A.I.D. has shifted its approach to identifying the sectoral nature of its projects. The system it currently uses identifies all aspects of a project using a set of activity codes. Projects may be coded as having a number of subsector characteristics, including characteristics from a number of sectors.

Using this system, the MSI team prepared Table 2-11, which shows the frequency with which evaluations in the database involve various aggregate activity codes. Table 2-12 provides a more detailed, subsector view of this distribution; 141 evaluations, or 53%, of the full set of 268 evaluations are included in this table. Projects covered by the other 127 evaluations had not been assigned activity codes by A.I.D.⁶ The sum of the observations on Tables 2-11 and 2-12 exceed the number of evaluations in the database because some evaluations were assigned more than one activity code.

As these tables indicate, evaluations frequently focused on projects linked to the agricultural sector. Health was also a frequent descriptor of projects examined by the evaluations. Education, private enterprise and natural resources codes were also used with considerable frequency to describe projects which were examined by these evaluations.

While MSI used the new activity codes were used as the primary means of characterizing the sectoral nature of the evaluations, some information was also obtained from A.I.D.'s budget office on functional account codes, which A.I.D. had previously used to track sectoral activity

⁶ MSI coded multi-project evaluations as including a particular sector if any project in the set that had been examined the multi-project evaluation had that particular sectoral code.

related to the FY89 and FY90 evaluations. Table 2-13 presents this information and compares it to the information on the distribution of evaluations by functional accounts in prior evaluation reviews. Roughly a quarter of the evaluations it examined were found to include projects that were funded from several functional accounts, including all of the projects involving education activities or the private sector.

Table 2-11. Frequency with Which Aggregate Activity Codes Were Found in 141 Evaluations

Aggregate Activity Codes	Number of Evaluations	Percent of Evaluation with Activity Codes
Agriculture	100	71%
Education	58	41%
Energy	4	3%
Infrastructure	2	1%
Health	70	50%
Human Rights/Democratic Initiatives	10	7%
Natural Resources	34	24%
Nutrition	25	18%
Private Enterprise	40	28%
Population	29	21%
Public Sector	20	14%
Project Development	3	2%

F. Special Interest Codes

The system A.I.D. uses to identify the sectoral characteristics of its projects also contains an element that is used to assess whether projects respond to special Agency and Congressional concerns. As is the case with sectoral activity codes, projects may be assigned a number of different codes. The review team coded an evaluation as having a special interest as long as at least one project considered by that evaluation was coded as having a particular special interest.

Of the special interest codes, training was the code most frequently associated with projects covered by evaluations the team reviewed. The frequency with which training and other special interest codes were associated with the 141 evaluations for which such data was obtained is shown below in Table 2-14.

**Table 2-12. Activity Codes with Which Projects Examined
by 141 Evaluations Were Associated**

Activity Codes	Number
AGRICULTURE GROUP	
Agnbusiness	10
Agricultural Credit	8
Agricultural Education	15
Irrigation	10
Agricultural Land Use and Settlement	5
Agricultural Marketing	11
Agricultural Management Planning and Policy	15
Pest Management	5
Agricultural Technology Development and Dissemination	21
Sub-Total Agriculture	(100)
EDUCATION GROUP	
Basic Education for Adolescents and Adults	7
Education for Children	10
Human Resources Development for Education?	14
Human Resources Development for Individuals	14
General Public Education and Extension	3
Sub-Total for Education	(58)
ENERGY GROUP	
Energy Management, Planning, and Policy	4
Sub-Total for Energy	(4)
INFRASTRUCTURE GROUP	
Rural Roads	2
Sub-Total for Infrastructure	2
HEALTH GROUP	
Child Spacing/High Risk Births	7
Diarrheal Disease Control/Oral Rehydration	10
HIV/AIDS	3
Immunization	9
Malaria	9
Women's Health	7
Acute Respiratory Infection	7
Health Systems Development	9
Vector Control	4
Water Quality	5
Subtotal for Health	(70)
HUMAN RIGHTS GROUP	
Democratic Institution Building	7
Strengthening Legal Systems	2
Technical Electoral Assistance	1
Subtotal for Human Rights	(10)
NATURAL RESOURCES GROUP	
Forestry	11
Hazardous Waste	1
Agricultural Land Development	1
Environmental Management	10
Soils	4
Water Quality Improvement	3
Water Resources Management	4
Subtotal for Natural Resources	(34)

Activity Codes	Number
NUTRITION GROUP	
Breast feeding	6
Growth Monitoring and Weaning Foods	8
Nutrition Management Planning & Policy	3
Vitamin A	4
Nutrition of Women	4
Subtotal for Nutrition	(25)
PRIVATE ENTERPRISE GROUP	
Business Development Promotion	13
Financial Markets	13
Trade and Investment Promotion	14
Subtotal for Private Enterprise	(40)
POPULATION GROUP	
Family Planning Contraceptives	5
Family Planning Program Development	12
Family Planning Service Delivery	12
Subtotal for Population	(29)
PUBLIC SECTOR GROUP	
Administration and Management	20
Subtotal for Public Sector	(20)
PROJECT DEVELOPMENT GROUP	
Project Development and Support	3
Subtotal for Project Development	3
TOTAL NUMBER OF ACTIVITY CODE ASSOCIATIONS FOR EVALUATIONS	395

Table 2-13. Relationship of Evaluation Coverage to Functional Accounts

Functional Accounts	Evaluations of Projects Funded Primarily from One Functional Account		Evaluations of Projects Funded Exclusively from the Equivalent of a Single Functional Account
	FY85-86 Review	FY87-88 Review	FY89-90 Review
Agriculture, Rural Development and Nutrition	52%	38%	56%
Educational and Human Resource Development	14%	7%	---
Health and Population	14%	26%	13%
Special Development Account, i.e., Private Sector, Energy and the Environment	19%	25%	---
Projects Funded from a Combination of Several Accounts			25%

Table 2-14. Frequency with Which Special Interests Appear in 141 Evaluations

Special Interest Codes	Number of Evaluations	Percent of Evaluations with Special Interest Codes
Training	94	67%
Institution Building	75	53%
Research	73	52%
Private Voluntary Organization	39	28%
Policy Reform	30	21%
Women in Development	29	21%

G. Sources of Funds

The project and non-project activities examined by the evaluations were funded through a number of different foreign assistance accounts, including the Development Assistance account, which until recently has been divided into several functional accounts as was shown in Table 2-13; the Economic Support Fund; and a number of smaller foreign assistance accounts.

- 159 (or 63%) of the 251 evaluations for which financial profile information was available examined projects funded through the Development Assistance account (DA), which constitutes 27% of all U.S. bilateral assistance;
- Another 22 (or 9%) of the evaluations examined projects funded through the Economic Support Fund (ESF), which accounts for roughly 66% of U.S. bilateral assistance.
 - Of the 22 evaluations that focused on ESF- financed activities, 15 were from the Near East Bureau; Africa and Asia each contributed 3; and the final 2 came from the Latin America and Caribbean Bureau.
- 3 (or 1%) of the evaluations focused on projects funded through the International Disaster Assistance account, a relatively small element of U.S. bilateral assistance.
- The final 67 evaluations (27% of the total) were funded through a combination of these accounts, with the most frequent funding combination in this category being DA and ESF.

- The largest set of evaluations in this final cluster, 40, came from Africa. The Latin America and Caribbean Bureau contributed another 16.

H. Funding Levels of the Activities Evaluated

Single-project and multi-project evaluations for which MSI acquired financial information covered projects with funding levels ranging from under \$999 thousand to over \$100 million, as Table 2-15 indicates. At the low end of this spectrum, 7 of the 13 evaluations in the under \$1 million category examined projects in the Africa Bureau. At the high end, 16 of the 18 evaluations that examined activities worth over \$100 million came, primarily from the Asia and Near East Bureaus.

The average value of the projects examined by the evaluations was \$26.2 million. On a bureau basis, the average value of these projects ranged from a high of \$69 million for the Near East to a low of \$16 million for Latin America and the Caribbean, as Table 2-16 shows.

Table 2-15. Value of Projects and Non-Project Assistance Activities Covered by the Evaluations

Value of Projects and Non-Project Activities	Number	Percent
Under \$999,999	13	5%
\$1 million \$9,999,999	74	30%
\$10 million to \$24,999,999	88	35%
\$25 million to \$49,999,999	38	15%
\$50 million to \$99,999,999	19	8%
Over \$100 million	18	7%
Total Number of Evaluations for which Data was available	250	100%

Table 2-16. Value of Projects and Non-Project Activities Covered by Evaluations

Bureau	Number of Evaluations	Total Value of Projects and Non-Project Activities Covered by Evaluation (In Thousands)	Average Value of Projects and Non-Project Activities Covered by Evaluation (In Thousands)
Asia	51	\$2,137,793	\$41,917
Near East	16	1,112,610	69,538
Latin America and the Caribbean	67	1,050,326	15,676
Africa	67	1,200,044	17,911
R & D	39	844,835	21,662
FHA	5	94,976	18,995
PRE	5	114,725	22,945
Total/Average	250	\$6,555,309	\$26,221

I. Share of the A.I.D. Portfolio Examined by FY 1989 and FY 1990 Evaluations

In order to assess the degree to which the evaluations A.I.D. had carried out in FY 1989 and FY 1990 covered the agency's portfolio, MSI queried A.I.D.'s budget office concerning the total number of projects active in FY 1990 and the total value of these projects. The answers that were received were compared to the number of evaluations MSI had examined and the total value of the projects assessed through those evaluations to create the rough estimates of the fraction of its portfolio A.I.D. evaluates.

The degree to which the evaluations carried out by regional and central bureaus, together with the 16 broad multi-project evaluations examined during this review cover A.I.D.'s portfolio is presented by Table 2-17. As this table suggests, the evaluations examined during the review covered a two year time period, whereas the study's figures on the number of activities in A.I.D.'s portfolio are from a single year, FY90. In order to estimate portfolio coverage, i.e., the percentage of the total number of activities, the answer was divided in half to represent portfolio coverage in a single year. Using this approach, the review team estimated that roughly 8% of A.I.D.'s portfolio was examined through evaluations in each of the two fiscal years covered by this study. Since a number of bureau evaluations actually covered more than one project, and given that Table 1-1 indicated that evaluations were carried out that were not included in this review, an 8% estimate for portfolio coverage may be somewhat conservative.

**Table 2-17. Active Projects Covered by FY89-FY90 Evaluations
as a Percent of All Projects Included
in Bureau Portfolios**

A.I.D. Bureau	Number of Active Projects During FY 90	Number of Projects Covered by Evaluation over a Two Year Period	Percent of Projects Covered by Evaluations
Europe	24	-	-
Asia	279	51	18%
Near East	143	17	12%
Latin America and the Caribbean	590	68	12%
Africa	556	67	12%
Research and Development (R&D)	205	39	19%
Food and Humanitarian Assistance (FHA)	72	5	7%
Private Sector (PRE)	24	5	21%
Policy Directorate	2	-	-
Other ⁷	7	-	-
Total	1,902	252	13% (for two years or 6.5% on an annualized basis)
Additional projects covered by 16 broad multi-project Evaluations. Each of which covered at least 4 projects. (i.e., 16 x 4)		64	
Expanded Total	1,902	316	17% (or 8% on an annualized basis)

Looking at A.I.D.'s evaluation coverage of its portfolio from a financial perspective, Table 2-18 indicates that on an annualized basis that coverage was also approximately 8%. Table 2-18, which considers the question of the value of the portfolio covered by evaluations excludes the value of the 16 broad multi-project evaluations examine by this study; the two bureau evaluations for which financial data could not be obtained, and a number of other evaluations (identified in Table I-1) which were not included in the study's database. Thus, an estimate of 8% coverage of the A.I.D. portfolio on a value basis is once again conservative.

⁷ Includes projects funded by Donor Coordination, TDP, Economic Affairs and the Special Requirements Fund.

**Table 2-18. Value of Projects Covered by FY89-90
Evaluations as a Percent of
Bureau Portfolios**

	Value of All Active Projects in FY 90	Value of Projects Covered by 250 Evaluations of Projects carried out over two years, for which a Value could be Established	Percent of Funds Covered by Evaluations
Europe	\$ 699,169	-	-
Asia	8,080,819	\$2,137,793	26%
Near East	11,520,251	1,112,610	10%
Latin America and the Caribbean	6,527,319	1,050,326	16%
Africa	6,377,963	1,200,044	19%
R&D	3,939,439	844,835	21%
FHA	1,104,107	94,976	9%
PRE	210,051	114,725	55%
Policy Directorate	58,120	---	---
Other	53,219	---	---
TOTAL	\$38,570,457	\$6,555,309	17% (for two years or 8% on an annualized basis)

CHAPTER THREE

THE COMPLETENESS OF A.I.D. EVALUATIONS

As the A.I.D. Evaluation Handbook makes clear, an evaluation report is not complete unless it contains an evaluation scope of work; a discussion of the data collection and analysis methods that were used, and an A.I.D. Evaluation Summary.⁸ The completeness of A.I.D.'s FY89 and FY90 evaluations, in all of these regards, is summarized below.

A. Evaluation Scopes of Work

The preparation of an evaluation scope of work is the responsibility of the organizational unit that sponsors an evaluation. As the A.I.D. Evaluation Handbook indicates:

" The scope of work is critical to obtaining the type of information needed. It must articulate as clearly and precisely as possible the questions managers need addressed through an evaluation. Experience clearly demonstrates that the time and effort required for writing a sound scope of work acceptable to host country as well as A.I.D. managers is easily justified by improvements in the quality, utility and acceptance of the evaluation results."⁹

In reviewing the 268 evaluations in the FY89 and FY90 evaluation database, the review team noted whether evaluation reports contained full or partial scopes of work. As Table 3-1 indicates, 74% of the evaluations were found to contain either a full or partial scope of work. The fact that scopes of work were not included in evaluation reports cannot be taken to mean that such documents did not exist. It simply means that 26% of the evaluations failed to comply with this A.I.D. evaluation requirement.

Single-project evaluations, as a group, were found to contain scopes of work more frequently than were multi-project evaluations. As Table 3-1 illustrates, one of the clusters of projects for which scopes of work seemed to be lacking was the set of 16 broad, multi-project evaluations that were listed in Table 2-1. In addition, reviews of "lessons learned," which tend to draw upon the findings of existing evaluations, frequently failed to include scopes of work. Among the regional bureaus, evaluations that focused on projects in Africa seemed to lack scopes more often than did evaluations of projects in other regions. The same was true for PRE, a bureau for which the database included only a small sample of evaluations.

⁸ A.I.D.'s FY85-86 evaluation review, which examined the completeness of these reports in greater detail used a longer list of items to judge report adequacy, including the presence of a table of contents and executive summary. MSI's rating form examined the subset of items included on the FY85-86 completeness list which seemed to be critical for A.I.D. evaluations, as opposed to reports in general.

⁹ Agency for International Development, A.I.D. Evaluation Handbook, Supplement to Chapter 12, A.I.D. Handbook 3, Project Assistance, Washington, D.C., 1989, Section 3.5.

**Table 3-1. Frequency with which Evaluations
Had a Full or Partial Scope of Work**

Bureaus	Evaluation for Which There was a Full or Partial Scope of Work	Total Number of Bureau Evaluations	Percent of Evaluations for Which There was a Scope of Work
Asia	42	51	82%
Near East	13	17	72%
Latin America and the Caribbean (LAC)	53	68	78%
Africa	45	67	67%
R & D	30	39	77%
FHA	4	5	80%
PRE	2	5	40%
Large Multi-project Evaluations	9	16	56%
All Evaluations	198	268	74%

Evaluations that examined projects with a high dollar value had scopes of work more frequently, on a percentage basis, than did evaluations that examined projects of lesser value. Of the evaluations that examined projects valued at under \$1 million, 61% contained scopes of work. At the other end of the spectrum, 83% of the evaluations of projects valued at over \$10 million contained scopes of work.

Recognizing A.I.D.'s interest in taking gender issues into consideration as projects are designed and in the course of evaluations, the review team noted evaluations indicating that women had been consulted as the scope of work for an evaluation was developed. The number of times such references were found may well understate the frequency with which this type of consultation occurred, since there is no requirement for scopes of work or evaluations to state whether women participated developing the initial plans for an evaluation. With that caveat in mind, MSI found that 12 evaluations (5%) reported that women had been consulted as evaluation scopes of work were prepared.

Comparing the evaluations included in this FY89 and FY90 evaluation review to evaluations included in earlier reviews, MSI found that the frequency with which scopes of work were included in evaluation reports rose steadily. As Table 3-2 indicates, that improvement was substantial.

Table 3-2. Share of Evaluations that Include Scopes of Work Across Several Evaluation Reviews

Evaluation Reviews	Number of Evaluations Received	Percent that Included a Scope of Work
FY 85-86 Review	212	49%
FY 87-88 Review	287	54%
FY 89-90 Review	268	74%

B. Descriptions of Evaluation Methods in Evaluation Reports

The methods used for gathering and analyzing evaluation data directly affect the validity and credibility of evaluation conclusions and recommendations. For this reason, the A.I.D. Evaluation Handbook requires that evaluation reports include a description of an evaluation's methodology, i.e., a brief summary of what information was gathered for an evaluation, how that information was obtained, and how it was analyzed.

In scoring the evaluations, the review team made a distinction between (1) evaluations which presented a discussion of both data collection and data analysis methods and (2) those which discussed only data collection. While evaluations in the former category were considered to be in full compliance with A.I.D.'s requirement, those in the latter group were not.

Overall, 234 (87%) of A.I.D.'s FY89 and FY90 evaluations were found to contain explanations of their methodology. Thirty (11%) of these 268 evaluations included methodology sections which discussed both data collection and data analysis and were thereby in full compliance with A.I.D.'s requirements. Another 204 (76%) of the evaluations included methodology sections that only discussed data collection.

Table 3-3 displays information on a bureau basis with respect to the inclusion of methodology sections in evaluation reports. Most of the bureaus for which the review team had relatively large numbers of evaluations did well with respect to this requirement. Evaluations of projects in the Africa Bureau did less well than the overall average in terms of including discussions of evaluation methods.

With respect to evaluation types, approximately the same proportion of interim and final or ex-post evaluations contained methodology sections. However, evaluations differed, as a function of number of different skills present on an evaluation team, as to whether they presented methods sections. 95% of the evaluations carried out by teams that included individuals with four or more skills also included methodology sections. Teams that were characterized as offering only one skill presented evaluation methods discussions 65% of the time.

Table 3-3. Degree to Which Evaluations Described the Data Collection and Analysis Methods That Were Used

Bureaus	Full Description of Evaluation Methods (Data Collection and Analysis Procedures)	Partial Description of Methods (Data Collection Only)	Total Number of Evaluations With Methods Sections	Total of Bureau Evaluations	Percent of with Methods Sections
Asia	8	39	47	51	92%
Near East	1	13	14	17	82%
Latin America and the Caribbean (LAC)	13	48	61	68	90%
Africa	1	52	53	67	79%
R & D	2	35	37	39	95%
FHA	1	4	5	5	100%
PRE	--	4	4	5	80%
Large Multi-Project Evaluations	4	9	13	16	81%
All Evaluations	30	204	234	268	87%

Table 3-4 compares the findings for this evaluation review to previous evaluation reviews with respect to the inclusion of evaluation methods discussions.

