

# 2003 ANNUAL REPORT ON OPERATIONS EVALUATION



## 2003 ANNUAL REPORT ON OPERATIONS EVALUATION



**Deepa Chakrapani**

*Director-General, Operations Evaluation: Gregory K. Ingram*

*Director, Operations Evaluation Department: Ajay Chhibber*

*Manager, Corporate Evaluation and Methods: Victoria M. Elliott*

 OPERATIONS EVALUATION DEPARTMENT

 THE WORLD BANK

© 2003 The International Bank for Reconstruction and Development / The World Bank  
1818 H Street, NW  
Washington, DC 20433  
Telephone 202-473-1000  
Internet [www.worldbank.org](http://www.worldbank.org)  
E-mail [feedback@worldbank.org](mailto:feedback@worldbank.org)

All rights reserved.

1 2 3 4 05 04 03

The findings, interpretations, and conclusions expressed herein are those of the author(s) and do not necessarily reflect the views of the Board of Executive Directors of the World Bank or the governments they represent.

The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of the World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

#### **Rights and Permissions**

The material in this work is copyrighted. Copying and/or transmitting portions or all of this work without permission may be a violation of applicable law. The World Bank encourages dissemination of its work and will normally grant permission promptly.

For permission to photocopy or reprint any part of this work, please send a request with complete information to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA, telephone 978-750-8400, fax 978-750-4470, [www.copyright.com](http://www.copyright.com).

All other queries on rights and licenses, including subsidiary rights, should be addressed to the Office of the Publisher, World Bank, 1818 H Street NW, Washington, DC 20433, USA, fax 202-522-2422, e-mail [pubrights@worldbank.org](mailto:pubrights@worldbank.org).

# Contents

<b>FOREWORD</b>	<b>v</b>
<b>ABBREVIATIONS AND ACRONYMS</b>	<b>ix</b>
<b>1. INTRODUCTION</b>	<b>1</b>
<b>2. MONITORING OF BANK OPERATIONS</b>	<b>3</b>
Instrument Level	4
Sector Level	9
Country Level	11
Summary	12
<b>3. EVALUATION OF BANK OPERATIONS</b>	<b>15</b>
Instruments	15
Sector Level	19
Country Level	20
Grants and Global Programs	20
Knowledge Initiatives: Knowledge Sharing, Research, and Training	21
Summary	23
<b>4. CORPORATE PRIORITIES IN MONITORING AND EVALUATION</b>	<b>25</b>
Helping Client Countries Build Results Orientation	25
Results Orientation in the Bank	26
<b>5. SELF-EVALUATION IN OED</b>	<b>29</b>
Relationship between OED and the Committee on Development Effectiveness	29
Recommendations from Previous AROEs	29
OED Surveys: PPARs and Sector and Thematic Studies Are the Focus This Year	30
Geographic Focus of OED Evaluations	31
<b>6. CONCLUSIONS AND RECOMMENDATIONS</b>	<b>33</b>
Conclusions	33
Recommendations	34

## **ANNEXES**

Annex A: Monitoring and Evaluation Framework at the Country Level	35
Annex B: Monitoring and Evaluation Units in the Bank	37
Annex C: OED Review of Rating Changes and Quality of Implementation Completion Reports	39
Annex D: Results of OED Self-Evaluation Surveys of Clients	43
Annex E: Glossary of Terms	51
Annex F: Management Response	55
Annex G: Chairman’s Statement, Committee on Development Effectiveness (CODE)	61

## **ENDNOTES**

**65**

## **BIBLIOGRAPHY**

**69**

## **BOXES**

Box 1.1: Monitoring and Evaluation: What Do They Mean?	2
Box 2.1: OED’s Suggestions for Enhancing the Evaluability of ESW/AAA	7
Box 2.2: QAG’s Country-Level Approach to Assessing ESW Quality	8
Box 2.3: Some Limitations of the MDGs as Monitorable Targets	10
Box 2.4: OED’s Recommendations on Monitoring Sector Progress	11
Box 2.5: Results-focused CASs Designed to Provide an M&E Framework	12
Box 2.6: A New and Comprehensive Approach to Country Portfolio Reviews	13
Box 3.1: Integrating Trust Fund M&E with Bank Operations	19
Box 4.1: Two-Tiered IDA Results Measurement System	27
Box 4.2: OED Suggestions on the Elements of an Effective Corporate Report on Results	28
Box 5.1: OED’s Independence is Assured Through its Mandate and Organizational Structure	30
Box 5.2: Forthcoming OED Retrospective on Country Assistance Evaluations	31
Box 5.3: Using OED Web sites to Enhance OED’s Impact	31