Table 3-4. Frequency with which A.I.D. Evaluations Discuss Evaluation Methods

Evaluation Methods	Number of Evaluations Received	Percent that Included a Discussion of Evaluation Methods
FY 85-86 Review	212	75%
FY 87-88 Review	287	76%
FY 89-90 Review	268	87%

As Table 3-4 suggests, evaluations are improving in terms of the frequency with which they include discussions of evaluation methods. This finding is tempered, however, by evidence in Table 3-3 which indicates that the majority of these discussions are incomplete. They do not conform to A.I.D.'s requirement to describe data analysis as well as data collection procedures.

C. A.I.D. Evaluation Summaries

A.I.D. Evaluation Summaries, which present A.I.D.'s comments on an evaluation as well as planned follow-up actions, have been required in one form or another for nearly twenty years. In the FY89 and FY90 database, 132 (49%) of the 268 evaluations were accompanied by A.I.D. Evaluation Summaries.

In this regard it is worth noting that the boxes of evaluation documents MSI received from A.I.D. contained a number of A.I.D. Evaluation Summaries which were not attached to evaluation reports.¹⁰ In some cases, the review team was able to determine that some evaluation reports and A.I.D. Evaluation Summaries that had different A.I.D. card catalogue numbers actually referred to the same evaluation. Where reports and summaries referred to the same evaluation, the review team linked them together and treated them as one document.

Of the 132 FY89 and FY90 evaluations that included A.I.D. Evaluation Summaries, 44 (33%) indicated that their preparation had been preceded by a formal evaluation review in which A.I.D. staff examined the evaluation's findings and recommendations and reached decisions concerning follow-up actions.

Table 3-5 presents a frequency distribution by bureau showing that evaluations of projects in the Asia Bureau were accompanied by evaluation summaries 71% of the time. This was far more frequent than the average for the Agency as a whole. Interim evaluations and evaluations of single projects did slightly better than the Agency-wide average of 49% for turning in A.I.D. Evaluation Summaries, while final and ex-post evaluations and multi-project evaluations did slightly worse.

Comparing the frequency with which evaluations in the FY89 and FY90 database were accompanied by A.I.D. Evaluation Summaries to the findings of prior evaluation reviews, it appears that A.I.D.'s performance is worse today than it was five years ago. There has been overall decline in the frequency with which A.I.D. Evaluation Summaries are submitted, as Table 3-6 illustrates.

Looking across several measures of compliance, it appears that there may be some tendency for evaluations that are complete in one regard to be complete in other ways as well. Thus, for example, there was a greater tendency for evaluations to include A.I.D. Evaluation Summaries when a scope of work was also present. Of the 132 evaluations for which evaluation summaries were submitted, 78% also contained either a complete or partial scope of work.

¹⁰ A.I.D.'s FY87-88 evaluation review also pointed out the fact that A.I.D. evaluations and their Evaluation Summaries were apparently being entered into A.I.D.'s library using different card catalogue numbers.

**Table 3-5. Frequency with which Evaluation Reports
were Accompanied by Evaluation Summaries**

Bureaus	Number for Which an Evaluation Summary was Submitted	Total of All Bureau Evaluations	Percent for Which an Evaluation Summary was Submitted
Asia	36	51	71%
Near East	9	17	53%
Latin America and the Caribbean (LAC)	32	68	46%
Africa	34	67	51%
R & D	18	39	46%
FHA	--	5	--
PRE	1	5	20%
Large Multi-project Evaluations	2	16	13%
All Evaluations	132	268	49%

**Table 3-6. Share of the Evaluations that are
Accompanied by A.I.D. Evaluation Summaries**

Evaluation Reviews	Number of Evaluations Reviewed	Percent that Included A.I.D. Evaluation Summary
FY 85-86 Review	212	68%
FY 87-88 Review	287	64%
FY 89-90 Review	268	49%

CHAPTER FOUR

EVALUATION COVERAGE AND OBJECTIVES

As the A.I.D. Evaluation Handbook points out:

"The primary purpose of monitoring and evaluation is to assist the managers of development activities to make well-informed decisions. Monitoring and evaluation must meet the information requirements of managers at different organizational levels within the Agency and, correspondingly, the information requirements of their counterparts. Although the types of information needed by managers at different levels are often similar or complimentary, each organizational level also has its own specific information requirements. Therefore, it is A.I.D.'s policy to support a variety of monitoring and evaluation activities to obtain the range of information needed by Agency and counterpart managers."¹¹

In practice, the types of management information that evaluations provide are a function of their sponsorship, timing, coverage and objectives. This section reviews MSI's findings with respect to each of these defining characteristics.

A. The Sponsorship of A.I.D. Evaluations

There is a great deal of flexibility within A.I.D.'s evaluation system with respect to evaluation sponsorship. While the system's basic design encourages line managers to arrange for the evaluation of their own projects on an interim and final basis, other organizational units also become involved in these activities. Bureau-level evaluation offices in regional and central bureaus, as well as CDIE itself, often sponsor evaluations. Several ambitious evaluations of entire mission portfolios were sponsored by the missions themselves.

This review of A.I.D.'s FY89 and FY90 evaluations illustrates the diversity of evaluation sponsorship found within A.I.D. As Table 4-1 indicates, overseas missions independently sponsored 163 (61%) of the 268 evaluations in the database. Regional and central bureaus which have direct responsibility for projects independently sponsored 73 (27%) of these evaluations. In addition, a number of instances were found where more than one organizational unit participated as an evaluation sponsor.

¹¹ A.I.D. Evaluation Handbook, Section 1. This perspective on the purposes of evaluation in A.I.D. was in a 1990 A.I.D. paper on "The A.I.D. Evaluation System: Past Performance and Future Directions", which was produced by what was at that time the Agency's Bureau for Program and Policy Coordination.

**Table 4-1. Organizational Sponsors of Evaluations
Examined Through This Study**

Organizational Unit within A.I.D.	Number	Percent
USAID Mission Acting Alone	163	61%
Regional Bureaus Acting Alone	22	8%
S & T Acting Alone	38	14%
Other Central Bureau (FHA or PRE) Acting Alone	13	5%
CDIE Acting Alone	7	3%
Combinations of any of the above, acting together	12	4%
Can't Tell	13	5%
Total	268	100%

In order to determine the frequency with which evaluations are "self-sponsored", i.e., financed by the bureau that is responsible for the funding and management of the project or projects those evaluations examined, MSI compared information on evaluation sponsorship to data on project "ownership" for the 251 evaluations on which it had financial data. Table 4-2 presents the results of this comparison. As the table suggests, 85% of A.I.D.'s evaluations are "self-sponsored." Evaluations that were not "self-sponsored" include, by way of example, six reviews of "lessons learned" sponsored by CDIE on behalf of the Agency as a whole.

**Table 4-2. Relationship Between Project
Financing and Evaluation Sponsorship**

Bureaus	Total Number of Evaluations	Self-Sponsored Evaluation (Evaluations sponsored by the Bureau that funded the activities)	Self-Sponsored Evaluations as Percent of All Evaluations for a Bureau
Asia	51	42	82%
Near East	17	16	94%
Latin America and the Caribbean (LAC)	67	57	85%
Africa	67	58	87%
R & D	39	35	90%
FHA	5	3	60%
PRE	5	3	60%
TOTAL	251	214	85%

MSI's findings concerning the "self-sponsorship" of evaluations are completely consistent with the guidance A.I.D. provides to its bureaus and missions. As noted already, the basic purpose of evaluation in A.I.D. is to provide managers, at various organizational levels, with the information they need.

B. Evaluation Types and Timing

A.I.D.'s evaluation system recognizes the immediate and longer-term purposes evaluations can serve, but it does not claim that every evaluation can serve all potential purposes, or be of equal utility to project managers overseas and those who supervise geographic and technical portfolios from Washington. Over the years, several categories of A.I.D. evaluations have evolved. Expectations concerning these evaluations are summarized below. As subsequent sections of this report indicate the coverage of evaluations did not always conform to expectations.

- **Interim evaluations** are undertaken during the financial life of projects and programs. These evaluations are often designed to provide information that can be used to guide on-going activities.
- **Final evaluations** are undertaken at the end of the financial life of projects, or when a follow-on project is contemplated. "Final" evaluations may take place "even though the project may have a year or more to run before its (financial termination date)."¹²
- **Ex-post evaluations** are undertaken at some point after A.I.D.'s funding for a project has ended. These evaluations often move beyond implementation issues to ask whether objectives were achieved and whether activities and benefits are being sustained in the absence of A.I.D. funding. Both final and ex-post evaluations are expected to provide information that helps with the formulation of future projects and programs.
- **Reviews of "lessons learned"** usually draw upon a number of evaluations in a particular geographic area or technical field. These evaluation studies attempt to aggregate the information generated by evaluations of a number of projects and programs and present it in a form that is of potential use to policy makers as well as to those who are designing new projects and programs.

In numerical terms, these evaluation categories have formed something of a pyramid for which large numbers of interim evaluations serve as a foundation. Historically, the vast majority of interim evaluations have dealt with a single project or program. Final and ex-post evaluations form a second and smaller tier of the pyramid. At the top rest the relatively few reviews of "lessons learned" that A.I.D. completes each year. In contrast to interim evaluations, reviews of "lessons learned," almost by definition, tend to examine multiple projects or programs.

¹² Ibid., Appendix D, p.8.

Of the 268 FY89 and FY90 evaluations, 159 (59%) turned out to be interim evaluations, as Table 4-3 indicates. Final evaluations constituted the second largest cluster, accounting for 25% of the database. Ex-post evaluations were few in number as were reviews of "lessons learned". In addition, 9% of the database could not be classified according to these four basic evaluation types.

Table 4-3. Distribution of FY 89 and FY 90 Evaluations by Type

Type of Evaluation	Number	Percentage
Interim	159	59%
Final	68	25%
Ex Post	6	2%
Lessons Learned	13	5%
Other	10	4%
Can't Tell	12	5%
TOTAL	268	100%

Comparing these findings to the findings of previous evaluation reviews, it appears that the share of interim evaluations was slightly lower than had been the case in prior years, as Table 4-4 indicates. Final evaluations also declined somewhat as did ex-post evaluations. The category in which the FY89 and FY90 review showed an increase was an "other evaluations" category, which includes reviews of "lessons learned".

Table 4-4. Percentage of Evaluations by Evaluation Type Across Several Evaluation Reviews

Evaluation Reviews	Interim Evaluations	Final Evaluations	Ex-Post Evaluations	Other Evaluations
FY 85-86 Review	60%	30%	7%	3%
FY 87-88 Review	69%	29%	-	2%
FY 89-90 Review	59%	25%	2%	14%

Table 4-5 shows the distribution of different types of evaluations by bureau.¹³ Of the final and ex-post evaluations, which together account for 28% of the database, more were focused on projects in the Latin America and the Caribbean Bureau than was the case for other bureaus. The share of interim and final evaluations in the FHA Bureau was also high, but the total number of evaluations focusing on activities within this bureau was small. Also of note is the fact that over half of the large multi-project evaluations that MSI analyzed as a separate group fall into the reviews of "lessons learned" category.

C. The Scope of A.I.D. Evaluations

The scope of an evaluation, for purposes of this report, refers to the number of projects examined by an evaluation and their concentration in a single country or distribution across several countries. The vast majority of FY89 and FY90 evaluations were found to limit their scope to a single A.I.D. project. Of the evaluations in the database, 84% fell into this category, as Table 4-6 illustrates. This finding is similar to that of the FY87-88 review, in which 89% of the evaluations dealt with a single project. Evaluations in the FY89 and FY90 database which focused on multiple projects were found to be evenly divided between those undertaken within a single country and those in which projects in several countries were examined.

Table 4-7 presents the distribution of evaluations in the database by their type and scope. A total of 143 evaluations, or 53%, were interim evaluations that dealt with progress in a single A.I.D. project. Of these interim evaluations, 101 (70%) can be characterized as examining projects that are "mission-owned", i.e., financed and managed by bilateral missions overseas rather than regional offices or Washington bureaus.

D. The Purposes of A.I.D. Evaluations

As this review proceeded, it became clear that the term "evaluation purpose" has two meanings, both of which warranted review:

- The first way in which an evaluation communicated its purpose was through a formal statement of its intentions, e.g., its intention to examine project performance, or impact, or both.
- The second way in which evaluation purposes were revealed, albeit implicitly, was in an evaluation's conclusions and recommendations. While some evaluations only commented on the project that had been evaluated, others derived implications and lessons at the sectoral level or on a multi-sectoral basis.

¹³ In this table, as in many others in this report, the 16 large multi-project evaluations, listed in Table 2-1, for which no financial data was collected, are displayed on a separate line. As previously noted, this approach helps in identifying the characteristics of the most ambitious of A.I.D.'s evaluations, while it slightly understates the share of evaluations dealing with projects financed by the Latin America and Caribbean Bureau (LAC). The slight distortion caused by this choice of data presentation was viewed as being worth the trade-off in terms of information potentially gained concerning A.I.D.'s largest and most complex evaluations.

Table 4-5. Types of Evaluation by Bureau

Bureau	Interim Evaluations		Final and Ex-Post Evaluations		"Lessons Learned" Syntheses		Other Evaluations/ Can't Tell		Total Number of Bureau Evaluations
	Number (159)	Percent of Bureau Total	Number (74)	Percent of Bureau Total	Number (13)	Percent of Bureau Total	Number (22)	Percent of Bureau Total	
Asia	39	76%	9	18%	1	2%	2	4%	51
Near East	6	35%	5	29%	1	7%	5	29%	17
Latin American and the Caribbean (LAC)	34	50%	27	40%	1	1%	6	9%	68
Africa	46	69%	16	24%	0	-	5	7%	67
R & D	26	67%	11	28%	0	-	2	5%	39
FHA	2	40%	2	40%	0	-	1	20%	5
PRE	2	40%	1	20%	1	20%	1	20%	5
Large Multi-Project Evaluations	4	25%	3	19%	9	56%	0	-	165

Table 4-6. Scope of Evaluations Included in the Review

Evaluation Scope	Number	Percent
Single Project	224	84%
Multiple Projects in a Single Country	19	7%
Multiple Projects in Several Countries	19	7%
Other	6	2%
TOTAL	268	100%

In the paragraphs below, the findings of the evaluation review with respect these two different perspectives on evaluation purposes are reviewed.

1. The Stated Objectives of Evaluations

While a good deal can be inferred about an evaluation's objectives from its timing and scope, A.I.D. also requires that each evaluation include a clear statement of its purpose.¹⁴ Among the FY89 and FY90 evaluations, MSI found that evaluation purposes were identified for virtually all (99%) evaluations. Differences between bureaus with regard to the inclusion of a clear statement of purpose were minor as were differences by type of evaluation.

In the evaluation literature as well as in A.I.D. documents dealing with evaluation, it is frequently suggested that interim evaluations limit their investigation to management and implementation issues, while final and ex-post evaluations move beyond these issues to examine questions of impact and attribution. In order to examine these propositions empirically, the review team clustered detailed statements of evaluation purposes into three primary groups and coded the FY89 and FY90 evaluations as belonging to only one group, i.e.:

- A set of evaluations that stated their intent to examine only management and implementation issues.
- A set of evaluations that cited an examination of management and implementation issues as well as other purposes, including, for example, an assessment of the prospects for replicating a project or program in other countries; and
- A set of evaluations that cited purposes other than, or even excluded an examination of management and implementation issues.

¹⁴ A.I.D. Evaluation Handbook, op. cit., Section 3.5 and Appendix B.

Table 4-7. Scope Of Evaluations by Evaluation Type

Type of Evaluation	Evaluation of a Single Project		Evaluation of Several Projects in One Country		Evaluation of Several Countries		Other Evaluations		Total by Evaluation Type
	Number (224)	Percent of Type	Number (19)	Percent of Type	Number (19)	Percent of Type	Number (6)	Percent of Type	
Interim Evaluation	143	90%	5	3%	7	4%	4	3%	159
Final Evaluation	59	87%	6	9%	3	4%	0	-	68
Ex-Post Evaluation	3	50%	1	17%	2	33%	0	-	6
"Lessons Learned" Syntheses	4	31%	3	23%	5	38%	1	8%	13
Other Evaluations/ Can't Tell	15	68%	4	18%	2	9%	1	5%	22

Of the 268 evaluations in the database, 221 (82%) stated that an examination of management and implementation issues was one purpose of the evaluation. However, these were normally not the only objectives on which evaluations focused. In fact, only 45 (17%) of the 221 that cited a management and implementation purpose, 45 (17%) indicated that this was the only purpose of the evaluation. Table 4-8 illustrates this point. It also indicates when other purposes, or combinations of purposes were identified. Table 4-9 identifies illustrative purposes, beyond an examination of management and implementation issues, for which evaluations were reportedly undertaken. Some evaluations cited more than one of these purposes.

Table 4-8. Purpose for which Evaluations Were Carried Out

Purposes for which Evaluations Were Carried Out	Number of Evaluations	Percent
Only Management/Implementation Purposes	45	17%
Management/Implementation Purposes and Other Purposes	176	65%
Other Purposes, without regard to Management/Implementation Concerns	45	17%
None of the Above	2	1%
All Evaluations	268	100%

Table 4-9. Frequency with which Illustrative Purposes Other than Management/Implementation Reasons Were Cited as at Least One Reason for Conducting Evaluations

Appropriateness of the Project Design	Number of Times Reason was Cited	Percent of Evaluation Citing this Reason
Decide whether to continue or terminate a project	34	13%
Assess overall Attainment of Project Purposes and Goals	18	7%
Determine Project Effectiveness in Achieving its Outputs and Purpose	28	10%
Redesign the Project	24	9%
Facilitate design of Follow-on Project	55	21%
Provide Input for the Design of Similar or Related Projects	44	16%
Assess Prospects for Replication Elsewhere	16	6%

With respect to the evaluation purposes claimed by different types of evaluations, data from the evaluation review suggest that the conventional wisdom on this topic is not completely accurate. Among the 159 interim evaluations A.I.D. undertook, 33 (21%) were scored as having only management and implementation purposes. The majority of A.I.D.'s interim evaluations, 111 (70%) out of 159 cited purposes in addition to a review of management and implementation issues, and 14 (9%) interim evaluations did not even include an examination of management and implementation issues.

Findings for final and ex-post evaluations also ran somewhat counter to conventional wisdom. Of the 68 final evaluations in the database, 57 (84%) included, rather than excluded, an examination of management and implementation issues among their purposes, as did 3 (50%) of the 6 ex-post evaluations in the database. Even reviews of "lessons learned" occasionally included an examination of management and implementation issues. Management and implementation issues were reportedly considered in 5 (63%) of the 13 evaluations in this category.

2. Implicit Objectives of Evaluations

All evaluations in A.I.D. are expected to formulate recommendations and to draw out the lessons that have been learned through an evaluation. Evaluations that reach beyond the projects they examine to comment upon future projects, sectoral issues or multi-sectoral matters are potentially of use to a variety of audiences within A.I.D. and in host countries. From a management information perspective, they are different from evaluations that comment only on the projects they examined.

The evaluation literature suggests (perhaps inaccurately) that interim evaluations of individual projects have little of relevance to say to anyone outside of the immediate project context. MSI's evaluation review sought to test the validity of this proposition, and found that indeed, irrespective of their stated objectives, the majority of A.I.D.'s evaluations reach no further in their implications than the immediate project or program they examined. At most, they claim that their conclusions and recommendations can be applied to other projects of a similar nature.