## **FIGURES**

Figure 2.1: Weak Evaluability of Lending Operations	4
Figure 2.2: Deterioration in the Quality Assurance Process during Project Design	5
Figure 2.3: Deterioration in the Quality of Reporting	6
Figure 3.1: Definition of Outcomes and the Quality of Arrangements for Evaluation Vary	17

## **TABLES**

Table 2.1: Framework for Monitoring in Bank Operations	3
Table 3.1: Framework for Evaluation in Bank Operations	16
Table 4.1: M&E Framework at the Corporate Level	26
Table 5.1: Geographic Focus of OED Evaluations (FY99–02)	32

## Foreword

There is increasing international pressure on the development community to demonstrate value added and to strengthen the results focus of aid management. In the last fiscal year, the Implementation Forum has given new impetus to the need for measuring and demonstrating results in the Bank. The *Better Measuring, Monitoring, and Managing for Development Results* initiative, launched in September 2002, aims to establish a framework for increasing the results orientation in both the Bank and client countries.

This *Annual Report on Operations Evaluation* recognizes the increased Bank emphasis on the results and effectiveness of its development activities, and its efforts to build results orientation in client countries in the past year. It takes stock of the Bank's frameworks both for monitoring and for evaluation, comments on their effectiveness, and highlights areas for improvement and greater management attention.

### MONITORING AND EVALUATION OF BANK OPERATIONS

Although monitoring and evaluation are usually discussed in tandem, they serve distinct and complementary functions. Monitoring facilitates managerial decisionmaking and determines whether adequate implementation progress has been made to achieve outcomes. Evaluation provides attribution and causality. It determines the extent to which outcomes have been achieved and helps demonstrate results.

This report raises three broad areas for Bank improvement: the need to enhance outcome orienta-

tion and evaluability of activities in the design stages; the need to strengthen the quality and realism of reporting on implementation progress, based on the performance indicators set out during the design stages; and the need to close gaps in self- and independent evaluation.

#### *Monitoring*

One clear institution-wide priority is to establish clear and realistic outcomes and to develop monitorable indicators, baseline data, and benchmarks of progress. Another imperative is to enhance the framework for monitoring and reporting on implementation progress against the specified outcomes. The Bank's quality assurance process should widen its focus to include greater outcome orientation and evaluability.

Management has a well-established monitoring framework for its projects, but the emphasis on setting clear outcomes and supporting monitorable indicators needs to be enhanced. Country strategy and project design processes and documents should

systematically set outcome objectives for the Bank, explicitly link Bank intervention to expected outcomes, and clearly define performance indicators. Progress reports should use these performance indicators to report on and to assess implementation progress. The newly piloted results-based Country Assistance Strategy aims to strengthen outcome orientation of Bank country-level assistance.

The outcome orientation of the Bank's nonlending operations, knowledge sharing activities, and sector strategies needs strengthening. Monitoring of implementation progress against outcomes is weak for grants and global programs, trust funds, sector strategies, and knowledge sharing initiatives. Management has initiated reforms to enhance monitoring of trust funds and global programs, but it is too early to comment on their effectiveness, and that of sector strategies is evolving.

### *Evaluation*

The framework for self-evaluation by management, and independent evaluation by OED, is strongest at the project level. At the country level, the framework for independent evaluation, comprising OED Country Assistance Evaluations, is well established. The recent introduction of CAS Completion Reports responds to a longstanding OED recommendation and closes gaps in self-evaluation at the country level. But there remain other gaps in both self- and independent evaluation. Self-evaluation of sector strategies, nonlending operations, trust funds, and knowledge initiatives is weak. New initiatives are under way to improve self-evaluation of grants and global programs, though it is too early to comment on their effectiveness. Independent evaluation of sector strategies and global programs needs strengthening.

### *Corporate Level*

Management expects to report on Bank performance at the corporate level primarily by developing the *Annual Review of Portfolio Performance* into an *Operational Performance and Results Review* (OPR). The OPR will need to evolve over time to become a com-

prehensive report on the results of all Bank activities, including corporate-level initiatives such as global programs, capacity building, and knowledge sharing. Initially, the OPR will be based on limited information on country-level outcomes, pending mainstreaming of the results-based Country Assistance Strategy initiative and other efforts to enhance outcome orientation of lending and nonlending activities. Management has also developed the International Development Association (IDA) results measurement framework to report to donors on progress on important development outcomes in IDA-eligible countries.