As Table 4-10 indicates, 112 (42%) of the 268 evaluations in the database limited their conclusions and recommendations to the projects they had examined. In 93 (35%) of the evaluations, evaluation teams also drew explicit lessons or conclusions from the projects they evaluated for projects which were similar to those they had examined, while 61 (23%) of the FY89 and FY90 evaluations reached beyond the project level to comment on broader issues. Among this latter group, 50 evaluations noted implications at the sectoral level while 11 reached conclusions that had implications for multiple sectors. On a bureau basis, Table 4-11 shows the frequency with which evaluations reached beyond the project level in their conclusions and recommendations.

Table 4-10. Evaluation Implications Beyond the Project Level

Levels of Conclusions	Number	Percent
Only the Project that was evaluated	112	42%
Only the project that was evaluated and similar projects	93	35%
Sectoral level issues as well as the project or program	50	18%
Multi-sectoral issues as well as the project or program	11	4%
Evaluation did not provide clear conclusions or recommendation	2	1%
All Evaluations	268	100%

While the overall percentage of evaluations that offered A.I.D. managers conclusions and recommendations reaching beyond the project level is somewhat low, a detailed examination of these 61 projects provides some useful insights about the kinds of evaluations that yield this type of information.

Table 4-11. Distribution of Evaluations with Implications Beyond the Project Level

Bureau	Sectoral-Level Conclusions		Multi-Sector Conclusions	
	Number	Percent	Number	Percent
Asia	13	26%	3	27%
Near East	2	4%	-	-
Latin American and the Caribbean (LAC)	15	30%	2	18%
Africa	12	24%	1	9%
R&D	3	6%	-	-
FHA	-	-	-	-
PRE	-	-	-	-
Large Multi-Project Evaluations	5	10%	5	46%
All Evaluations	50	100%	11	100%

Of the 50 evaluations that reached conclusions at the sectoral level, 40 (80%) were evaluations of single projects, as were 4 (36%) of the evaluations that reached multi-sectoral conclusions. None of the 19 evaluations of multiple projects across several countries provided multi-sectoral conclusions and only 1 of these evaluations reached conclusions at the sectoral level.

With respect to evaluation types, 27 (54%) of the 50 evaluations that reached sectoral conclusions were interim evaluations, and 22 of these evaluations were of "mission owned" bilateral projects rather than projects funded and managed by regional offices or Washington bureaus. In addition, 5 (46%) of the 11 evaluations that reached multi-sectoral conclusions were interim evaluations, and four of these were "mission-owned" bilateral projects. Two (18%) of the final evaluations in the database reached multi-sectoral conclusions, while none of 6 ex-post evaluations in the database did so. Reviews of "lessons learned" reached multi-sectoral conclusions in 3 out of 13 evaluations, and sectoral conclusions in 2 evaluations.

CHAPTER FIVE

TEAM COMPOSITION AND QUALIFICATIONS

The quality and ultimate utility of evaluations rest heavily on the evaluation teams A.I.D. selects. This chapter describes the composition of A.I.D.'s evaluation teams. It also examines the skills team members bring to their task and the way in which technical knowledge and evaluation skills are blended on teams.

A. The Size of A.I.D. Evaluation Teams

The majority of A.I.D.'s FY89 and FY90 evaluations included in this review were carried out by teams of two to four people. Table 5-1 shows the distribution of A.I.D.'s evaluation teams by their size. A small number of evaluations were found to have very large evaluation teams, with two of these having 17 team members each. At the other end of the spectrum, 28 (10%) of the evaluations were found to have been carried out by a single individual.

Table 5-1. Distribution of Evaluation Teams by Size

Number of People on Team	Number of Evaluation Teams of This Size	Percent of Teams in this Size Category
1	28	10%
2	50	19%
3	49	18%
4	54	20%
5	28	11%
6 or 7	29	11%
8 or more	8	3%
Can't Tell	22	8%
TOTAL	268	100%

B. Team Structure and Composition

This section examines the degree to which various U.S. and host country organizations, including A.I.D., contribute team members to evaluations undertaken by A.I.D. The presence of A.I.D. staff members, and personnel from host country ministries on evaluation teams opens the question of the degree to which A.I.D.'s evaluations are "external", i.e., carried out by individuals

who have not been involved in a project's design or implementation, versus "internal", i.e., carried out by teams which include members of the project's design or implementation team. This section takes up both of these issues.

1. The Organizational Composition of A.I.D. Evaluation Teams

This review separately examined the U.S. and host country composition of A.I.D. evaluation teams. On the U.S. side, contractors who were associated with U.S. firms were found to be involved in more evaluations than were representatives from any other U.S. entity. The next most frequent member of evaluation teams, from an institutional perspective, was A.I.D. itself. These levels are consistent with the findings of the FY 87-88 evaluation review. Table 5-2 illustrates the frequency with which various U.S. entities participated in FY89-90 evaluations. This table indicates when particular groups worked alone on an evaluation and when they worked as part of an evaluation team that included representatives from several U.S. organizations.

On a bureau basis, A.I.D. staff served as team members on 45% of the evaluations of projects in the Africa Bureau. This percentage is higher than the 29% rate of A.I.D. staff participation on teams for evaluations of projects in the Near East Bureau and the 13% rate of A.I.D. staff participation on teams for evaluations of LAC Bureau projects, as Table 5-3 points out.

Table 5-2. Nature of U.S. Participation on Evaluation Teams

U.S. Participation on Evaluation Teams	Number	Percent
A.I.D. Staff Only	21	8%
A.I.D. Staff with Others	50	18%
Subtotal for A.I.D. participation on team	(71)	(26%)
U.S. Contractors (Firms) Only	153	57%
U.S. Personal Services Contractors (PSCs) only	10	4%
U.S. Universities only	5	2%
Combinations of Non-A.I.D. Evaluators	16	6%
Subtotal for no A.I.D. participation on team	(184)	(69%)
U.S. Aspects of Team Composition Could not be Determined	13	5%
TOTAL	268	100%

Table 5-3. A.I.D. Participation by Bureau.

Bureau	Total Number of Bureau Evaluations	Evaluations in Which A.I.D. Staff Participated as Team Members	Percent of Evaluations That Included A.I.D. Team Members
Asia	51	13	25%
Near East	17	5	29%
Latin America and the Caribbean (LAC)	68	9	13%
Africa	67	30	45%
R & D	39	6	15%
FHA	5	0	--
PRE	5	0	--
Multi-project Evaluations	16	8	50%
TOTAL	268	71	26%

There was no host country involvement on the evaluation team in 180 (67%) of the 268 evaluations. Table 5-4 shows the frequency with which personnel from the host country ministry that sponsored a project served as evaluation team members. It also shows the frequency with which other host country representatives served on evaluation teams.

Table 5-4. Host Country Participation on Evaluation Teams

Host Country Participation on Evaluation Teams	Number of Evaluations	Percent of Evaluation
Personnel from the Sponsoring Ministry	29	11%
Personnel from Other Clearly Defined Host Country Institutions, e.g. Other Ministries, Firms, etc.	35	13%
Other Host Country Personnel, Institutional Affiliation Unclear	24	9%
Subtotal for Host Participation Teams	88	33%
No Host Country Involvement on Team	180	67%
Total	268	100%

Table 5-5 shows the frequency with which host country personnel participated in A.I.D. evaluations on a bureau basis. As this table indicates, host country representatives participated in evaluations as team members for projects in the Asia Bureau far more frequently than was the case for other bureaus. As Table 5-6 also indicates, the set of broad multi-project evaluations (analyzed as a separate group) included host country team members on a relatively frequent basis. However, the majority of all multi-project evaluations on which host country personnel participated were those multi-project evaluations, whether large or small, that were undertaken in a single country. Only one instance was noted where a host country team member participated in a multi-project evaluation that was carried out in several countries. In terms of the types of evaluations in which host country personnel participated, Table 5-6 indicates that 56% of the evaluation teams on which host country personnel served were teams for interim evaluations.

Table 5-5. Frequency with Which Bureau Evaluations Include Host Country Personnel as Team Members

Bureau	Host Participation as Evaluation Team Member	Total Bureau Evaluations	Percent of Bureau Evaluation in Which Host Personnel Served on Teams
Asia	25	51	49%
Near East	6	17	35%
Latin America and the Caribbean (LAC)	20	68	29%
Africa	25	67	37%
R & D	2	39	5%
FHA	2	5	40%
PRE	0	5	-
Multi-project Evaluations	8	16	50%
TOTAL	88	268	26%

Of interest also is the fact that there was a good deal of overlap between A.I.D. and host country participation on evaluation teams. Of the 71 evaluations in which A.I.D. staff participated, 32 (45%) also included host country team members. Stated in a slightly different way, A.I.D. staff served as team members in 32 (36%) of the 88 evaluations in which host country nationals were reported to have been on the evaluation team.

2. The Frequency With A.I.D. Evaluations are "Internal" Versus "External" in Character

In the A.I.D. Evaluation Handbook the advantages and drawbacks of "internal" evaluations are carefully outlined and A.I.D.'s policy on this issue is set forth. Simply put, A.I.D. encourages "internal" evaluations, or the participation of members of a project's implementation staff, on interim evaluations. A.I.D. specifically discourages this type of involvement on final and ex-post evaluations:

Table 5-6. Host Country Participation as Evaluation Team Members

Host Country Personnel on Teams	Mid-Term Evaluations	Final or Ex Post Evaluations	Other Evaluations	Total Number of Teams on Which Host Personnel Participated	Percent of Evaluation Teams on Which Host Country Personnel Participated
Sponsoring ministry personnel participated as team members (whether or not other host country personnel were on the evaluation team.)	18	10	1	29	11%
Other host country personnel with clear organizational affiliation participated as team members (There were no sponsoring ministry personnel on these teams.)	18	9	8	35	13%
Host country personnel participated on evaluation teams but their organizational affiliations were not clear	13	8	3	24	9%
Subtotal for Host Country Participation	(49)	(27)	(12)	(88)	(33%)
Evaluation team had no host country members	110	47	23	180	67%
TOTAL	159	74	35	268	100%

"To avoid conflicts of interest, final or ex-post evaluation teams must be composed entirely of individuals with no previous connection (from initial design through implementation) with the activity being evaluated. This includes both U.S. and host country personnel. (Nevertheless), including A.I.D. direct-hire staff on evaluation teams who are not associated with the project...is encouraged wherever possible. Their participation serves as a direct link to Agency operations, expediting the transfer of experience and lessons learned from the evaluation." ¹⁵

¹⁵ A.I.D. Evaluation Handbook, op. cit., Section 3.5.6.

Using this guidance, the review team coded evaluations as being "internal" or "external" in character. The results of this coding suggest that, overall, 66 (25%) of the evaluations included in the database were "internal" in character, while 193 (72%) were "external". Nine evaluations could not be scored in this regard. These proportions are quite similar to those found in A.I.D.'s FY85-86 evaluation review. This earlier review reported that 21% of A.I.D.'s evaluations were "internal" while 77% were "external" in character.

If A.I.D.'s guidance concerning the involvement of its own staff and the staff of host country ministries were strictly followed, one would expect to find no cases where a final or ex-post evaluation was coded as being an "internal" evaluation. Conversely, it would also be reasonable to expect that a large number of interim evaluations would be scored as being "internal". That is not, however, what the MSI evaluation review found. Of the 159 interim evaluations in the database, 41 (26%) were scored as being "internal" in character. This is a low frequency compared to what A.I.D.'s evaluation guidance suggests. On the other hand, the review team found that 18 (26%) of A.I.D.'s 68 final evaluations were "internal", as were 2 of the 6 ex-post evaluations in the database.¹⁶

The issue of A.I.D. and host country participation in A.I.D. evaluations is taken up again, from a utilization perspective as well as from an institutional development perspective, in Chapter Six, which focuses on the degree to which A.I.D.'s evaluations have a participatory style.

C. The Skills Provided By Evaluation Team Members

The evaluation process requires not only general knowledge of a technical area, but also the skills that are needed to gather evidence concerning changes which are often difficult to measure in developing country environments. The need for a variety of skills on an evaluation team makes it difficult for one-person evaluation teams to provide results equal, in quality terms, to the evaluation products produced by somewhat larger teams.

Irrespective of whether evaluations use informal or survey research approaches for gathering data, teams generally need to have at least one member who understands the strengths and weaknesses of alternative information gathering approaches and the validity and reliability of the evidence which various approaches yield. For that reason, A.I.D.'s Evaluation Handbook requires that relevant evaluation skills be present on its teams:

"In general, an evaluation team requires technical specialists as well as at least one evaluation specialist...A social scientist with field research experience or a management specialist with development project experience can often serve as the evaluation specialist."¹⁷

¹⁶ A.I.D. participation on a team was scored as being "internal." The same was true for A.I.D.'s FY85-86 evaluation review.

¹⁷ Ibid.

While the A.I.D. Evaluation Handbook requires that such skills be present on teams, it does not currently require that evaluation reports describe the skills which team members bring to an evaluation. In the absence of a clear requirement to identify team skills, it was not surprising to find evaluations that failed to provide this type of information. The fraction of A.I.D. evaluations which failed to specify team member skills was, quite high. Of the 268 evaluations, 87 (33%) contained no information concerning the skills of evaluation team members. On a bureau basis, evaluations of projects in the Near East Bureau were notably deficient in this regard, with over 40% lacking information on the skill composition of evaluation teams.

As to the skills provided by team members in evaluations that reported on this matter, Table 5-7 shows the frequency with which various disciplines and skills were available on evaluation teams. Sectoral disciplines, e.g., health and agriculture, are represented on many evaluation teams as are other technical skills such as economics. With respect to skills listed in A.I.D.'s guidance on the presence of evaluation skills, the data suggest that this requirement is more often filled by individuals with a background in social sciences or management than it is by individuals who are described as having a specialized knowledge of evaluation.¹⁸

Table 5-7. Distribution of Skills Across 181 Evaluation Teams Where Team Skills Were Identified

Disciplines Represented	Number of Times Each Discipline was Present on 181 Evaluations Teams	Percent of Instances Where Disciplines were Stated
Economics	62	16%
Accounting/Finance	19	5%
Sectoral Disciplines	130	35%
Social Sciences	47	12%
Statistics/Mathematics	6	2%
Business/Trade	16	4%
Management/Administration	46	12%
Evaluation	19	5%
Development Disciplines	7	2%
Other Disciplines	27	7%
Total Occurrences	379	100%

¹⁸ Unlike the other disciplines listed in Table 5-5, academic degrees in evaluation are not common. Individuals who concentrated on evaluation in an academic environment would most likely be listed as having a degree which falls within the general province of the social sciences.

Clustering the three skill areas that correspond to A.I.D.'s broad definition of "evaluation disciplines", i.e., evaluation, management or social sciences, the review team found that 96 (53%) of the 181 evaluation teams for which information on team skills was available met A.I.D.'s requirement concerning the presence of an evaluation specialist on each team. There were few differences between bureaus or by evaluation type or scope with regard to the presence of an evaluation specialist on teams.

D. The Blending of Skills on A.I.D. Evaluation Teams

In addition to valuing the presence of relevant skills on an evaluation team, the literature on evaluation suggests that there is a benefit to having a mix of skills on such teams. In order to assess the degree to which A.I.D. evaluation teams incorporate a mix of skills, the review team created a composite variable that was used to code evaluations as having one, two, three or four or more different skills present on an evaluation team. Table 5-8 presents findings with respect to team complexity.

Table 5-8. Complexity of Evaluation Teams

Number of Disciplines/Skills Present on an Evaluation Team	Number of Evaluations at this Skill Complexity Level	Percent of Evaluations at this Level
One Discipline/Skill	55	21%
Two Disciplines/Skills	71	26%
Three Disciplines/Skills	38	14%
Four or more Disciplines/Skills	17	6%
Can't Tell	87	33%
TOTAL	268	100%

Among those evaluations where the skill composition of teams was known, 70% used teams with a blend of skills. The remaining 30% used teams that drew upon only one discipline or skill. The number of evaluations for which only one skill was listed was about double the number in which a single individual carried out the evaluation. On some teams, including a few relatively large teams, there were several individuals all of whom had the same basic skill. Team size and skill diversity are not necessarily related in A.I.D. evaluations.

In order to determine whether teams that offered evidence of skill diversity were more likely to include evaluation specialists, using A.I.D.'s broad definition of that term, the review team compared the number of skills on a team to its measure of whether evaluation skills are present on teams. Figure 5-1 illustrates this relationship for the 96 (36%) A.I.D. evaluations that included at least one individual with evaluation skills. As the figure suggests, the more skills that are brought to bear on an evaluation, the more likely it is that evaluation skills will be part of the mix. Evaluation teams consisting of four or more people almost always included at least one team member who met A.I.D.'s requirement concerning the presence of evaluation skills on teams.

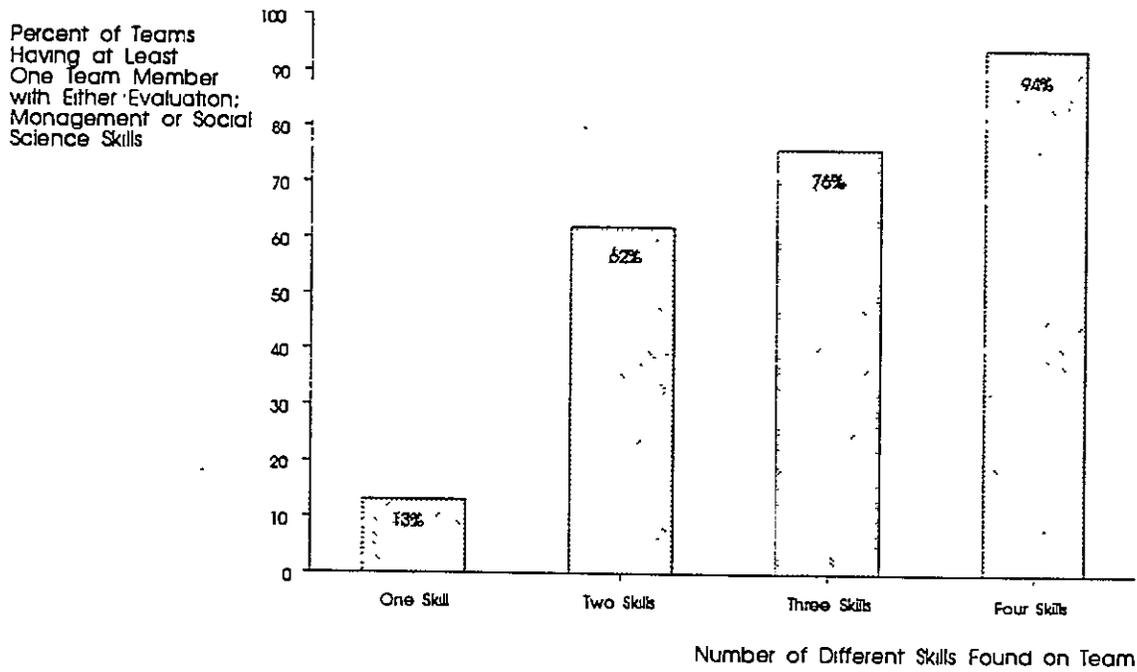


Figure 5-1. The Relationship Between Skill Complexity and Adherence to A.I.D.'S Evaluation Skills Requirement

E. Gender Considerations and Team Composition

While there are no A.I.D. evaluation guidelines in this regard, A.I.D. generally encourages its staff to take such steps as are necessary to ensure that project designs as well as evaluations are carried out in a manner that is sensitive to gender considerations. In practice, this involves the collection and analysis of data on the ways in which projects may provide different degrees of access to services or have differential effects on men and women.