## **BUILDING MONITORING AND EVALUATION CAPACITY IN CLIENT COUNTRIES**

Bank and borrower monitoring and evaluation (M&E) are closely related: the Bank relies on country data for measuring the performance of its interventions at the project, sector, and country levels. But limitations in country M&E—data, systems, and capable evaluators—have encouraged the Bank and other donors to take an enclave approach to M&E, with little emphasis on building sustainable country capacities and systems. This has resulted in a multiplicity of donor M&E burdens on borrowers (CDF Secretariat 2003). The Bank has been active in helping borrowers strengthen their results orientation and evaluation capacity in order to enhance government performance. Yet the Bank has made limited progress in mainstreaming evaluation capacity building in its operations.

### *Self-Evaluation in OED*

OED's annual self-evaluation this year surveyed one relevant audience—the evaluatees—on their reactions to two OED products: sector and thematic studies and Project Performance Assessment Reports (PPARs). The results of the survey suggest that OED should clarify and better present its methodology, including its evidentiary base, for sector and thematic studies. It should also better target and disseminate

its sector and thematic and project evaluations to Bank staff, and enhance consultation with borrowers. This year, OED has also complemented its annual survey with detailed reviews in two areas of its work: a CAE retrospective and a review to enhance the use of the Web site as a tool to serve OED's business strategy. These reviews are currently under way.

Although OED has largely implemented the recommendations that evolved from the 2002 AROE, there is a need for more effective internal dissemination of OED products through enhanced targeting and active participation in Regional and sector events. As OED moves to rate CAS Completion Reports, it should ensure that there are more transparent procedures to resolve differences in project and country ratings between management and OED. A review of the geographical distribution of OED evaluations over the last four fiscal years suggests that the country distribution of OED's work is largely consistent with patterns of operational activities, as measured by number of exiting projects and share of lending.

## RECOMMENDATIONS TO MANAGEMENT AND TO OED

### *Management*

- (a) Issue an Operational Policy/Bank Procedure that sets out the mandate, framework, roles, and responsibilities for self- and independent evaluation.

- (b) Mainstream greater outcome orientation in CASs and a self-evaluation framework for Bank assistance at the country level.
- (c) Strengthen the M&E of the development outcomes for activities financed by trust funds.
- (d) Strengthen outcome orientation and evaluation of nonlending activities, at the aggregate level and at the task level (either individually or in clusters).

### *OED*

- (a) Enhance borrower consultations on sector and thematic studies and PPARs.
- (b) Improve internal dissemination of sector and thematic studies through better targeting and more active participation in sector and Regional events.
- (c) Institute a transparent procedure to resolve differences in ratings between OED and management for country and project-level evaluations.



Gregory K. Ingram  
*Director-General, Operations Evaluation*





## Abbreviations and Acronyms

AAA	Analytical and advisory service
ACS	Activity Completion Summary
AIS	Activity Initiation Sheet
ACTTF	Accounting Trust Funds Division
AFR	Africa Region
APL	Adaptable Program Loan
ARCS	Audit Reports Compliance System
ARDE	Annual Review of Development Effectiveness
AREC	Annual Report on Evaluation Capacity Development
AROE	Annual Report on Operations Evaluation
ARPP	Annual Review of Portfolio Performance
BP	Bank Procedures
CAE	Country Assistance Evaluation
CAS	Country Assistance Strategy
CASCR	Country Assistance Strategy Completion Report
CASPR	Country Assistance Strategy Progress Report
CDF	Comprehensive Development Framework
CES	Country Evaluation Summary
CGIAR	Consultative Group on International Agricultural Research
CODE	Committee on Development Effectiveness
CPPR	Country Portfolio Performance Review
CPR	Country Portfolio Review
CRS	Client reporting system
DANIDA	Danish International Development Agency
DEC	Development Economics Vice Presidency
DECRA	DEC Research Advisory Unit
DGF	Development Grant Facility
DGO	Director-General, Operations Evaluation
EAP	East Asia and Pacific Region
ECA	Eastern Europe and Central Asia Region
ECD	Evaluation capacity development
ES	Evaluation summary
ESW	Economic and sector work
FSAP	Financial Sector Assessment Program
FSE	Financial Sector Network

GCR	Grant Completion Report
HDN	Human Development Network
HIPC	Heavily Indebted Poor Countries
IBTF	Initiating Brief for a Trust Fund
ICM	Implementation Completion Memorandum
ICR	Implementation Completion Report
IDA	International Development Association
IDF	Institutional Development Fund
IFC	International Finance Corporation
IPR	Implementation Progress Report
JSA	Joint Staff Assessment
LEG	Legal Department
LIL	Learning and Innovation Loan
MAR	Management Action Record
MDG	Millennium Development Goals
MNA	Middle East and North Africa Region
MRP	Memorandum and Recommendation of the President
MTR	Mid-Term Review
OED	Operations Evaluation Department
OPCS	Operations Policy and Country Services
PAD	Project Appraisal Document
PIP	Project Implementation Plan
PPAR	Project Performance Assessment Report
PREM	Poverty Reduction and Economic Management Network
PRSP	Poverty Reduction Strategy Paper
PSI	Private Sector and Infrastructure Network
PSR	Project Supervision Report
QACU	Quality Assurance and Compliance Unit
QAE	Quality at entry
QAG	Quality Assurance Group
QER	Quality Enhancement Review
QSA	Quality of Supervision Assessment
RSB	Research support budget
S&T	Sector and thematic
SAP	Systems, Applications, and Products in Data Processing
SAR	South Asia Region
SD	Sector director
SPIR	Sector Policy and Implementation Review
SRM	Strategy and Resource Management
SSP	Sector Strategy Paper
TFO	Trust Fund Operations (Department)
TM	Task manager
TQC	Trust Funds Quality Assurance and Compliance Unit
VPU	Vice-presidential unit
WBI	World Bank Institute
WBIEG	World Bank Institute Evaluation Group

## 1. Introduction



The *Annual Report on Operations Evaluation (AROE)* fulfills the OED mandate to assess monitoring and evaluation (M&E) of development effectiveness of World Bank activities. It focuses on the status, effectiveness, and progress in improving the M&E framework for the Bank. It does not report on evaluation findings, which are reported in the *Annual Review of Development Effectiveness*.

The 2002 AROE took a 10-year retrospective view of the evolution of controls and M&E at the Bank. It noted the Bank's progress since 1992 in establishing greater staff accountability, focusing on portfolio implementation, and intensifying managerial and Board oversight of development effectiveness. The Bank has implemented a risk assessment and control framework; established and strengthened fiduciary policies and compliance; and created a professional fraud and investigation unit, a Quality Assurance and Safeguard Compliance Unit (QACU), and a Quality Assurance Group (QAG) to track risks and quality in real time. Simultaneously, OED's 1997 Renewal Strategy has increased the scope and role of independent evaluation.

Several recent developments have heightened the pressure on countries, the Bank, and donor agencies to demonstrate the results of development assistance. The Comprehensive Development Framework, the Heavily Indebted Poor Countries initiative, the IDA 13 Replenishment Agreement, and the Poverty Reduction Strategy Papers all represent efforts to associate development assistance flows more directly with demonstrable results. The need to demonstrate value added and strengthen the results focus of aid management was a major stimulus of

the development consensus reached at the Monterrey Conference on Financing for Development in March 2002. Since then, the Johannesburg and Doha Conferences have reaffirmed the commitment of development agencies—including the Bank—to demonstrate and disseminate results.

In the Bank, the 2002 Strategic Forum (now the Implementation Forum) gave new impetus to the need for measuring and demonstrating results. In March 2002, a vice-president and senior adviser to the managing directors were tasked to ensure a coordinated approach for the Bank's results work. In September 2002, the Bank launched the *Better Measuring, Monitoring, and Managing for Development Results* (or *Managing for Results*) initiative, which aims to establish a framework for increasing the results orientation both in the Bank and in client countries. An implementation plan for the Managing for Results initiative was presented to the Committee on Development Effectiveness (CODE) of the Bank's Board in December 2002. A Results Secretariat was created in February 2003 in Operations Policy and Country Services (OPCS). While the development community and clients together seek to achieve the Millennium Development Goals (MDGs) and other targets of development success, the challenge for Bank

management is to translate these goals into a set of outcomes, along with monitorable indicators, benchmarks of progress, and adequate baselines for demonstrating improvements in its own performance.