One approach for ensuring that gender is considered as a factor as an evaluation is carried out is to incorporate this concern into an evaluation scope of work. Another involves having one member of an evaluation team take special responsibility for ensuring that gender issues are examined. A third approach is to send out evaluation teams which have both male and female team members.

Of the 221 (82%) evaluations in which the gender of team members could be determined, 105 (48%) had at least one female team member. Of these evaluations, 71 had one woman on the evaluation team, while the remaining 34 teams included more than one woman. As to the gender of the evaluation team leader, 187 evaluation teams had male team leaders and 32 had female team leaders. In addition, in 9 (3%) of the 268 evaluations one evaluation team member had been given a special responsibility for ensuring that gender issues were considered in the course of an evaluation.

CHAPTER SIX

EVALUATION STYLE: A.I.D. AND HOST COUNTRY PARTICIPATION IN THE EVALUATION PROCESS

From a management information perspective, the value of an evaluation lies in the degree to which its findings and recommendations are understood and utilized by the immediate and secondary audiences for such reports. Because the value of an evaluation is so closely linked to utilization, both the general literature on evaluation and A.I.D.'s own evaluation literature place a premium on audience identification and on the direct participation of A.I.D. and sponsoring host ministry staff in some types of evaluations.

The A.I.D. Evaluation Handbook discusses a variety of ways in which the primary audience for an evaluation can participate in the evaluation process, thus heightening the probability that this audience will utilize the evaluation's results. In summary, the participatory steps that A.I.D. views as facilitating the eventual utilization of evaluation results include:

- Participation by A.I.D. and host ministry personnel in the preparation of evaluation **scopes of work**, including the clear **identification of an evaluation's audience**.
- Encouraging and facilitating **donor and sponsoring host ministry participation on interim evaluation teams**, as discussed in Chapter Five.
- Interaction with an evaluation team at all stages of the evaluation process, e.g., using such approaches as **User Advisory Panels**.
- Organizing pre-evaluation **Team Planning Meetings** that bring an evaluation team and the evaluation's sponsor and primary audience together to discuss the evaluation's objectives.
- A.I.D. and host country involvement in **oral debriefings** in which evaluation teams present their findings and recommendations and **participation in the review of written drafts** of evaluation reports, and
- Conducting a formal **A.I.D. Evaluation Review**, in which the sponsor and primary audience for an evaluation, including senior managers, examine an evaluation's findings and make decisions based on its recommendations.

The following paragraphs present the findings of this review with respect to the various audience identification and participation measures outlined above.

A. Participation in the Development of Scopes of Work and Audience Identification

While the A.I.D. Evaluation Handbook makes it clear that both donor and host ministry staff need to participate in the development of evaluation scopes of work, there is no way to tell from A.I.D. evaluation reports whether an evaluation's intended audience participated in the

development of a scope of work, even when scopes of work are included as evaluation report annexes. All that can be determined is whether Scopes of Work are included in such documents. This question was examined in greater detail in Chapter Three of this report.

Turning to the question of audience identification, the A.I.D. Evaluation Handbook notes that this identification is one of the tasks of an evaluation scope of work. Evaluation reports are expected to state what is known about the intended audience for an evaluation. In the case of interim evaluations of bilateral projects, for example, A.I.D. normally expects that the primary audiences will include the mission and its implementation contractor as well as the sponsoring host country ministry. At other times, the audience for an evaluation may be a regional bureau or the Agency as a whole.

Table 6-1 displays findings concerning the audiences identified in A.I.D.'s FY89 and FY90 evaluations. As the table indicates, 57% of these evaluations explicitly identified their audiences. While this percentage is low relative to the share of evaluations that included a scope of work, it is high compared to the findings of A.I.D.'s 1988 examination of a small number of evaluations as part of a study of the utilization of A.I.D. evaluation reports.¹⁹

Table 6-1. Frequency with Which Evaluations Identified Specific Audiences

Audiences for An Evaluation	Number of Evaluations	Percent
Only the Mission or AID/W Office that Funded the Project(s) Which were Evaluated	74	28%
The Implementation Team (Mission or Regional Bureau and the Sponsoring Ministry)	35	13%
Only a Sponsoring Ministry	2	1%
Only a Regional Bureau	11	4%
Only Project Beneficiaries	1	-
Only CDIE or the Office of the Administrator	5	2%
Audiences in Other Combinations	24	9%
Subtotal for Audience Identified	(152)	(57%)
Audience Not Identified	116	43%
Total	268	100%

¹⁹ Yin, Robert K., et. al. "Preliminary Study of the Utilization of A.I.D.'s Evaluation Reports." Washington, D.C., The Cosmos Corporation, 1988, p. 39. This report found that only 2 of 33 CDIE evaluations identified their audiences clearly.

Data from the FY89 and FY90 evaluation review suggest that some types of evaluations conformed with A.I.D.'s requirement concerning audience identification more frequently than others. Thus 54% of the interim evaluations and 59% of the evaluation of single projects -- which in some cases are the same evaluations -- identified their audiences, whereas 75% of all multi-project evaluations and reviews of "lessons learned" included an identification of their audiences.

On a bureau basis, evaluations of Asia and PRE Bureau projects were the most consistent with respect to the identification of evaluation audiences. Evaluations of Asia Bureau projects included clear descriptions of their audiences 73% of the time. In the PRE Bureau, for which only a small sample of evaluations was examined, audiences were identified 80% of the time. At the other end of the spectrum were evaluations of projects in the R&D Bureau, with 46% of these evaluations identifying their audiences.

B. A.I.D. and Host Ministry Participation on Evaluation Teams

As noted earlier, A.I.D.'s expectations concerning audience participation as evaluation team members is not the same for final and ex-post evaluations as it is for interim evaluations. A.I.D. actively encourages the participation of its staff and the staff of the counterpart host country ministry as team members in interim evaluation. It specifically discourages the participation on final and ex-post evaluation teams of any A.I.D. or host ministry staff member who has been closely involved in a project's design or implementation.

With respect to the participation of A.I.D. staff and the staff of sponsoring host ministries on interim evaluations, the A.I.D. Evaluation Handbook defines two objectives. The first objective, which applies equally to A.I.D. and host ministry staff, focuses on the utilization of evaluation results:

"Combining project staff with outside evaluators is encouraged for interim process evaluations...This adds to the perceived legitimacy of the evaluation and facilitates more rapid use of the findings and recommendations."²⁰

"Moreover, the findings of evaluation(s) will have more credibility for host country managers if they have had a direct role in carrying out these activities."²¹

The second purpose served by involving host country, and particularly host ministry staff, focuses on institutional development:

"The capability to collect and analyze useful data on a timely basis to guide decision-making is certainly a key component of such institution building. Therefore, A.I.D.'s

²⁰ A.I.D. Evaluation Handbook, op. cit., Section 3.5.6.

²¹ *Ibid.*, Section 2.5.

monitoring and evaluation activities provide an excellent opportunity for improving the capabilities of host country counterpart organizations to collect, analyze and use data."²²

With this guidance in mind, the review team first took a broad look at overall A.I.D. and host country participation on evaluation teams. The results of this review were presented in Chapter Five, which indicated that A.I.D. staff participated as team members in 26% of the 268 evaluations, while host country personnel participated in 33%.

The review team then examined in greater detail the participation of A.I.D. and host ministry staff in interim evaluations of "mission-owned" projects, i.e., those projects which are financed and managed by A.I.D.'s bilateral missions. The 106 evaluations in this subset represent instances in which the participation of both A.I.D. staff and host ministry staff as evaluation team members is clearly recommended in the A.I.D. Evaluation Handbook. Thus, the data for this subset of evaluations provides a good measure of whether A.I.D.'s guidance in this regard is being followed.

In its examination of the 106 evaluations in the subset of "mission-owned" interim evaluations, MSI found that A.I.D. staff had participated as evaluation team members in 32 (30%) of the evaluations in this subset, while 43 (41%) of these evaluations had host country participation on evaluation teams. A still smaller group of 15 (14%) of the evaluations in this subset included host country representatives from the sponsoring ministry on their evaluation teams.

Table 6-2 displays host country participation in the 106 evaluations included in the subset of "mission-owned" interim evaluations. As the table indicates, regional bureaus failed to include host country personnel on evaluation teams for interim evaluations of "mission-owned" projects more than 50% of the time. In the Near East Bureau, the share of interim evaluations which did not include host country personnel on the team was substantially higher.

²² Ibid.

Table 6-2. Host Country Participation in Interim Evaluations of "Mission-Owned" Projects

Bureau	Number of "Mission-Owned" Interim Evaluations	Sponsoring Ministry Participation or Evaluation Teams		Other Host Country Personnel on Evaluation Teams		No Host Country Personnel on Evaluation Teams	
		Number	Percent	Number	Percent	Number	Percent
Asia	38	3	7%	14	38%	21	55%
Near East	6	-	-	1	17%	5	83%
Latin America and the Caribbean (LAC)	28	2	7%	7	25%	19	68%
Africa	31	10	32%	4	13%	17	55%
Broad Multi-Project Evaluations	3	0	-	2	66%	1	34%
Total	106	15	14%	28	27%	63	59%

57

With respect to sponsoring ministry participation, the Africa Bureau was almost twice as likely as the Asia or LAC Bureaus to include sponsoring ministry personnel on interim evaluations of "mission-owned" projects. While neither the Asia or LAC Bureau included sponsoring ministry personnel as team members in a large percentage of their interim evaluations of "mission-owned" projects, both bureaus included other host country personnel on teams at least 25% of the time. In comparison, the review team found no instances in the Near East Bureau where sponsoring ministry personnel had been included as team members on interim evaluations. In only one of six interim evaluations of "mission-owned" projects in the Near East Bureau was there any host country team participation whatsoever.

Looking across the 15 interim evaluations of "mission-owned" projects in which sponsoring ministries provided as evaluation team members, the review team noted that A.I.D. staff also participated in 11 (73%) of these evaluations. Of the 15 evaluations in this cluster, 14 were found to be single-point-in-time, or "snapshot" evaluations, which used unstructured or impressionistic methods for collecting interview data. In addition, 12 of the 15 were found to include site visits.

As the foregoing suggests, A.I.D. and host country participation, particularly sponsoring ministry participation, on interim evaluation teams for "mission-owned" projects, occurs much less frequently than the A.I.D. Evaluation Handbook suggests is desirable. Conversely, in final evaluations of "mission-owned" projects, A.I.D. and sponsoring host ministry participation on teams was occurring, in spite of guidance discouraging such participation:

- While 62% of the evaluations of "mission-owned" projects in which A.I.D. participated as a team member were interim evaluations, 38% were final or ex-post evaluations, and
- In much the same manner, 60% of the evaluations in which host ministry personnel participated as team members were interim evaluations, while 40% were final or ex-post evaluations.

C. Users Advisory Panels

While the A.I.D. Evaluation Handbook does not specifically discuss such panels, it does recommend that the A.I.D. officer who is responsible for an evaluation maintain "periodic and open communication with the (evaluation) team." User advisory panels can function in this manner for evaluations that involve multiple audiences or complex technical issues.

Within the set of evaluations it reviewed, the review team found 12 cases in which user advisory panels had been integrated into an evaluation process. Of these 12 cases, 10 were mission-sponsored evaluations that examined a single project. Seven of the evaluations that used advisory panels did so in connection with interim evaluations while 4 were associated with final evaluations.

The small number of evaluations that used advisory panels suggests, on a prospective basis, the need for a different indicator to assess the degree to which the sponsor of an A.I.D. evaluation maintains an open line of communication with an evaluation team throughout the process.

D. Team Planning Meetings

Team Planning Meetings (TPMs), which bring together an evaluation team and its sponsor or audience at the start of the evaluation process, are believed to serve a number of useful functions. They are less cumbersome than user advisory panels, yet, from a utilization perspective, TPMs can effectively increase the interest or "stake" an evaluation's audience has in an evaluation's findings and recommendations. It does so by fostering sponsor and audience involvement in refining the questions an evaluation will examine and the means it will use to answer them. The A.I.D. Evaluation Handbook states that TPMs are "highly recommended for orienting the evaluation team." In this handbook, TPMs are described as facilitating the development of "a basic consensus among team members concerning the objectives of their assignments (which) will expedite work on the evaluation and contribute to producing a useful report."²³

Of the evaluations in the FY89-90 database, 52 (19%) stated that TPMs had been held. It is possible that a larger share of A.I.D.'s evaluations involved TPMs, but from evaluation reports alone, this could not be determined. Nor could the composition of the group that participated in a TPM be determined from evaluation reports. In some cases the evaluation sponsor and audiences may have been involved. In other cases TPMs may have only involved evaluation team members. MSI found no important difference by evaluation type, i.e., interim; final, etc., with respect to the use of TPMs, although their use did seem to relate positively to team size, with evaluations involving large team employing this participatory step more frequently than did those with smaller teams.

On a bureau basis, evaluations that focused on projects in Asia and in the Near East used TPMs more frequently than was the case in other bureaus, as Table 6-3 indicates. As this table also points out, broad multi-project evaluations, which the review treated as a separate category, included TPMs more frequently than did evaluations that examined fewer projects.

²³ Ibid, Section 3.6.2.

**Table 6-3. Evaluations for Which a Team Planning Meeting (TPM)
was Part of the Evaluation Process**

Bureaus	Evaluation for Which a TPM was Held	All Bureau Evaluations	Percent of Evaluations that Had a TPM
Asia	12	51	24%
Near East	4	17	24%
Latin America and the Caribbean (LAC)	13	68	19%
Africa	8	67	12%
R & D	7	39	18%
FHA	2	5	40%
PRE	0	5	--
Broad Multi-project Evaluations	6	16	38%
All Evaluations	52	268	19%

E. A.I.D. and Host Country Participation in Oral Debriefings and the Review of Written Drafts of Evaluation Reports

Oral briefings and the circulation of draft copies of evaluation reports are perhaps the most common methods used to ensure that the audience for an evaluation is aware of its findings, conclusions and recommendations and has an opportunity to interact with the evaluation team on these matters. These steps, while important from a utilization perspective, are not always documented in evaluation reports. Nor are such events always reported in A.I.D. Evaluation Summaries.

This review of A.I.D.'s FY89 and FY90 evaluations found evidence of A.I.D. and host country participation in oral briefings in 87 (32%) of the 268 evaluations in the database. Evidence of A.I.D. and host country involvement in the review of draft versions of evaluation reports was found for 94 (35%) of these evaluations. There was roughly an 85% overlap between these two activities, i.e., in most of the cases where A.I.D. and host country personnel had participated in an oral briefing they had also examined draft versions of evaluation reports. Data on the frequency with which various parties were reported to have participated in these activities are summarized in Table 6-5.

As to the types of evaluations that reported on A.I.D. and host country involvement in oral briefings or their review of draft reports, it appears that participation of this sort is more likely in interim evaluations and in evaluations that examine multiple projects in a single country than in other types of evaluations. On a bureau basis, there was slightly more reported audience involvement in these activities in evaluations that focused on projects in the Asia, Africa and LAC Bureaus than in evaluations which focused on projects in other bureaus.

Table 6-4. Evaluation Participants Involvement in Reviewing Evaluation Findings

Participants	Participated in an Oral Review of the Evaluation		Reviewed the Written Draft of the Evaluation Report	
	Number	Percent	Number	Percent
Only the Mission or AID/W Office that Funded the Project(s) Which Were Evaluated	26	10%	-	-
The Implementation Team (Mission or Regional Bureau and the Sponsoring Ministry)	44	16%	37	14%
Other Participants Alone or in Combinations	17	6%	57	21%
Subtotal for Participation	(87)	(32%)	(94)	(35%)
Can't Tell	181	68%	174	65%
All Evaluations	268	100%	268	100%

F. Formal Evaluation Reviews

Evaluation reviews in A.I.D. differ from oral briefings on evaluation findings in the following way. In an oral briefing an evaluation team describes its findings and recommendations to members of the evaluation's audience. An A.I.D. Evaluation Review has a broader scope. In an A.I.D. Evaluation Review, those who would be expected to implement evaluation recommendations are encouraged to examine their merits, and if action is warranted, to discuss, if not decide, how and when it will be taken.

Among the FY89 and FY90 evaluations in the database, 47 (18%) made reference to a formal evaluation review, as Table 6-7 indicates. The absence of information on formal evaluation reviews cannot, however, be interpreted to mean that evaluation reviews only occur 18% of the time. Information on formal evaluation reviews was most often found in A.I.D. Evaluation Summaries and, given that 49% of A.I.D.'s FY89 and FY90 evaluations included such summaries, the true frequency with which formal evaluation reviews occur is probably understated.

Table 6-5. Frequency with Which Evaluations Indicate that an Evaluation Review Meeting was Held by A.I.D.

Bureaus	Total of Bureau Evaluations	Number for Which an Evaluation Review was Known to Have Been Held	Percent for Which an Evaluation Review was Held
Asia	51	9	18%
Near East	17	2	12%
Latin America and the Caribbean (LAC)	68	11	16%
Africa	67	11	16%
R & D	39	13	33%
FHA	5	-	-
PRE	5	1	20%
Broad Multi-project Evaluations	16	-	-
All Evaluations	268	47	18%

G. A Composite View of A.I.D. and Host Country Participation

In order to obtain an overall picture of A.I.D. and host country "end-user" participation in A.I.D. evaluations, MSI developed a composite rating. This rating integrates several aspects of the A.I.D. guidance that encourages user participation in the evaluation process. It is thus biased in favor of, and should only be applied to, interim evaluations. Given the degree to which A.I.D.'s guidance concerning user participation in the evaluation process focuses on host country personnel, particularly the staff of ministries that are involved as co-sponsors in A.I.D.-financed project, the review team only applied its composite participation rating to those 106 projects that are interim in nature and are also "mission-owned" bilateral projects. The rating scale for participation that MSI developed has three levels: high, medium and low, as defined below.

High This rating was applied only in situations where both A.I.D. and the sponsoring ministry participated on the evaluation team and, in addition, each party either participated in oral briefings or reviewed a draft of the evaluation.

Medium Evaluations in this category include those in which A.I.D. either had a representative on the evaluation team or participated in oral briefings or reviewed a draft report and a representative of the sponsoring ministry participated at the same level, i.e., was involved in one of three possible participatory activities.

Low A rating of "low" on participation was given to evaluations where either A.I.D. or the host ministry, but not both, had a representative on the evaluation team. A code of "low" was also assigned if one, but not both, of these parties participated in oral briefings on the evaluation or reviewed a written draft.

Can't Tell Evaluations in this category did not clearly indicate that either A.I.D. staff or representatives of sponsoring ministries participated as team members. In addition evaluations provided no evidence suggesting that either of these parties had participated in oral briefings or reviewed written drafts of evaluations.

Of the 106 evaluations which were both interim in nature and "mission-owned," 6 (6%) were scored high on MSI's composite rating on participation. Another 35 evaluations (33%) scored medium on this composite rating while 29 (27%) scored low. Of the 106 evaluations in this subset, 36 (34%) did not provide enough information on participation to be scored.