Although monitoring and evaluation are usually discussed in tandem, they serve distinct and complementary functions, as noted in box 1.1. Monitoring determines whether adequate progress on implementation is being made to achieve outcomes. It

facilitates managerial decisionmaking by providing information on adjustments that may be needed to the implementation process (Kusek and Rist 2001). Evaluation provides attribution and causality, and serves as a basis for accountability and learning.

This year's AROE recognizes the increased corporate emphasis on M&E in the Bank since last year, and it focuses specifically on areas that require greater improvements and management attention.<sup>1</sup> It thereby responds to the Board's view in July 2002 that the AROE could be more useful if it were to focus specifically on areas where progress was most needed, such as lagging areas in M&E and the corporate scorecard, partnerships, and the monitoring of trust funds. Management has acknowledged some of these weaknesses or gaps.

This AROE takes stock of the Bank's frameworks for monitoring in Chapter 2 and for evaluation in Chapter 3, commenting on their effectiveness and pinpointing areas for improvement. Chapter 4 focuses on M&E at the corporate level.

The report draws on reports from Operations Policy and Country Services (OPCS) and the Quality Assurance Group (QAG); ongoing and completed OED evaluations; and interviews with staff in Regional front offices, network anchors, the Development Economics Vice-Presidency (DEC), and the World Bank Institute (WBI).

### **Box 1.1: Monitoring and Evaluation: What Do They Mean?**

*Monitoring* is the regular and continuous tracking of inputs, outputs, outcomes, and impacts of development activities against stated results. It can be undertaken at any level of operation: project, program, sector, country, or global (OED 2002b). Monitoring involves establishing (a) a set of outcome indicators along with monitorable targets, benchmarks of progress, and baselines, as well as data collection, analysis, and reporting guidelines; (b) quality assurance processes, before and during implementation; and (c) concurrent reviews of implementation progress. The purpose of monitoring is to provide management decisionmaking with information to enhance implementation.

*Evaluation* is the systematic and objective measurement of the extent to which operational programs and activities produce desired results. Unlike monitoring, evaluation attempts to establish causality and attribution (OED 2002b; World Bank 1998). It serves as the basis of accountability and learning by staff, management, and clients. Evaluative information is used for the formation of new directions, policies, and procedures. Ex-post evaluation in the Bank, undertaken on or after activity completion, involves three aspects: completion reporting by activity heads, internal or external reviews conducted by management (self-evaluation), and independent evaluation undertaken for the Board or other governing bodies.

## 2. Monitoring of Bank Operations



This chapter describes the Bank's monitoring framework at the instrument, sector, and country levels (summarized in table 2.1).<sup>1</sup> Each of these monitoring processes constitutes an important building block for the institution-wide results agenda. The chapter comments on the effectiveness of specific aspects of the framework, and highlights recent improvements and unresolved areas for management attention.

**Table 2.1: Framework for Monitoring in Bank Operations**

Type of intervention or financing instrument <sup>a</sup>	Defining outcomes and monitorable indicators	Review and quality assurance	Tracking and reporting on implementation progress	Managing implementation progress
Lending operations	Project Appraisal Document (PAD) or Memorandum & Recommendation of the President	Management and peer reviews, Quality Enhancement Reviews (QER), and annual QAG review of quality at entry <sup>b</sup>	Project Implementation Plan (PIP), Project Supervision Report (PSR)	Mid-Term Review, Country Portfolio Performance Reviews (CPPR), annual QAG reviews of supervision (QSA)
ESW/AAA	Concept Paper	Peer and management review, QER (few), QAG ESW reviews	SAP monitoring	Peer review, management review, QAG ESW reviews
Trust funds <sup>c</sup>	Initiating Brief for trust funds	VPU, TFO review	Status reports	
Sector strategy	Sector Strategy Paper (SSP), SSP Updates	QAG, network, and OPCS review	Implementation Progress Reports (IPR)	Sector Board (SB) reviews
Country program <sup>d</sup>	Country Assistance Strategy (CAS)	SB, network, and OPCS review	CAS Progress Report (CASPR)	CPPR, management reviews
Grants/global programs	Grant proposal	Task manager review	Financial reporting	Task manager review

*Note:* ESW = economic and sector work; SAP = Systems, Applications, and Products in Data Processing; TFO = Trust Fund Operations.

a. Grants and Global Programs and Knowledge Initiatives are discussed in Chapter 3. One other means of financing not discussed here is guarantees.

b. QAG's quality at entry assessment is undertaken after project approval.

c. Trust funds are a source of funding that is used to supplement Bank financing of various Bank and recipient development activities.

d. Country program support includes CAS, PRSP, JSA, HIPC and other development services.

## INSTRUMENT LEVEL

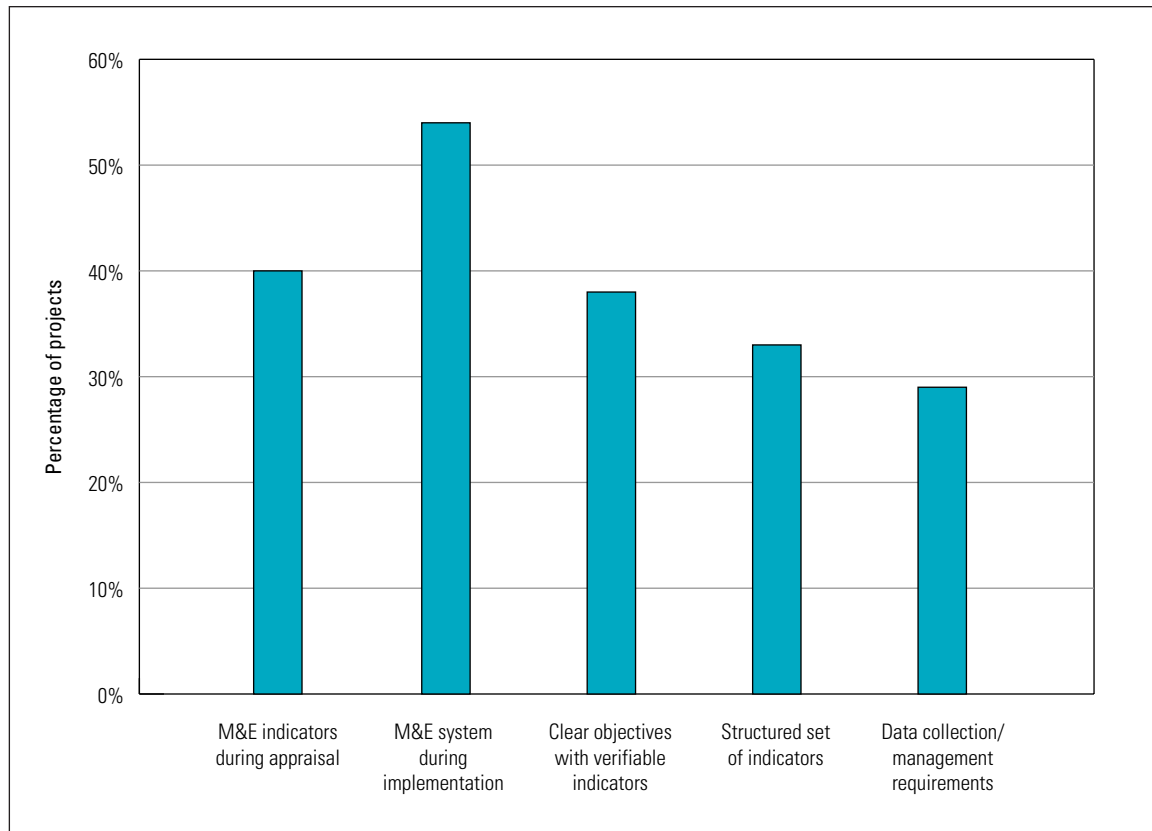
### *Lending Operations (Investment and Adjustment Projects)*

OED has reported on improvements in project outcome ratings over the last two years, and QAG has reported on improvements in overall quality at entry since FY99. The framework for monitoring investment projects is well established, but there remain two areas of weakness: establishing outcome indicators and monitoring and reporting on implementation progress. The quality assurance processes at entry and implementation should place greater emphasis on the evaluability and design of appropriate monitoring systems. Monitoring of adjustment loans against the specified outcomes needs to be strengthened, and supervision reporting on all projects should be improved.<sup>2</sup> Since lending operations are the main tools for affecting the outcome of the Bank's assistance strategy, the absence of monitorable outcome indicators and targets at the

project level and lack of adequate supervision reporting could undermine the country-level results initiative.