On a bureau basis, evaluations of projects in the Africa Bureau received the greatest number of high scores for participation, i.e., 5 of the 6 evaluations receiving this score were evaluations of Africa Bureau projects. The missions involved in these highly participatory evaluations were Malawi, Mali, Kenya, Swaziland and Somalia, with one evaluation each. The one final evaluation that received a high score was a LAC Bureau evaluation in Jamaica. In all other regards, the evaluations which were found to be highly participatory were very much in the mainstream of A.I.D.'s current evaluations.

CHAPTER SEVEN

EVALUATION FRAMEWORKS AND METHODS

This chapter examines the conceptual frameworks that A.I.D.'s FY89 and FY90 evaluations used to form judgements about project and program performance. It also considers the methods the evaluations used to gather and analyze information.

A. The Basis Evaluations Use to Judge Performance

The determination that a project or program has been successful or unsuccessful requires a judgement. If the basis by which projects are judged is transparent, then the audience for an evaluation has a rational foundation against which to consider evaluation conclusions and recommendations. Where evaluations fail to share definitions of "success" that are applied in an evaluation, it is as if evaluators are asking their audience to trust their judgement, rather than share in a open and verifiable process.

The A.I.D. Evaluation Handbook encourages evaluation teams to make judgements about performance by comparing the information they collect concerning a project to notions about what that project was expected to accomplish. Project design documents, especially a project design tool called the Logical Framework, articulate the Agency's substantive expectations about project accomplishments. Project costs are a second frame of reference to which project performance can be compared. This section examines the degree to which A.I.D.'s FY89 and FY89 evaluations utilized either or both of these conceptual frameworks as a basis for making judgements about performance.

1. Use of the Logical Framework in A.I.D. Evaluations

A.I.D.'s evaluations system is linked to the approach it uses to design projects by a device called the Logical Framework. The Logical Framework, which A.I.D. has been using since the early 1970s, is a tool for organizing information about a project's hierarchy of objectives. It uses the following terminology to characterize levels of objectives and the relationships among their levels:

- Inputs (actions and resources),
- Outputs (immediate results),
- Purpose (reason for the project and primary outcome), and
- Goal (higher level, e.g., sectoral or national objective, to which the project contributes)

In a Logical Framework, indicators of performance and specific targets in terms of the quality, quantity and timing of results are specified. Since a Logical Framework identifies the

criterion a project design team established for assessing performance, it is often viewed as offering a "fair" basis for judging the adequacy of a project or program's performance.

The A.I.D. Evaluation Handbook makes numerous references to project Logical Frameworks. It includes a requirement for evaluation reports to include Logical Frameworks as an evaluation appendix where such frameworks exist. It also calls for interim evaluations "to review actual versus planned progress towards the outputs, purpose and goal of (a) project."²⁴

In order to gain a clearer understanding of whether, and to what degree, Logical Frameworks were used as the basis for A.I.D.'s current evaluations. The review team included questions on this issue in its rating form. Two elements of the rating form were used to determine whether and to what degree evaluation teams used the criteria established in a project's Logical Framework as the basis for assessing project performance:

- Evaluations were coded on whether they explicitly referred to a project or program's design, discussed a project's inputs, outputs, purpose and goal; and addressed their status in the course of the evaluation.
- Evaluations were also coded on the types of indicators they used to measure performance. Indicators taken from project Logical Frameworks were specifically noted on the rating form.

Of the 268 evaluations in the FY89 and FY90 database, 180 (67%) evaluations made explicit reference to using a project's design as the basis for assessing project performance. On average, 70% of the evaluations in regional bureaus and in the R&D Bureau indicated that the project design served as the conceptual basis for judging project performance. Only the Near East Bureau fell well below this average, reporting that project designs were used to assess progress in 47% of the bureau's evaluations.

Quite understandably, the use of project designs as a basis for assessing performance was found to be more likely in evaluations of single projects, on both an interim and final basis, than it was in multi-project evaluations. Of the 224 single project evaluations in the database, 163 (73%) reported that project designs were used as the basis for judging project performance, in contrast to 14 (37%) of the 38 multi-project evaluations in the database which made this claim. Interim and final evaluations both utilized project designs as the basis for judging project performance about 73% percent of the time, while ex-post evaluations used them somewhat less frequently.

With respect to the types of indicators used to measure project and program performance:

- 252 (94%) of the evaluations were coded as using indicators drawn from the project context as a basis for measuring performance, and

²⁴ A.I.D. Evaluation Handbook, Section 3.3.5.

- 139 (52%) of the evaluations explicitly stated that performance was being assessed using indicators taken from a project's Logical Framework.

The overlap between these two measures consisted of 117 evaluations scored as using indicators from the project context and, more specifically, using indicators from a project's Logical Framework. Given this overlap, it appears that 135, or about half of all evaluations that used project-related indicators to measure success, drew their indicators of performance out of the project context without reference to project Logical Frameworks.

On a bureau basis, indicators taken from Logical Frameworks were used to assess performance in over 50% of the evaluations in all regional bureaus except the Near-East. They were also used in slightly more than 50% of the evaluations of projects in the R&D Bureau, as Table 7-1 indicates.

The use of indicators drawn from project Logical Frameworks was found to be higher for interim evaluations, where they were used for 62% of the time, than it was for final and ex-post evaluations, which used indicators drawn from Logical Frameworks 50% and 33% of the time, respectively. Reviews of "lessons learned" were found to use indicators taken from Logical Frameworks 15% of the time. In contrast to a strong tendency to use indicators of performance drawn from a project's context, a small fraction of the evaluations, 31 (or 12%), appeared to rely totally indicators of performance that were not explicitly linked to the project context for drawing conclusions about project performance.

Table 7-1. Frequency with Which A.I.D. Evaluation Use Performance Indicators from Project Logical Frameworks as the Basis for Judging Success and Failure

Bureau	Total Number of Evaluations	Evaluations That Used Performance Indicators from Project Logical Frameworks	Percent of Bureau Evaluations
Asia	51	27	53%
Near East	17	4	24%
Latin America and the Caribbean (LAC)	68	39	58%
Africa	67	40	60%
R&D	39	20	51%
FHA	5	2	40%
PRE	5	2	40%
Broad Multi-project Evaluations	16	5	31%
Total	268	139	52%

2. The Comparison of Performance to Cost

In order to determine whether evaluations used cost as one of the bases for their assessment of performance, the review team noted when costs were considered in relation to project benefits. Its rating form also pursued the somewhat more complicated question of whether the costs and apparent benefits of a project were considered in relation to other options which had been or were currently available to A.I.D., i.e., was the project (given its benefits in relation to costs) a better investment than something else.

With respect to the first of these questions, MSI found that 55 (21%) of A.I.D.'s FY89 and FY90 evaluations examined project effects in relation to costs. This suggests that roughly 80% of all A.I.D. evaluations make no comment on whether the cost-effectiveness analyses undertaken at the point of a project's design still seemed valid. Still fewer evaluations examined costs and benefits in relation to some definable alternative; 29 (11%) of the 268 evaluations in the database were coded as having examined this second question.

Not surprisingly, final and ex-post evaluations were found to consider project effects in relation to cost about twice as often as interim evaluations, i.e., 32% of the time in the former case as compared to 16% for the latter. In addition, evaluations that examined projects whose funding level was under \$10 million were found to be somewhat more likely to have included a cost analysis as part of the evaluation than were evaluations examining projects that had been funded at much higher levels, i.e., over \$50 million. Roughly 30% of the evaluations that examined smaller projects included cost analyses, while 12% of the evaluations of very large project considered cost. The presence of economists on evaluation teams did not appear to be related to whether cost data was examined or not.

On a regional basis, MSI found that 29% of the evaluations of projects in the Near East Bureau and 25% of the evaluations of projects in the LAC Bureau considered cost factors. The frequency with which these two bureaus examined cost issues was higher than was found for other bureaus. PRE, a bureau for which there was only a small sample of evaluations, also considered costs in relation to performance on a fairly consistent basis.

B. Evaluation Designs and Methods

The approaches evaluations use to gather and analyze data have a clear and direct relationship to the credibility of their conclusions and the regard in which their recommendations are held. This is not to suggest that there is only one "right way" to gather and analyze information. That is not the case. There are in fact many legitimate approaches for gathering and analyzing information in evaluations. What is not credible or acceptable is an evaluation that asserts conclusions and makes recommendations without offering a comprehensible factual basis for such statements.

This section of the report examines the evaluation designs and information gathering methods which were used in A.I.D.'s FY89 and FY90 evaluations. At the end of the section a composite measure is used to identify those evaluations that were the strongest from a methodological standpoint.

1. Evaluation Designs Used in A.I.D. Evaluations

The design, or basic structure, of an evaluation conveys the choices an evaluation team makes concerning the degree of rigor that is to be applied in seeking answers to questions regarding changes brought about by project activities. In most evaluations, teams face questions about changes that have occurred since a project began. Their ability to answer those questions depends, in good part, about what is known about the pre-project situation, e.g., whether the kinds of goods and services a project provides were already available; general economic conditions, etc. Despite the fact that A.I.D. requires its staff to include baseline information on such conditions in its project design documents, the information acquired during the design process is often too general to be used to attribute or isolate casual relationships regarding the changes brought about by A.I.D. projects.

Evaluation designs that attempt to measure the specific changes that can be attributed to A.I.D. projects are relatively expensive. They tend to demand answers to exactly the same question, from almost exactly the same farmers or children and at least two points in time, i.e., before the project started and after project goods or services have been provided. Evaluation designs of this sort must be put in place at the time a project is funded. As other reviews of A.I.D. evaluations have already suggested, such designs are rarely used. More common are studies that gather data on an "after only" basis and attempt to use reason to deduce whether changes occurred and whether such changes are attributable to A.I.D.'s projects or to other factors in the environment.

As a practical matter, the factors that influence the choice of a basic evaluation design include the questions to be answered by an evaluation and the financial resources which can be devoted to obtaining those answers. During the active financial life of an A.I.D. project, questions about impact may not be answerable, obviating the need for at least a portion of A.I.D.'s interim evaluations to employ complex evaluation designs. Similarly, where projects introduce a service that was not previously available, but which is known to have specific effects, e.g., vaccinations, relatively simple evaluation designs are often adequate. Complex, multiple-point-in-time evaluation designs are most useful when A.I.D. undertakes projects where the likely outcomes are somewhat uncertain and where there are a number of other factors in a project situation that could plausibly bring about the very changes on which A.I.D.'s project is focused.

In reviewing A.I.D.'s FY89 and FY90 evaluations, the review team coded evaluations with respect to the basic evaluation design they applied. Experimental and quasi-experimental designs are those which, as described above, attempted to systematically acquire data at two points in time (e.g., before and during or after the period when goods and services are delivered to the project's intended beneficiaries). Single point in time, or "snap shot", designs, on the other hand, are those evaluation approaches that collect data at a single point during the life of a project, often only after it has provided the goods and services it was intended to deliver.

In addition to noting whether evaluations used single- or multiple-point-in-time approaches for measuring changes brought about by project activities, the review team also recorded when evaluations did neither. A number of reviews of "lessons learned," for example, did not collect

new data. Instead, they drew conclusions by comparing the findings of existing evaluation reports.

Table 7-2 presents MSI's findings concerning the distribution of A.I.D.'s FY89 and FY90 evaluations in terms of their basic design. As the table indicates, single-point-in-time or "snap shot" studies are by far the most common, accounting for 87% of all of the FY89 and FY90 evaluations. Single-point-in-time evaluations are even more common when only interim evaluations are considered. This type of design was used in 145 (91%) of the 159 interim evaluations in the database.

Table 7-2. Distribution of Evaluations According to Their Basic Approach or Design

Evaluation Design	Number	Percent
"Snapshot" of Project Performance (no before/baseline measures and no comparison/control groups)	234	87%
Experimental or Quasi-experimental (before and after measures, sometimes involving comparison/control groups)	21	8%
Analyses/syntheses based on the results of other studies and other designs	13	5%
All Evaluations	268	100%

Attempts to apply more complex and expensive experimental and quasi-experimental designs to obtain information for more than one point in time accounted for only 8% percent of all evaluations and were most often carried out in Latin America and Asia. Interestingly, 9 (43%) of the 21 evaluations that were coded as being experimental or quasi-experimental were interim evaluations rather than final or ex-post evaluations.

A.I.D. is already aware of the high frequency with which single-point-in-time evaluation designs are being used in evaluations. It's heavy reliance on single-point-in-time or "snap shot" evaluation designs reflects practical constraints. Recognizing this situation, A.I.D. has shifted some of its evaluation emphasis toward the identification of ways in which single-point-in-time studies can produce high-quality data. It has invested, for example, in the development of guidelines concerning the use of innovative evaluation methods, e.g., the use of group interviews and other approaches categorized as "rapid appraisal methods."

2. Sources of Information for A.I.D. Evaluations

Turning to the question of the quality of the information obtained by evaluations, irrespective of their design, the review team coded evaluations for all of the types of information

they used. Their rating forms used detailed questions to identify the patterns and categories of information sources used in evaluations. The first of these source clusters focused on pre-existing information, i.e., records and documents that an evaluation team could draw upon instead of, or in addition to, any new data it collected. Two categories of new data were also identified: (a) direct measures, e.g., birth weight, and (b) site visits or case studies through which interview or observational data could be obtained.

Most evaluation teams identified the sources of data they used to reach conclusions. All but 13 (4%) of the 268 evaluations in the database specified their sources of information. Of these evaluations, virtually all included pre-existing data as one of their sources. The pre-existing data category included several types of information, e.g., progress reports on projects, earlier evaluations, etc. These types of pre-existing information were cited far more frequently than were baseline data. Of the 268 evaluations in the database, 54 (20%) were coded as having baseline data.

Since virtually all evaluations utilized pre-existing data, the analytic question for the review team became whether evaluations used only pre-existing data or whether pre-existing information was used in combination with one or more types of new data. Table 7-3 provides the answer to that question. As the table indicates, the most frequent combination involved the use of pre-existing data and a site visit or case study. Site visits and case studies were also used in combination with direct measures and pre-existing data. When aggregated, the information in this table indicates that:

- A total of 198 (74%) of the evaluations A.I.D. undertook in FY89 and FY90 involved site visits or case studies in combination with pre-existing data and, at times, also in combination with direct measures.
- Another 102 (38%) of the evaluations used some type of direct measure of progress in combination with pre-existing data and, in some cases, in combination also with site visits or case studies.
- Of the 40 evaluations that used only pre-existing data, 37 (92%) used single-point-in-time, or "snapshot," evaluation designs and 31 (76%) were interim evaluations.

3. Types of Data Acquired Through A.I.D. Evaluations

In addition to considering the sources of evidence that evaluations used to develop their findings and conclusions and frame their recommendations, MSI also attempted to discern whether evaluations were gathering data that could be used to make valid statements about the changes that occur in or are brought about by A.I.D. projects. Also of interest in this regard was the degree to which evaluation teams acquired three specific types of data: (a) trend data; (b) data directly from beneficiaries; and (c) gender-disaggregated data. The review team also noted whether evaluations collected data on cross-cutting issues, i.e., sustainability, environmental impact and the gender-specific effects of projects and programs. Information on the collection of data on cross-cutting issues is presented in Chapter Eight, together with information on

evaluation findings in these areas. Information on the collection of trend, beneficiary and gender-disaggregated data in evaluations is discussed below.

Table 7-3. Frequency with Which Various Sources of Evidence Were Used by Evaluations

Types of Evidence Obtained & Used in Evaluations	Number	Percent
Primarily data that pre-existed the evaluation, e.g., progress reports from projects (minimal interviews/no site visits)	40	15%
Pre-existing data plus direct measured, e.g., physical evidence/records	17	6%
Pre-existing data plus site visits or case studies	113	43%
Pre-existing data, plus direct measures and case studies or site visits	85	32%
Sources of evidence not specified	13	4%
All Evaluations	268	100%

a. Trend Data

To the degree that A.I.D. relies, for practical reasons, on single-point-in-time evaluation designs, the use of trend data offers evaluators a means of widening their perspective on a project or program. At times, trend data can substitute for baseline data, which projects often lack.

The review team coded evaluations on whether they used trend data to assess changes that had occurred in or been brought about by A.I.D. projects. Some evaluations cited progress reports and earlier evaluations as trend information sources. Still other evaluations indicated that secondary data sources, e.g., government statistics, etc., had been used. All of these sources offered evaluation teams opportunities to assess changes over time. In particular, they added perspective and depth to evaluations that used a "snap shot" design.

With respect to the specific question of whether evaluations utilized some form of trend data, the review team coded 197 (74%) of the 268 evaluations in the database as using some type of trend data to formulate their findings and conclusions. The use of trend data was, as might be expected, more frequent in final and ex-post evaluations than it was in interim evaluations. Of the 74 final and ex-post evaluations in the database, 63 (85%) reported using trend data, as compared to 108 (68%) of the 159 interim evaluations in the database.

b. Data from Project Beneficiaries

The credibility of an evaluation's conclusions about impact tends to be strengthened when evaluation teams gather information from a project's beneficiary population about a project's outputs as well as its impact. Information from beneficiaries is often just as useful during interim evaluations as it is in final and ex-post evaluations.

Table 7-4 shows, on a bureau basis, the frequency with which A.I.D. evaluations collected data directly from project and program beneficiaries. Of the 286 evaluations in the database, 158 (59%) acquired data directly from project or program beneficiaries. A total of 131 (83%) of the 158 evaluations that gathered beneficiary data examined projects that delivered services directly to people, i.e., they were primarily service projects rather than institution-building or policy-reform endeavors.

Table 7-4. Frequency with Which Evaluations Gathered Data Directly from Project or Program Beneficiaries

Bureau	Number of Evaluations	Total Number of Bureau Evaluations	Percent
Asia	32	51	63%
Near East	11	17	65%
Latin America and the Caribbean (LAC)	50	68	75%
Africa	36	67	54%
R & D	12	39	31%
FHA	4	5	80%
PRE	4	5	80%
Broad Multi-project Evaluations	9	16	56%
All Evaluations	158	268	59%

On a bureau basis, evaluations of projects in the Africa Bureau were found to collect beneficiary data less frequently than was the case for other regional bureaus. With respect to types of evaluations, beneficiary data were obtained for 83 (52%) of the 159 interim evaluations in the database. Final evaluations and ex-post evaluations did somewhat better in this regard, including beneficiary data 69% and 83% of the time, respectively.

While the percentage of evaluations that included beneficiary data did not vary dramatically as a function of evaluation design, the review team found that 84% of the evaluations used only structured interview approaches gathered beneficiary data, whereas only 54% of the evaluations that used only impressionistic methods gathered data from beneficiaries. Of those evaluations that used a blend of both interview techniques, 75% gathered beneficiary data.

c. Gender-Disaggregated Data

As A.I.D.'s general guidance makes clear, the collection of information on a gender-disaggregated basis during project design, and in monitoring and evaluation activities, is essential, if A.I.D. is to understand the gender-specific effects of its project and programs. In order to determine whether A.I.D. evaluations were in conformance with this guidance, the review team coded evaluations with respect to whether they had collected data, of any sort, on a gender-disaggregated basis. Overall, only 56 (22%) of A.I.D.'s FY89 and FY90 presented data on a gender-disaggregated basis.

On a regional basis, evaluations of projects in the Near East Bureau did the best in this regard, with 35% of its evaluations presenting gender-disaggregated data. Evaluations that used structured approaches for collecting interview data as well as those that used direct measures, e.g., birth weight records, as a source of data were slightly more likely to include gender-disaggregated data than were other evaluations. Not surprisingly, evaluations whose purposes focused on questions other than project management and implementation issues, e.g., project impact or replicability, were also more likely to include gender-disaggregated data.