**Setting clear outcomes and monitorable indicators:** The Project Appraisal Document (PAD) for investment loans and the Memorandum and Recommendation of the President (MRP) for adjustment loans develop the rationale for the proposed investment. The PAD uses a log frame to specify outcomes, intermediate indicators, and monitorable targets. Despite the improvement in overall quality at entry of projects since FY99, QAG assessments have noted shortcomings in setting interim benchmarks, confusion among staff between output and outcome indicators, and the lack of baseline data against which to gauge change, particularly on structural and sector policy reforms (QAG 2003a, p. 12). For adjustment loans, there is no explicit requirement to set outcomes; the emphasis is on tracking actions specified in the policy matrix (QAG 2003a, p. 12; World Bank 2001b, pp. xvi, 88). A new policy on adjustment lend-

**Figure 2.1: Weak Evaluability of Lending Operations**



Source: OED data.

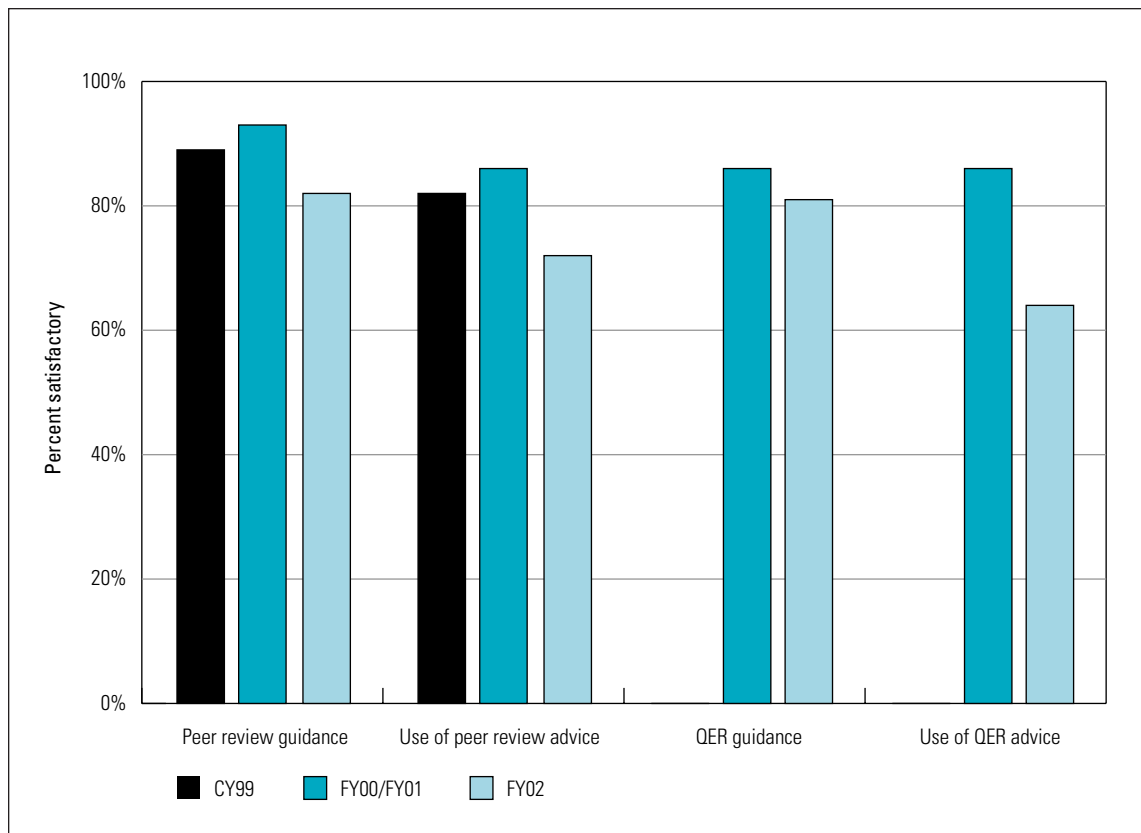
ing, to be issued by the end of calendar year 2003, is expected to include specific provisions to strengthen the articulation of development objectives as desired outcomes and to assess fulfillment of conditions in the context of achieving these outcome objectives.

**Monitoring has been weak:** A small proportion of projects build good M&E into their design. Only 42 percent of projects assessed by QAG in FY02 had appropriate monitoring systems. Consistent with this finding, OED data indicates that 40 percent of evaluated projects between July 1997 and March 2003 incorporated M&E indicators at the appraisal stage. OED also found that 54 percent of projects included an M&E system for the implementation phase, 38 percent had clear project and component objectives verifiable by indicators, 33 percent had a structured set of indicators, and 29 percent had requirements for data collection and management (figure 2.1). To help redress this weakness in design, DEC is planning to assist operational staff in setting up baseline data for a few pilot projects in order to

facilitate periodic impact evaluation. This is expected to enhance the evaluability of project outcomes.