MSI also examined the relationship between the presence of women on evaluation teams, or the assignment of the responsibility for ensuring that gender issues were considered in the evaluation to a particular team member, and the frequency with which evaluations collected and presented data on a gender-disaggregated basis. The data provided by the evaluation review on these topics comes from a very small set of evaluations that reported on such matters.

For the full set of 268 evaluations, there were 105 teams that included female members and 116 that had no female team members. Insufficient information was available on team composition to code the other 47 teams. Of the teams that included female team members, 31 (30%) presented data on a gender-disaggregated basis, whereas only 19 (16%) of the all male teams presented data on a gender-disaggregated basis. It appears that the simple presence of women on evaluation teams leads to almost a doubling of the frequency with which gender-disaggregated data is presented.

Similarly, the presence of a female team leader, or the assignment of an explicit responsibility for gender considerations to a specific individual on an evaluation team seemed to make a difference in the frequency with which data were presented on a gender-disaggregated basis. Of the 32 evaluations that had female team leaders, 11 (34%) presented data on a gender-disaggregated basis. While this percentage is not high, it is higher than the 20% share of teams led by men that presented gender-disaggregated data. Only 11 teams included individuals who were assigned special responsibilities for gender considerations. Yet 7 (78%) of these teams used

gender-disaggregated data as compared to only 20% of the 258 teams which did not assign the responsibility for considering gender issues to a specific member of the evaluation team.

While all three of these measures suggest that the involvement of women or of an individual who has special responsibilities for considering gender issues within the context of an evaluation improves the likelihood that gender-disaggregated data will be presented, the small number of cases in which gender-disaggregated data were presented at all places some constraints on the degree to which generalizations can validly be made using these data.

4. The Use of Interviews in A.I.D. Evaluations

Interviews as a means of gathering new information during an evaluation cuts across evaluation designs and often enhances evaluations that depend heavily on other sources of information. Overall, 259 (97%) of the 268 FY89 and FY90 evaluations the review team examined used interviews to some degree. Even those studies that did not use either direct measures or site visits as a means of collecting data appear to have supplemented the pre-existing information they examined with some interviews.

In coding A.I.D.'s evaluations, the review team distinguished between structured interview approaches, in which standardized interview forms are used, and impressionistic approaches, which include all types of informal and unstructured interview approaches, whether used with individuals or groups. As Table 7-5 indicates, impressionistic approaches to interviews were

Table 7-5. Interviewed Approaches Used by Evaluation Teams

Interview Approaches	Number	Percent
Impressionistic (informal interviews for which no formal instrument or interview structure used)	188	70%
Structured (interviews utilized a formal instrument or followed structured guidelines)	31	12%
Combination of Impressionistic and Structured Interview Techniques	40	15%
Subtotal Involving Interviews	(259)	(97%)
No interviews or interview techniques cited	9	3%
All Evaluations	268	100%

used far more often than were structured approaches. Alone, or in combination with structured interview techniques, impressionistic approaches were used in 228 (85%) of all evaluations. In comparison, structured interview techniques were used, alone or in combination with impressionistic approaches, in 71 (26%) of A.I.D.'s FY89 and FY90 evaluations.

On a bureau basis, there were some variations with respect to the use of impressionistic and structured interview techniques. The use of only structured techniques was highest in the R&D and LAC Bureaus. Yet, evaluations in each of these bureaus used this approach only in 15% of their evaluations. The most frequent use of only impressionistic methods was in the Near East Bureau, where 82% of the interviews were carried out using this type of approach. In evaluations of projects in the Latin America and Caribbean Bureau, impressionistic approaches were used alone 66% percent of the time, and the R&D Bureau used these approaches were used alone even less frequently, i.e., 62% of the time.

While there was not much difference with respect to the interview techniques used in evaluations by evaluation type, single project evaluations were found to rely on only impressionistic methods more frequently than were multi-project evaluations. This was also the case for single point in time, or "snapshot" evaluations as compared to those that obtained information for multiple points in time. The review team also noted that in the evaluations carried out by A.I.D. staff, acting alone rather than serving as part of a mixed team, the evaluators depended on only impressionistic interview methods far more heavily than did evaluation teams which included representatives of other U.S. entities, e.g., contractors, universities, etc.

In order to determine how well evaluation teams documented their use of structured evaluation methods, the review team examined whether evaluations included copies of the formal questionnaires that had been administered, as well as sampling plans used to select individuals to be interviewed during an evaluation. With respect to the first of these questions, MSI found that only 48 evaluations included questionnaires or outlines of interview plans. A total of 40 (56%) of these evaluations were among the 71 which had been coded as using structured interview approaches, alone or in combination with impressionistic methods. What this indicates is that roughly 56% of the A.I.D. evaluations that used structured methods for their interviews also included the questionnaires they employed.

As to the issue of sampling plans, they were included in only 41 evaluations. MSI found that 30 (42%) of these evaluations were among the 71 that used structured interview methods, alone or in combination with impressionistic methods. Another 10 evaluations that used only impressionistic methods only were coded as containing the equivalent of sampling plans, i.e., descriptions of how "focus groups" or community interviews were set up. These 10 evaluations represented 5% of the 188 evaluations that used only impressionistic methods.

With respect to regional bureaus, evaluations of projects in the LAC Bureau more frequently included both questionnaires (21% of the evaluations) and sampling plans (25% of the evaluations) than did evaluations for other regional bureaus. Evaluations of projects in the R&D Bureau, which had formal questionnaires for 23% of its evaluations, did nearly as well as LAC in this regard, but these evaluations included substantially fewer sampling plans.

What is particularly interesting to note with regard to the inclusion of formal questionnaires and sampling plans is the frequency with which these kinds of evaluation tools were incorporated in evaluations in the "snapshot" category. Single-point-in-time, or "snapshot" evaluations, are sometimes thought of as always being impressionistic in nature. Yet that is clearly not always the case.

While single-point-in-time evaluations do include a large number of projects that use only impressionistic interview methods, i.e., 171 (73%) of 234 evaluations in this category, some single-point-in-time studies use more structured approaches. Of the total of 71 evaluations that used structured interview approaches with any evaluation design, 57 (80%) were associated with single-point-in-time, or "snapshot" evaluations. Similarly, of the 48 evaluations that included formal questionnaires in connection with any type of evaluation design, 41 (85%) were single-point-in-time evaluations. Finally, 31 (76%) of the 41 evaluations which described a sampling plan in connection with interviews also used a "snapshot" or single-point-in-time evaluation design.

Looking broadly across the data, it is clear that the range of detailed methods used in single-point-in-time, or "snapshot" evaluations is quite wide. Further efforts to document and understand that range may be warranted.

5. The Quality of Methods used to Obtain Evidence In A.I.D. Evaluations

In order to develop a broad picture of the quality of evaluation methods reported in A.I.D. evaluations, the review team created a composite measure out of several of the more detailed technical characteristics of evaluations discussed above. This composite measure scored evaluation methods as providing evidence at one of three quality levels: poor, adequate and good.²⁵ Definitions of these levels are provided below:

Poor	Evaluations were scored as being poor on methods if they did not specify their data sources or if they used only pre-existing data, i.e., did not collect any new data. Projects were also scored as having been deficient in their methodology if they collected new data using structured data collection techniques but provided neither their instrument nor their sampling plan.
Adequate	Evaluations scored as being adequate on evaluation methods were those that used structured methods to collect new data and provided either a complete or partial data collection instrument or their sampling plan. Evaluations that collected new data using impressionistic data collection methods, but which did not obtain beneficiary data, were also scored as being adequate.

²⁵ While this methods composite treated interim and final evaluations, which account for the bulk of the evaluations in the data base, with an even hand, the way in which it score evaluations which used only pre-existing data operated as a bias against reviews of "lessons learned" and other studies which drew only upon previous evaluations.

Good The term good was reserved for those evaluations that used structured data collection approaches and that presented both a data collection instrument and sampling plan. It was also applied to those evaluations that used impressionistic data collection methods and which also obtained beneficiary data.

Applying this composite score to the 268 evaluations in the FY89 and FY90 database, MSI found that 69 (26%) of the evaluations received scores of "poor" on the methods composite; 180 (67%) were scored as being "adequate" and only 16 (6%) were scored as being "good."

Of the 16 evaluations that received a rating of "good" on the methods composite, 7 (44%) were evaluations of projects in the LAC Bureau and another 4 (25%) were evaluations of projects in the R&D Bureau. The majority of evaluations rated as "good" on the methods composite were interim evaluations and evaluations that dealt with single projects. However, this distribution was only proportional to the frequency of these kinds of evaluations appeared in the database. The presence of individuals with evaluation skills, broadly defined, on evaluation teams also appears to be positively related to higher scores on the evaluation methods composite, but only slightly so.

CHAPTER EIGHT

THE FINDINGS OF A.I.D. EVALUATIONS

This section examines the conclusions and recommendations offered by A.I.D. evaluations. With regard to the general manner in which evaluations handled findings, conclusions and recommendations, MSI found that:

- Roughly 40% of the 268 evaluations in the database clearly distinguished findings (or facts) from conclusions (or interpretations).
 - But only 22 (8%) of the evaluations presented possible alternative interpretations of the facts they had gathered.
- A larger share of A.I.D.'s evaluation, 89%, clearly identified their recommendations, and
 - Of the 157 interim evaluations in the database, 58 (40%) recommended modifications in project Logical Frameworks.

In the paragraphs below, MSI reviews in greater detail the general conclusions that evaluations reached about project and program performance as well as their specific conclusions with regard to three cross-cutting issues: the sustainability of A.I.D. projects and programs; environmental impacts; and project effects on a gender-specific basis.

A. Findings and Conclusions About Basic Project Performance

In order to aggregate evaluation conclusions concerning project performance across a large number of evaluations, MSI coded each evaluation in terms of the answers that the evaluation provided to a set of questions about general project or program performance, including whether:

- Project outputs had been or were being provided;
- The project's purpose had been or was still likely to be achieved.
- The project was judged as being efficient, i.e., costs in relation to performance were as planned.
- Unplanned project effects, of either a positive or negative character had been noted.
- On an overall basis, the evaluators had judged the project or program to be successful.
- There was a good fit between the project or program and a mission's Country Development Strategy Statement (CDSS), where applicable.

The answers provided by A.I.D.'s FY89 and FY90 evaluations are discussed in the following paragraphs.

1. Actual Performance As Compared to Plans

Overall, the review team found that 205 (76%) of A.I.D.'s FY89 and FY90 evaluations reached conclusions about progress in creating project outputs. Of the 268 evaluations, 165 (62%) reached conclusions concerning the realization of a project or program's purpose. A much smaller share, (22%) reached conclusions at the goal level of a project. When only final and ex-post evaluations are considered, the percentage that reached conclusions at the purpose and goal levels of projects did not change substantially. The percentages, considering only final and ex-post evaluations were 59% and 27%, respectively. As these data suggest, interim evaluations turned out to be just as likely to consider purpose-level achievements and almost as likely to reach conclusions at the goal level.

At both the output and purpose levels, evaluators concluded that, in most instances, all or some aspects of a project's outputs and purpose had been achieved. Very few evaluations concluded that no progress whatsoever had been made. At the output level, 3 (1%) of the evaluations reached this conclusion, while 10 (4%) came to a similar conclusion at the purpose level. On a bureau basis, Table 8-1 shows the frequency with which evaluators reported that projects had achieved some or all of their outputs and either fully or partially realized their purpose.

As Table 8-1 indicates, the tendency for evaluations to reach conclusions at the output, but not the purpose level, was more pronounced in the regional bureaus and in the R&D Bureau than it was, for example, in broad multi-project evaluations. At both the purpose and the output levels, single-project evaluations were a good deal more likely to reach conclusions about project progress than were evaluations of multiple projects. While there were very few evaluations with experimental or quasi-experimental designs in the database, those that were included seemed to address the question of output- and purpose-level performance more frequently than did single-point-in-time, or "snapshot" evaluations.

Evaluations were also found to differ in the degree to which they reached conclusions at various performance levels as a function of the evaluation purposes they had cited. At the output level, for example:

- Evaluations stating that their only purpose was the investigation of management and implementation issues reached conclusions about output level performance 67% of the time.
- Evaluations identifying a broader range of purposes reached conclusions about project performance at the output level 84% of the time.

At the purpose level, the gap between evaluations that did and did not examine more than management and implementation issues was similar.

Table 8-1. Degree to Which Evaluations Report That Projects are Achieving Their Outputs and Realizing their Purposes

Bureaus	Total of Bureau Evaluations	All or Some of the Project Outputs Were Achieved	Percent of All Projects That Reportedly Achieved Some or All Outputs	All or Some of the Purpose Level Objectives Were Achieved	Percent of All Projects That Reportedly Achieved Some or All of Their Purpose Level Objectives
Asia	51	38	75%	29	57%
Near East	17	12	71%	8	47%
Latin America and the Caribbean (LAC)	68	59	87%	47	69%
Africa	67	50	75%	39	58%
R & D	39	32	82%	22	56%
FHA	5	4	80%	3	60%
PRE	5	3	60%	3	60%
Broad Multi-Project Evaluations	16	4	25%	4	25%
All Evaluations	268	202	75%	155	58%

Turning to project efficiency, i.e., the degree to which project or program outputs were in line with anticipated costs, the review team found that of 268 evaluations in the database, 65 (24%) contained information on the efficiency of projects. This percentage was roughly the same as that found in the FY87-88 evaluation review. In 63 of the evaluations in the FY89 and FY90 database, all or some components of projects were considered to be reasonably efficient. In three of these evaluations, projects were found to be inefficient in this sense.

2. Unplanned Effects

In addition to examining the planned results of projects, the A.I.D. Evaluation Handbook encourages evaluation teams to look broadly at all of the positive and negative effects of projects before reaching conclusions concerning their impact.²⁶ To assess the degree to which A.I.D.

²⁶ A.I.D. Evaluation Handbook, op. cit., Section 3.4.

evaluations incorporate this guidance, MSI noted when evaluations identified unanticipated effects of the projects and programs they examined. Table 8-2 summarizes data on the frequency with which evaluations discussed the unplanned positive and negative effects of projects and programs. Evaluations of projects in the Near East Bureau seemed to incorporate this type of information more frequently than did evaluations of projects in other regional bureaus. On a percentage basis, large multi-project evaluations identified unanticipated positive and negative effects more frequently than did smaller evaluations carried out for projects in any of the bureau clusters shown on the table.

Table 8-2. Frequency with Which Evaluation Identified Unanticipated Effects of Project Activities

Bureaus	Total of Bureau Evaluations	Evaluations Reported Unanticipated Positive Effects		Evaluations Reputed Unanticipated Negative Effects	
		Number	Percent	Number	Percent
Asia	51	4	8%	5	10%
Near East	17	3	18%	3	18%
Latin America and the Caribbean (LAC)	68	8	12%	9	13%
Africa	67	7	10%	9	13%
R & D	39	4	10%	3	8%
FHA	5	1	20%	--	--
PRE	5	1	20%	--	--
Large Multi-Project Evaluations	16	5	31%	5	31%
All Evaluations	268	33	12%	34	13%

Consistent with the finding that broad multi-project evaluations tended to record unanticipated project effects somewhat more frequently than did smaller evaluations, the review team found that, in general, multi-project evaluations reported on unplanned effects, particularly positive effects, more frequently than did single-project evaluations. In addition, final and ex-post evaluations, as well as reviews of "lessons learned," commented on unplanned effects more frequently than did interim evaluations. Coverage of unplanned effects in an evaluation also appeared to be slightly better when evaluation teams were relatively large and when they

included at least one individual with evaluations skills, as broadly defined in the A.I.D. Evaluation Handbook.

3. The "Fit" Between Projects and a Broad Mission Program (CDSS)

In A.I.D.'s overseas missions, project success is in part a function of the degree to which it successfully implements elements of an overall development strategy. To gauge this aspect of project success, MSI noted when evaluations indicated whether there was a good, poor or mixed "fit" between a project and a mission CDSS. This measure makes sense only for the 177 evaluations of "mission-owned" projects in the database. Within this subset, 47 (27%) of the evaluations of "mission-owned" projects commented on the fit between the project and the mission's CDSS. Of these evaluations, 36 (20%) reported that there was a good "fit;" while 11 (6%) indicated that the "fit" was mixed. The majority, 130 (73%) did not address this question.

4. Overall Performance of A.I.D.'s Projects and Programs

In addition to coding evaluations on a series of discrete success measures, the review team noted when evaluations made overall statements about the success of a project or the lack thereof. This overall measure was viewed as subsuming and balancing out whatever inconsistencies may have existed at more discrete levels at which performance was judged.

For 249 (93%) of A.I.D.'s FY89 and FY90 evaluations, evaluation teams reached an overall conclusion about project performance. For 241 (97%) of these evaluations, evaluation teams concluded that some or all of the project or programs objectives had been achieved. Only 8 (4%) of these evaluations were judged to be unsuccessful. Table 8-3 shows the frequency, by bureau, with which evaluators stated that some or all of the objectives of projects or programs had been or were being achieved.

Table 8-3. Frequency with Which Evaluations Judged that Overall Projects Were Achieving Some or All of the Objectives

Bureau	Evaluations Which Report That Overall Projects are Achieving Some or All of Their Objectives	Total Number of Bureau Evaluations	Percent Which Overall are Achieving Some or All of Their Objectives
Asia	45	51	88%
Near East	14	17	82%
Latin America and the Caribbean (LAC)	61	68	90%
Africa	58	67	87%
R & D	38	39	97%
FHA	5	5	100%
PRE	5	5	100%
Broad Multi-project Evaluations	15	16	94%
All Evaluations	241	268	90%

As this table suggests, many evaluations which did not reach conclusions with respect to the degree to which the outputs or the purpose of a project had been achieved, nevertheless they reached an overall conclusion concerning project or program success. This pattern pertained for evaluations which used project Logical Frameworks as a basis for making judgements about performance and for evaluations which did not utilize this tool. Table 8-4 illustrates the degree to which overall conclusions as compared to detailed conclusions were reached in evaluations. It also shows how such conclusions differed as a function of the timing or type of evaluations in which they were drawn. As Table 8-4 makes clear, the number of evaluations in which overall positive conclusions are reached is quite a bit higher, for every type of evaluation, than is the number which reached more specific conclusions. The reason for this gap is not at all clear.

Table 8-4. The Effect of Project Status/Evaluation Timing on Evaluation Conclusions Concerning Achievements

Achievements Cited by The Evaluations	Interim Evaluations (N=159)		Final & Ex-Post Evaluations (N=74)		Other Evaluations and Those Score Can't Tell With Respect to Timing (N=35)	
	Number	Percent of All Interim Evaluations	Number	Percent of All Final or Ex-Post Evaluations	Number	Percent of All of The Evaluations
Some or All of the Project Outputs Were Achieved	127	80%	59	80%	16	46%
Some or All of the Project(s) Purpose Level Objectives Were Achieved	100	63%	44	60%	11	31%
Overall the Project(s) Were Achieving some or All Objectives	142	89%	69	93%	30	86%

Interim evaluations were only slightly less likely to reach overall positive conclusions than were final and ex-post evaluations. Similarly, multiple-project evaluations in a single country were slightly less likely than single-project evaluations to reach overall positive conclusions. Yet these differences were not strong enough to explain why so many evaluations which had not drawn any intermediate level conclusions, i.e., about outputs, purpose, efficiency or unplanned results, reached overall positive conclusions with respect to project success.

B. Findings and Conclusions about Cross-Cutting Issues

In addition to noting what evaluations said about basic project performance, the review team coded evaluations with respect to what they said about three cross-cutting issues: sustainability, environmental impact and gender considerations. These findings and conclusions are discussed in the following paragraphs.

1. The Sustainability of A.I.D. Projects and Programs

The term sustainability has several meanings, including the financial and managerial sustainability of projects; the degree to which benefits continue to flow to a project's target group; and the appropriateness, from a long-term perspective, of their relationship to natural resources. For purposes of this review, sustainability was used primarily to connote the continuation of project activities and hence their effects. Environmental impacts were handled separately, and are discussed in subsection (b) below.

Of the 268 evaluations in the FY89 and FY90 evaluation database, 116 (43%) intentionally assessed the sustainability of A.I.D. projects and programs in at least a minimal way. Of the 116 evaluations which did so, 54 (47%) provided the definition of sustainability they intended to apply. This percentage compares favorably to evaluations included in the FY87-88 review, where 36% were reported to have examined this issue. As Table 8-5 indicates, evaluations of projects in the Africa Bureau addressed this issue somewhat more frequently than did evaluations in other regional bureaus. In addition, both FHA and PRE, appeared to consider sustainability quite frequently, as did large multi-project evaluations.

Table 8-5. Frequency With Which Bureau Evaluations Explicitly Examined Sustainability Issues

Bureau	Evaluations That Explicitly Examined Sustainability	Total Number of Bureau Evaluations	Percent of Evaluations That Examined Sustainability
Asia	20	51	39%
Near East	6	17	35%
Latin America and the Caribbean (LAC)	30	68	44%
Africa	32	67	48%
R & D	11	39	28%
FHA	3	5	60%
PRE	3	5	60%
Broad Multi-project Evaluations	11	16	69%
All Evaluations	116	268	43%

Evaluations did not appear to differ with respect to whether they addressed sustainability depending upon whether one or many projects were involved. Nor were final and ex-post evaluations found to be significantly more likely to address sustainability than were interim evaluations. Those evaluations that examined sustainability did, however, appear to be distinguishable in two ways. First, a positive relationship was found between evaluation team size and discussions of sustainability, i.e., of the 55 evaluations that had a single evaluation team

member, 24 (44%) addressed project sustainability. In contrast, 71% of the 17 evaluations with four or more team members addressed sustainability. Evaluations that used multiple sources of data, i.e., existing records, plus direct measurement, plus site visits, also appeared to address sustainability more frequently than other evaluations.

On the broad question of whether project benefits would continue after A.I.D. funding ceases, 122 (44%) evaluations indicated that some benefits would continue, while only 77 (29%) suggested that all project benefits would continue. Table 8-6 shows the different ways in which evaluations were coded on these two questions.

Table 8-6. Frequency with Which Evaluations Conclude That Some or All Project Benefits Would be Sustained

Evaluation Conclusions Concerning Sustainability	Some Benefits Will Continue After A.I.D. Funding Stops		All Benefits Will Continue After A.I.D. Funding Stops	
	Number	Percent of All Evaluations	Number	Percent of All Evaluations
Issue Not Addressed	146	54%	191	72%
Low Probability	61	23%	54	20%
Moderate Probability	38	14%	17	6%
High Probability	23	9%	6	2%
All Observation	268	100%	268	100%

At a more detailed level, the review team coded evaluations with respect to the probability that projects were wholly or partially sustainable. Scores of high, medium and low were given. The resulting distribution is shown in Table 8-7. What is noteworthy is that the share of projects predicted to be sustainable and likely to continue providing benefits does not equate with the share were rated as being relatively successful in terms of achieving their overall substantive objectives.

In this vein it is worth noting that in 52% of the evaluations that concluded that project outputs and purposes were being achieved, the question of whether benefits would continue after A.I.D. assistance ends was not even addressed. Similarly, when final and ex-post evaluations were examined in detail, the differences between what evaluations say about performance in terms of basic project objectives and what they say about sustainability was easy to see. Of 44 final evaluations where teams concluded that projects would partially or completely achieve their purpose, only 8 (18%) were judged to be highly sustainable. Another 9 were coded as being low with respect to sustainability, and in 19 of these evaluations, the question of sustainability was simply not addressed.

Table 8-7. Frequency with Which Evaluations Reported That Some or All Projects Are Sustainable

Sustainability Rating	Some Aspects of the Project(s) Are Sustainable		All Aspects of the Project(s) Are Sustainable	
	Number	Percent	Number	Percent
High	61	23%	54	20%
Medium	38	14%	17	6%
Low	23	8%	6	2%
Not Scored	146	55%	191	72%
All Evaluations	268	100%	268	100%

To determine what sources of funds would allow projects to continue to provide benefits, the review team noted when evaluations identified other funding sources that were contributing to a project during the A.I.D. financing period as well as those sources that evaluations indicated would continue to finance project activities after the period of A.I.D.'s funding ended. Table 8-8 indicates the frequency with which host country budgets, other donors and the private sector, were mentioned. As the table indicates only a small fraction of A.I.D.'s evaluations identify how projects will be financed on a sustained basis. With regard to this table, MSI also noted that while host country budgets were the most frequently cited source of funds, other than A.I.D., private sector funding was mentioned nearly as often.

Table 8-8. Frequency With Which Evaluations Identify Other Sources of Project Funding

Other Sources of Project Funds	Project Funding During the Life of the A.I.D. Project		Project Funding After A.I.D. Funding Ceases	
	Number of Evaluations	Percent of All Evaluations (N=268)	Number of Evaluations	Percent of All Evaluations (N=268)
Host Country Government	83	31%	54	20%
Other Donors	24	9%	28	10%
Private Sector (User Fees, Etc.)	66	25%	71	26%
Other Sources	16	6%	18	7%
None Identified	85	32%	149	56%

Of particular interest to A.I.D. in recent years has been the incorporation of user fees into projects as a means of ensuring the long term provision of goods and services. Overall, only 17% of the projects examined through FY89 and FY90 appear to incorporate user fees into long-term financing questions, as Table 8-9 indicates. User fees were more frequently associated with projects in the LAC and Near East Bureaus than was the case for other bureaus. Evaluations of PRE Bureau projects suggest that user fees are also being incorporated into this bureau's projects.

Table 8-9. Frequency with Which Bureau Projects Reportedly Utilize User Fees During the Life of a Project

Bureau	Number of Evaluations That Cite User Fees as a Source of Funds During the A.I.D. Project Life)	Total Number of Bureau Evaluations	Percent Citing Private Sector Funding
Asia	5	51	10%
Near East	4	17	24%
Latin America and the Caribbean (LAC)	18	68	26%
Africa	10	67	15%
R & D	3	39	8%
FHA	--	5	--
PRE	2	5	40%
Broad Multi-project Evaluations	3	16	2%
All Evaluations	45	268	17%

As a counterpoint to questions about sources of funds, evaluation comments concerning the affordability of goods and services provided by projects were notable. Of the 268 evaluations in the database, 64 (24%) provided clear answers on this question. Of these, 26 (41%) said that services and goods provided by A.I.D. projects were affordable while 38 said that they were not. To the degree that user fees are expected to account for a significant portion of a project's long term financing, the affordability of project goods and services may be a critical issue.

2. Environmental Impact

Table 8-10. Degree to Which Evaluations Addressed Environmental Concerns

Bureau	Evaluation in Which Environmental Concerns Were Addressed	Total Number of Bureau Evaluations	Percent of Evaluations That Addressed Environmental Concerns
Asia	8	51	16%
Near East	2	17	12%
Latin America and the Caribbean (LAC)	8	68	12%
Africa	13	67	19%
R & D	6	39	15%
FHA	--	5	--
PRE	1	5	20%
Broad Multi-project Evaluations	3	16	38%
All Evaluations	41	268	15%

A small fraction of A.I.D.'s FY89 and FY90 evaluations examined environmental issues. As Table 8-10 indicates, 41 (15%) of the evaluations in the database set out to address environmental impact issues. This percentage is lower than was found for evaluations in the FY87-88 evaluation review, where 25% of the evaluations appear to have examined environmental questions. Africa Bureau evaluations included an examination of environmental impact more frequently than was the case in other bureaus, but even here the overall percentage fell below 20%. A somewhat higher than average share of the broad multi-project evaluations considered environmental impact.

While 41 evaluations set out to examine environmental issues, a much smaller number reached conclusions about environmental effects. Only 25 evaluations indicated that planned environmental outcomes had been achieved to any degree. With respect to unplanned environmental effects, 4 evaluations reported on negative effects while 5 reported that the unplanned environmental effects of projects had been either positive or neutral.

3. Gender Considerations

In MSI's review of A.I.D.'s FY89 and FY90 evaluations, a number of questions were examined to determine the degree to which gender-specific issues were being pursued. As it turns out, 70 (26%) of the evaluations in the database of 268 evaluations included information that facilitated an understanding of women's participation in projects or the degree to which women received benefits from A.I.D. financed project and programs. This is fewer than the 33% of evaluations in the FY87-88 review that were reported to have examined gender issues in some detail.

From a data perspective, only 7% of the evaluations in the database reported that the objectives of projects examined were articulated on a gender-disaggregated basis. As already noted, in Chapter Seven, only 11% of A.I.D.'s FY89 and FY90 evaluations collected new data on a gender-disaggregated basis.

From a design perspective, 8% of A.I.D.'s evaluations reported that the projects that were evaluated had identified obstacles to women's participation in their designs. A similar percentage reported that projects contained strategic plans for overcoming such obstacles. A slightly smaller percentage of evaluations, 6%, reported that, during the course of the project, obstacles to women's participation had been eliminated.

As to the types of benefits that evaluations discussed, on a gender-specific basis, the review team examined two: training, and other project services and benefits. With respect to training, the review team found that 180 (67%) of the 268 evaluations in the database reported that training had been provided by the projects they examined. Only 66 (25%) stated that women had been trained by these projects. The percentage of evaluations that reported that women had been trained was slightly higher for projects in the Asia and Near East Bureaus than it was for other bureaus.

Gender-disaggregated information on other project services and benefits reflected a similar situation. Of the 268 evaluations in the database, 178 (66%) reported that services and benefits had been provided directly to people in the projects they evaluated. In 50 (19%) of these evaluations was there information suggesting that women had received project services or other benefits. In 42 evaluations, there was adequate information to determine the share of benefits received by women, as Table 8-11 indicates. It should be noted that given the small number of projects that reported data of this type, it is difficult to generalize from these answers to A.I.D.'s full portfolio of projects.

Table 8-11. Share of Project Benefits Which Evaluations Report Are Received by Women

Share of Project Benefits Received by Women	Number of Evaluations	Percent
0 - 20%	12	4%
21% - 40%	15	6%
41% - 60%	5	2%
61% - 80%	2	1%
81% - 100%	8	3%
Can't Tell	226	84%
All Evaluation	268	100%

CHAPTER NINE

CONCLUSIONS AND RECOMMENDATIONS

MSI's review of A.I.D.'s FY89 and FY90 evaluations indicates that, broadly speaking, A.I.D.'s evaluation system is healthy. By and large, the A.I.D. evaluation system serves the intended purpose of providing management-useful information to relatively large numbers of mid-level staff who design and administer A.I.D. projects. At the same time, there are improvements that can and should be made. A.I.D.'s current efforts to introduce strategic evaluations that will benefit the Agency's top managers are appropriate in this regard. Detailed conclusions and recommendations for A.I.D. are provided in the following thirteen paragraphs.

1. Portfolio Coverage

Information from this evaluation review and from other sources indicates that:

- A.I.D.'s evaluation coverage of its portfolio is substantial for a U.S. government program.

On an annual basis, A.I.D. evaluates 150-200 projects and program, or 7% - 9%, of its portfolio through a mix of single-project and multiple-project evaluations. As of the end of FY90, A.I.D.'s portfolio contained 1910 projects that had a combined life-of-project value of \$38 billion. In each of the two years covered by the FY89 and FY90 evaluation review, CDIE acquired around 125 single- and multi-project evaluations that were complete enough to include in the database. The combined life-of-project value of the projects evaluated in each year covered by this review was approximately \$3 billion.

This level of coverage was approximately the same as was found in earlier reviews. As A.I.D. itself has reported, it achieves this level of evaluation with expenses that come to less than .02% of its annual budget. These expenditures are significantly lower than the 1% of budget which some federal agencies set aside for evaluation.²⁷

Given the size of A.I.D. programs, and in the nature of the evaluation work it undertakes, anecdotal information suggests that A.I.D.'s evaluation performance compares that of other U.S. government agencies. It also compares well to the performance of the World Bank. Operations evaluations at the Bank, which do not necessarily involve field work, have, on average, examined 130 projects, with a combined value of \$11 billion, each year for the past fifteen years.²⁸

²⁷ A.I.D., "The A.I.D. Evaluation System: Past Performance and Future Directions" op. cit., p. 16.

²⁸ World Bank, Evaluation Results for 1988. Washington, D.C.: International Bank for Reconstruction and Development 1990, p. 2.

2. Report Completeness

Findings from this evaluation review, together with the results of prior evaluation reviews suggests that:

- **With respect to their completeness, A.I.D. evaluations are not making steady improvements over time.**

Since FY86-87, A.I.D. has been publishing the results of its bi-annual evaluation reviews. While each of these reviews has had a unique scope of work, they have overlapped in their coverage of several evaluation factors which, while not providing a complete longitudinal picture of completeness, serve as a partial basis for assessing progress with regard to the completeness and coverage, but not the quality of A.I.D.'s evaluations. The changes that have occurred over the six years covered by three evaluation reviews appear to be somewhat random.

Over the six-year period, the share of evaluations that include scopes of work has risen from 49% to 74%. Those that provide at least a minimal discussion of evaluation methods has risen from 75% to 87%. At the same time, the number of evaluations for which required A.I.D. Evaluation Summaries were provided declined from 68% to 49%.

In addition to leaving out scopes of work and descriptions of evaluation methods, MSI found that many evaluations failed to identify the projects they examined using A.I.D.'s formal, seven-digit project numbers. The problems these omissions caused, from an evaluation review perspective, were compounded when evaluation teams cited the number of the A.I.D. project that paid for their work. Projects that paid for evaluations were often large centrally funded efforts that focused on sectoral and technical issues. When evaluations do not cite correct project numbers for the projects they evaluated the result can be the incorrect allocation of evaluations by bureau, country or office.

RECOMMENDATION: Set annual targets for improvements in the completeness of A.I.D. evaluation reports and decentralize to bureau and mission evaluation officers the responsibility for meeting and reporting on this aspect of performance.

-- *Develop a standard set of measures of evaluation report completeness that are to be used thereafter in all evaluation reviews and related quality-control activities to assess performance.*

3. Constraints on the Complete Reporting of Evaluation Methods

While reporting on methods has improved, compared to prior years, it is often inadequate. MSI found it difficult to judge evaluation quality when there was only scant reporting on evaluation methods. Yet evaluations that provided only a brief, single-page description of their methods were fully in compliance with the A.I.D.'s Evaluation Handbook. Whether intentionally

or not, A.I.D.'s guidance is discouraging the presentation of full and complete methodology sections in evaluation reports.

RECOMMENDATION: That A.I.D. revise its evaluation guidance to eliminate constraints on the full and complete reporting of evaluation methods in appropriate annexes, if not within the body of evaluation reports.

4. Clear Identification of Team Expertise

Given the types of evaluations A.I.D. is carrying out, the findings of this evaluation review indicated that:

- Too frequently, A.I.D. evaluations ask their readers to make a leap of faith concerning the foundation upon which their conclusions and recommendations rest.

Only a small fraction of A.I.D.'s evaluation teams include individuals with evaluation skills as they are broadly defined in the A.I.D. Evaluation Handbook. Furthermore, as the large number of A.I.D. evaluations that utilize impressionistic data collection approaches at only a single-point-in-time suggest, A.I.D. is relying heavily on the "expert judgement" of its evaluation teams. Yet, in a third of these evaluations, no description of the skills and expertise of the evaluation team members was presented.

RECOMMENDATION: That A.I.D. clarify and reinforce its requirement for A.I.D. evaluations to discuss team composition.

- *Revised instructions should explicitly call for the inclusion of information about the pertinent experience of team members who are serving as "experts" in some capacity on an evaluation team.*
- *Descriptions of team composition should also be required to identify which, if any, members of an evaluation team have skills and experience that conforms to A.I.D.'s broad definition of evaluation skills.*

5. Internal and External Evaluations

For a number of very good reasons, A.I.D. strongly encourages the participation of its staff as well as the staff of sponsoring host ministries in interim evaluations. Conversely, it discourages the participation of these same actors in final and ex-post evaluations. In the former case, A.I.D.'s position focuses on the value to these parties of the knowledge they gain about the projects they are administering. In the latter case, A.I.D.'s position is based on the need for objectivity.

- **Irrespective of what A.I.D.'s guidance says in this regard, A.I.D. staff and host ministry personnel currently participate as team members on both interim and final evaluations.**

As a practical matter, this review of A.I.D.'s evaluations found that there is little difference in the range of questions that interim and final evaluations examine, or in the methods they use to answer questions. The difference between them lies in how far along a project is in its product or service delivery cycle.

Given the many practical similarities between A.I.D.'s interim and final evaluations, one might argue that, for both a mission and for the ministry with which it is working, participation in final evaluations is simply another step in a continuing process that helps to build knowledge for the future as well as for the present.

RECOMMENDATION: That A.I.D. revisit the rationale for discouraging the participation of A.I.D. staff and sponsoring host ministry personnel who have worked with a project on final and ex-post evaluation teams. If the current policy still seems appropriate, enforce it. If it does not, change it.

6. Host Country Participation in A.I.D. Evaluations

When the evaluation review broadened the question of the level of A.I.D. and host country participation to cover all aspects of the evaluation process, it quickly became apparent that:

- **Host country personnel are not being brought into the evaluation process to the degree that any reasonable "stakeholder analysis" of projects and programs would suggest is appropriate.**

Not only do sponsoring host ministry personnel infrequently serve on evaluation teams, there is little in A.I.D.'s evaluations to suggest that they are being included in oral briefings, or being asked to review draft evaluation reports.

From a "stakeholder" perspective, the absence of sponsoring host ministry personnel participation in interim evaluations of "mission-owned" projects is only one aspect of the issue. Broad multi-country evaluations that result in the adoption of strategies A.I.D. then applies on a sectoral, continental or world-wide basis almost never include ministry representatives. Yet, the staff of these ministries have an overwhelming need to understand strategic options and to learn from the experiences of other countries.

RECOMMENDATION:

That A.I.D. re-examine its commitment to including sponsoring host ministry personnel on evaluation teams. If the commitment remains a serious one, improve A.I.D.'s performance in this regard.

- *Identify current barriers to host country involvement in all aspects of A.I.D.'s evaluation processes and define methods of overcoming them.*
- *Experiment with the inclusion of host ministry personnel on broad multi-project and multi-country evaluations that have a strategic or sectoral orientation from which they could benefit and to which they could contribute.*

7. Evaluation Timing

The evaluation review confirmed the findings of other evaluation reviews concerning A.I.D.'s evaluations from an evaluation timing perspective.

- **The A.I.D. evaluation system continues to emphasize the decentralized production of management-useful interim and final evaluations of individual projects.**

Broadly speaking, this is as it should be.

The A.I.D. evaluation system is functioning as anticipated when it is producing a distribution of evaluations that resembles a series of levels on a pyramid. As A.I.D. already knows from prior evaluation reviews, its system produces a relatively large number of mid-term evaluations, a smaller number of final evaluations, a handful of ex-post evaluations and occasional studies that review the lessons other evaluations have produced. Of the four levels in this pyramid, the bottom two levels, which contain interim and final evaluations of A.I.D. projects and programs, appear to be the most stable.

After twenty years, interim and final evaluations are an accepted and integral part of A.I.D.'s management system. For the most part, they are scheduled and funded by the management units (e.g., overseas missions or the technical offices in central bureaus) that administer the projects and programs being evaluated. While the production of interim and final evaluations in A.I.D. does not occur on "automatic pilot," it fits this image about as well as might be hoped in a large bureaucracy.

A.I.D.'s investments, over these same years, in ex-post or "impact" evaluations, reviews or syntheses of "lessons learned," strategic or issue-oriented evaluations, and other types of evaluations, have been less systematic than is the case for interim and final evaluations. Accordingly, there are fewer evaluations in these categories. Further, in contrast to the kinds of family resemblances that can be found among interim and final evaluations over time, the

substantive nature and coverage of evaluations that are clustered in the top levels of A.I.D.'s evaluation pyramid under the labels "ex-post," "lessons learned," and "other" can be quite different from review to review. New program and policy assessments, and new operations and management assessments, which CDIE is introducing this year, illustrate both CDIE's responsibility for advancing evaluation practice in A.I.D. and the way in which evaluations in the top two layers of the A.I.D. evaluation pyramid tend to mutate over time in response to Agency needs.²⁹ As their initiation is not coming at the cost of more routine evaluations upon which mid-level managers depend, their introduction represents a net gain for the evaluation system.

8. Substantive Coverage in Evaluations

MSI's examination of the range of questions addressed in different types of A.I.D. evaluations indicated that:

- The range of issues that are being addressed in A.I.D.'s interim evaluations is broader than that term might suggest.

The conventional wisdom, which holds that interim evaluations of individual projects tend to concern themselves only with management issues, and thereby have little of value to say to those who are concerned with sectoral issues and policies, is not accurate.

- Only 30% of A.I.D.'s interim evaluations limited their inquiry to management and implementation issues. Many interim evaluations, like final evaluations, examined issues such as the appropriateness of a project's design and the probability it could be replicated elsewhere.
- Of the FY89 and FY90 evaluations that reached conclusions and offered recommendations at the sectoral level, 54% were interim evaluations as were 46% of the evaluations that offered conclusions that were multi-sectoral in nature.

To the degree that interim as well as final and ex-post evaluations produced at the mission level reach findings that may have broader implications, they should be captured and fed into the Agency's program and policy decision-making process on a timely basis.

RECOMMENDATION: That A.I.D. develop an "early alert" procedure, which can be used by mission or bureau staff, to notify CDIE whenever an evaluation reaches sectoral, multi-sectoral or other broad-gauged conclusions. Such a system should allow CDIE to feed pertinent results of evaluations, from any source, into Agency-wide decision-making processes on a more timely basis.

²⁹ A.I.D., Handout: "Administrator Strengthens Role of Evaluation in A.I.D.", 1991.

9. Evaluation Approaches

As a practical matter, the majority of A.I.D.'s evaluations are single-point-in-time studies. Longitudinal evaluations, which collect consistent information before and after a project delivers the goods and services it is designed to provide, are few and far between.

- **Given that the overwhelming majority of A.I.D. evaluations are single-point-in-time endeavors, there is a need in A.I.D. for standards of "evaluation quality" that are specific to single-point-in-time evaluations.**

A separate definition of "evaluation quality" for longitudinal evaluations may or may not be needed, since the evaluation literature is replete with discussions of the standards to be applied when experimental and quasi-experimental evaluation design are used.

Looking across the 234 single-point-in-time evaluations in the FY89 and FY90 database, the thing that was striking was their methodological diversity.

- Some used trend data to expand their understanding of changes which may have occurred, while others did not.
- Some gathered beneficiary data, while others did not.
- Some used structured interview techniques, and included their interview forms and sampling plans in their reports, while others did not.
- Some gathered data on a gender-disaggregated basis, while others did not.
- Some included site visits and direct measures of performance, while others used only data from existing reports.
- Some drew performance indicators out of project Logical Frameworks, while others did not.
- Some examined performance in relation to cost, while others did not.

These dimensions are only a few of the factors that could be considered in assessing evaluation quality for single-point-in-time studies.

Developing broad standards and composite measures of quality, is not, however, an easy matter. Each of the bi-annual evaluation reviews has taken on this challenge to some degree, as have other CDIE activities over the years. When MSI developed a composite methods rating for this study, using only a few of the factors listed above, it found that high quality almost had to be separately defined for several evaluation subtypes, e.g., those that use impressionistic data-collection methods versus those that use more structured techniques.

RECOMMENDATION: That A.I.D. develop "working models" of what high and low quality in single-point-in-time evaluations means in the A.I.D. context.

-- *Test these models retrospectively against existing evaluations and prospectively with a sample of upcoming evaluations prior to issuing standards and measuring conformance with them in future bi-annual evaluation reviews.*

10. The Credibility of Overall Assessments of Performance

In examining the findings of A.I.D.'s evaluations, MSI discovered a substantial discrepancy between detailed evaluation findings and overall judgement about project success, which suggests:

- There is a tendency in A.I.D. evaluations to give projects the benefit of the doubt when making overall judgements about performance.

The facts that bear out this conclusion can be summarized briefly:

- In 60% of A.I.D.'s final evaluations no conclusion was reached about project achievement at the purpose level. Yet in 93% of A.I.D.'s final evaluations, teams reached the overall judgement that the projects were succeeding.
- Of the 268 evaluations MSI examined, only 43% addressed sustainability issues. Only 99 (37%) were reported to have a moderate to high probability of being sustained. Yet 90% of all evaluations were reported, on an overall basis, to be succeeding.

Where achievement at the purpose level in final evaluations is not being reported, and in all evaluations where sustainability is either not addressed or reported to be low, it is difficult to understand what evaluation teams mean when they report that, on balance, projects are succeeding. The standards of evidence for such judgments need to be clearer and higher than this review suggests they are currently.

RECOMMENDATION: That A.I.D. define standards of evidence concerning project performance and sustainability that must be met in order to concluded that projects are either succeeding or failing.

11. The Coverage of Cross-Cutting Issues in A.I.D. Evaluations

The findings of this and prior evaluation reviews indicated that:

- A.I.D. evaluations pay only a very limited amount of attention to cross-cutting issues that are of interest on an Agency-wide basis.

Statistics concerning evaluation coverage of cross-cutting issues are even less encouraging. While A.I.D. has tracked evaluation coverage of sustainability, environmental impact and gender issues for several years, evaluations are not technically required to address these or other cross-cutting issues.

With that caveat in mind, the review team found that the share of evaluations that examined the question of program sustainability, at even a minimal level, rose from 36% for FY87-88 to 43% for FY89-90. The share of evaluations that considered environmental issues and impacts declined from 25% in FY87-88 to 15% in FY89-90. Evaluation coverage of gender issues also appears to have declined somewhat. Whereas 33% of the FY87-88 evaluations considered women in development issues in some way, only 23% of the FY89-90 evaluations presented data on a gender-disaggregated basis and only 19% of those projects that provided direct benefits to people reported on whether women had received some portion of those benefits.

RECOMMENDATION: That A.I.D. decide whether evaluations are to be required to address cross-cutting issues and, for those cross-cutting issues where the answer is "yes," issue special guidance, or revise the A.I.D. Evaluation Handbook, to make both the requirements and appropriate procedures clear.

12. Improving Evaluation Quality

To the degree that MSI was able to examine evaluation quality issues using A.I.D. evaluation reports, it found that bureaus were relatively even, at the aggregate level, in the degree to which they focused on such matters in evaluations. While the Asia Bureau may have been better at one aspect of the overall task, Africa was good at something else, just as LAC, the Near East, R&D and the other central bureaus all had their strengths.

- - Relatively even performance at the bureau level with respect to evaluation completeness, coverage and methods issues, however, masks substantial quality control problems within bureaus. In each bureau for which MSI examined a substantial number of evaluations, some were quite good and others were very bad.

A quality control system that brings evaluations to a uniformly higher standard may need to be administered on a "real time" basis, i.e., as scopes of work are developed or when evaluations are in draft, rather than after the fact. Such a system need not be complicated. Theoretically, it could be constructed on a checklist basis.

While CDIE can monitor quality across the Agency, it may not be appropriate for CDIE to try to administer an evaluation quality-control program at the mission level or in offices within AID/Washington bureaus. A "real time" quality control system would, almost by definition, need to be administered at the bureau level.

RECOMMENDATION: That A.I.D. develop an approach for administering a simple and effective "real time" evaluation quality control system.

- *Test the approach in sample missions and offices across bureaus, rather than pilot testing such a system in a single bureau. Ownership of such a system must be broadly based.*

Looking beyond the findings of this evaluation, there is an opportunity CDIE may wish to act upon.

Pursuant to the October, 1990, announcement by A.I.D.'s Administrator on strengthening evaluation, CDIE's role in monitoring the quality of A.I.D.'s evaluation work is expected to increase. A.I.D.'s bi-annual evaluation reviews can play a role in this effort, but only if A.I.D. clarifies its expectations with respect to what evaluations will include and cover, and how their quality will be assessed.

RECOMMENDATION: That A.I.D. utilize the opportunity that bi-annual evaluation reviews provide to develop an adequate coverage and quality monitoring system.

- *Design a standardized evaluation review scope of work that has one section which deals with basic quality and coverage indicators that are to be measured in the same way on a longitudinal basis. Other sections can vary with each evaluation review.*
- *Draw upon the experience of past evaluation reviews in determining what is measurable using A.I.D. evaluations as the source of data.*
- *Pre-test any new system for assessing quality and coverage in an off year, e.g., with a sample of FY91 evaluations so that modifications can be made before the system must be used for a full FY91-92 evaluation review.*

13. Maintaining A High Quality Evaluation Library

The degree to which bi-annual evaluation reviews can accurately characterize the evaluation coverage of A.I.D.'s portfolio depends in good part on the quality of the evaluation library CDIE maintains on behalf of the Agency.

- **A.I.D.'s evaluation library and its automated information systems currently follow two filing practices which impede the conduct of bi-annual evaluation reviews and could impede the conduct of other quality control endeavors.**

As was first noted in A.I.D.'s FY87-88 evaluation review, the A.I.D. library often assigns different card catalogue numbers to evaluation reports and to the A.I.D. evaluation summaries that are intended to accompany them. A single card catalogue number is both adequate and appropriate.

A.I.D.'s automated information systems provides misleading information on evaluations when it reflects evaluation team deficiencies and errors with respect to project identification. Evaluations that are being entered into A.I.D.'s automated listings at times lack project numbers and at other times include misleading numbers, i.e., the number of the project that paid for the evaluation rather than the number of the project that was evaluated. This procedure encourages the incorrect assignment of evaluations to missions, bureaus and AID/Washington offices during evaluation reviews and would have the same impact on other quality-control activities.

RECOMMENDATION: That A.I.D. establish library and information system filing procedures that correct the two problems identified above.

-- *As a special, one-time effort, recode and recatalogue any FY91 evaluation documents which have been assigned multiple card catalogue numbers or inappropriate project numbers, so that future evaluation monitoring activities can be carried out in an orderly and efficient manner.*

ANNEXES

SCOPE OF WORK

A.I.D. EVALUATION SYNTHESIS: COVERAGE AND PERFORMANCE BACKGROUND

PPC/CDIE develops and issues Agency guidance on program and project evaluation, and also summarizes, synthesizes and disseminates lessons learned from development experience. In A.I.D.'s decentralized program management and evaluation system, most of the evaluation work, and the resulting evaluation reports (ERs), are generated by field Missions and some AID/W offices. CDIE is concerned with the coverage of these reports, their focus, quality and usefulness for a range of program management and decision-making needs.

Since 1982, CDIE has sponsored periodic reviews of all A.I.D. evaluation reports. These reviews have addressed such questions as the incidence of specific categories of findings, the quality of the reports, and the substantive analysis and summarization of findings and lessons learned contained in the reports.

As a continuation of this "evaluation synthesis" effort, and using approximately 350 reports submitted mainly during FY1989 and FY1990, CDIE seeks a review and analysis of several predefined elements that constitute important aspects of the coverage of the Agency's portfolio by these evaluations, the issues on which the evaluations focussed, and the treatment in the evaluation reports of three cross-cutting concerns. CDIE expects the results of this review to serve three major purposes:

- - Identification or clarification of areas where PPC may need to take further action in developing and issuing evaluation guidance;
- - Support for CDIE's ability to track changes in the coverage, quality, focus and usefulness of Agency-wide evaluation work;
- - Support for CDIE's ability to develop evaluation standards and models for future application.

ARTICLE I -- TITLE

"A.I.D. Evaluation Synthesis, 1989-1990: Coverage and Performance"

ARTICLE II -- OBJECTIVE

The objective of this delivery order is to provide PPC/CDIE a written report on the coverage and focus of approximately 350 evaluation reports submitted by A.I.D. units mainly during FY1989 and FY1990, relative to a set of pre-defined elements, together with the database and associated documentation from which the report and tabulated data are generated.

104

ARTICLE III -- STATEMENT OF WORK

The contractor will undertake and complete the following tasks:

A. Assemble and categorize evaluation reports

1. Based on PPC/CDIE/DI printouts, listings of evaluation contracts, actual evaluation reports, and discussions with bureau evaluation offices, assemble a list of evaluation reports completed during FY1989 and FY1990. With the assistance of CDIE, acquire copies of any reports missing from CDIE-DI's collection. Contractor will make arrangements necessary to transport reports from PPC/CDIE to contractor's place of business and to return these to PPC/CDIE upon completion of the work.

2. Refine a checklist of approximately 30 descriptive elements against which the contractor will review and process all evaluation reports and their associated A.I.D. Evaluation Summaries. The elements will form a database (dBase III + or other appropriate and CDIE-approved software) to be managed by the contractor during the performance of this work. Initial elements for the checklist are listed in the Annex to this Statement of Work. In preparing the checklist and constructing the database, the contractor will consult with CDIE regarding any further refinements or clarification of the elements as may be necessary prior to the final processing of the reports and entry of data.

3. Process evaluation reports and available Evaluation Summaries in accordance with the checklist into program categories and generate descriptive statistics.

B. Assess evaluation coverage of assistance portfolio

1. Using portfolio lists and values in A.I.D.'s Congressional Presentations and related tabulations of bilateral assistance program sectors and subsectors, assess the coverage of the portfolio represented by the relevant evaluation reports. This assessment will address coverage of individual country portfolios and coverage with respect to sectors and subsectors in the overall Agency portfolio.

2. This assessment will make particular note of the evaluation coverage of non-project assistance.

C. Determine the focus of evaluation reports

1. For no more than 350 reports, the contractor will assign and enter into the database data that describe the principal focus of each evaluation, based on information contained within the evaluation report. For this purpose, the contractor will develop and refine a typology of criteria or standards, that will serve to identify project, non-project and program evaluations that focus primarily on strategic, program and impact issues from those that focus primarily on narrower project management and implementation issues. This typology will also include criteria elements for assessing the extent to which gender, environmental and sustainability issues were treated in the evaluation reports.

105

2. In developing the typology, the contractor will refer to relevant frameworks developed by bureaus (e.g., program logframes, objective trees), as well as use in the evaluation reports of methodologies for cross-project and cross-program comparison relevant to the use of evaluation in strategic planning and program decision-making (e.g., cost-effectiveness, cost-benefit, relative impact, analysis of alternatives and prospective evaluation methods). A draft typology will be submitted to PPC/CDIE o/a three weeks following the signing of the contract. The draft and final typology will be subject to approval by PPC/CDIE.

3. The contractor will submit to PPC/CDIE o/a four weeks following the signing of the contract a report containing a preliminary selection of evaluation reports that meet criteria in the draft typology regarding a focus on program, strategy and impact issues.

4. For each evaluation report record in the database, the contractor will assign and enter into the database appropriate data on the criteria element, as derived from information contained in the evaluation report and its associated Evaluation Summary. Depending on the element, the data will consist of descriptive terms or characters, or numerical values, including scores that measure the degree to which the evaluation report meets the criteria.

Since almost all the evaluation reports were generated through A.I.D.'s decentralized evaluation system, contractor will recognize that the reports vary in terms of their focus, the specific questions addressed in each evaluation, the scope and depth of the analysis, and the methodology and data used to support each report's findings.

5. Contractor will submit a computer-generated report of data on the criteria elements for ten evaluation reports o/a ten weeks following the signing of the contract. This will constitute the interim report for this delivery order. Contractor will use this report as a means for clarifying and resolving with PPC/CDIE any remaining pre-tabulation issues or problems.

6. Following agreement between PPC/CDIE and the contractor on final report specifications, the contractor will develop report formats and programs as necessary to generate no more than 50 final summary tables that organize and tabulate data on all evaluation reports in terms of overall frequency distributions, percentages and other descriptive statistics, and in terms of bureaus, countries, sectors and subsectors corresponding to the evaluation reports.

D. Identify models of program evaluation and assess process aspects

1. From among the evaluation reports that most fully meet criteria of strategic and program evaluation, the contractor will select a sample of no more than 10 reports for further in-depth study.

2. Contractor will develop a protocol for studying the principal aspects of the process (e.g., development of SOWs, selection of evaluation methods, team selection and composition) that led to these 10 program and strategic evaluations. On the basis of information contained in the evaluation reports, scopes of work, and from personal and telephone interviews, the contractor will prepare an analysis of significant process aspects of these 10 reports. This analysis will be

incorporated into the final report for this delivery order, and will group these aspects into categories that are useful for the derivation of evaluation standards.

E. Prepare written report and oral presentation to A.I.D. staff

The contractor will prepare a written report on the results of the review and analysis. This report will include 1) assessment of coverage of assistance portfolio; 2) assessment of evaluation study focus; 3) assessment of models; and 4) assessment of evaluation processes.

The contractor will participate in a two-hour meeting during which the contractor will present the major findings of the review, and answer questions from A.I.D. staff regarding the report and its methodology.

ARTICLE IV -- REPORTS

As discussed above, the contractor will submit the following reports to PPC/CDIE:

- A. A preliminary report o/a four weeks after signing of this contract presenting an initial selection of evaluation reports meeting criteria regarding a focus on program, strategic and impact issues.
- B. An interim report o/a 10 weeks after signing of this contract, in five copies.
- C. Three verbal reports on progress submitted toward the end of each consecutive month following the signing of this contract.
- D. A final written report submitted to PPC/CDIE o/a the beginning of the fourth month following the signing of this contract. This report will be submitted first as a draft to the CDIE project officer. Following any changes required, the contractor will submit a final report in one unbound copy and 10 bound copies, together with the word processor disc used for the production of the final report.
- E. The database on diskette containing data on all evaluation report records together with relevant documentation (e.g., variable names or descriptions, decodes) developed to generate tables and other reports for this delivery order.

107