**Quality assurance at entry:** QAG has reported improvements in overall quality at entry from FY99 onward. Although QAG's quality at entry process applies only to a sample of 100 projects each year, its findings provide demonstration effects and offer opportunities for learning. The peer review process and Quality Enhancement Reviews, along with management review meetings, are two important tools for quality assurance during project design. Although these tools have increased quality, task teams appear to be reducing their use of the advice, as shown in figure 2.2 (QAG data). A recent internal management survey of task team leaders notes that 73 percent of respondents were dissatisfied with peer review processes, and 69 percent with Regional quality assurance arrangements.<sup>3</sup> The peer review and Quality Enhancement Review (QER) process should work more effectively to ensure greater clarity of outcome objectives and performance indicators at the time of appraisal.

**Figure 2.2: Deterioration in the Quality Assurance Process during Project Design**



Source: QAG data.



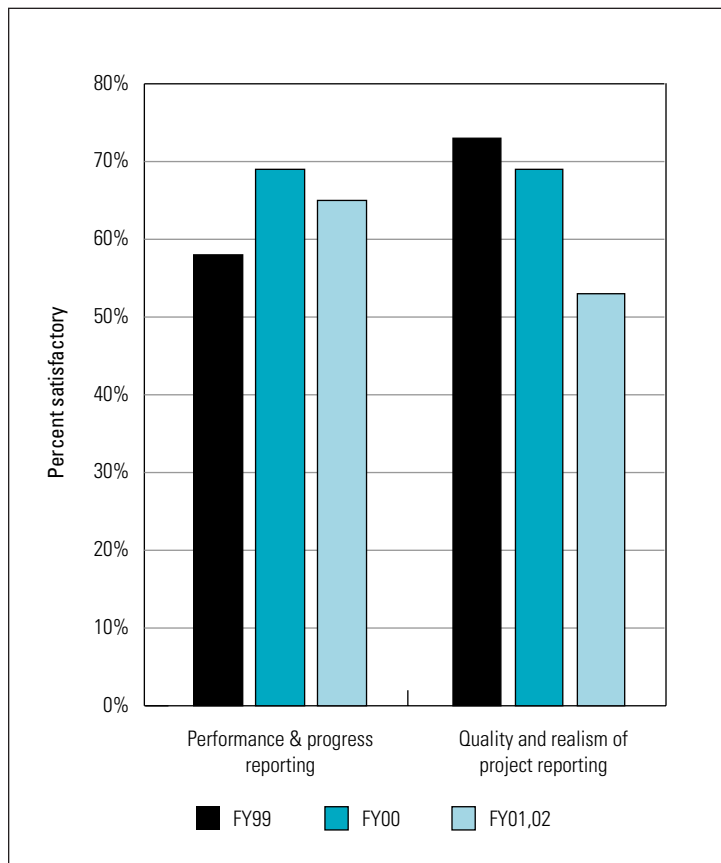
**Tracking and reporting on implementation progress:** For investment loans, the Project Implementation Plan (PIP), prepared by the borrower, was expected to provide the framework for monitoring implementation progress.<sup>4</sup> The Bank uses Project Supervision Reports (PSR) and supporting documentation (aide-memoires and back-to-office reports) for monitoring and reporting on implementation progress that should be based on the updated PIP. But neither the PIP nor the PSR is fully effective. QAG reports suggest that the PIP is little used and is seldom updated, and questions its usefulness. As a result, management has made the PIP optional for projects approved on or after July 1, 2003. QAG also reports weaknesses in *performance and progress reporting during supervision* and deterioration in the *quality and realism of project reporting during implementation* (figure 2.3). This finding is supported by OED data. Although 70 percent of

projects evaluated between 1997 and 2003 reported on project implementation progress and problems, only 36 percent used performance indicators. In projects that involve extractive industries, for instance, a recent OED review highlighted the paucity of information on implementation of safeguards in PSRs and other monitoring documentation. This weakness in reporting makes it difficult for the Bank to monitor the implementation of and compliance with safeguard policies (OED 2003e). For adjustment loans, QAG assessments of the quality of supervision normally rely on policy matrices because interim indicators and benchmarks for measuring progress toward outcome goals are not required and often not provided (QAG 2003a).

The PSR is not aligned with the borrowers' monitoring framework. Most government systems are ill-adapted to take on demanding monitoring requirements. This has led donors to establish enclave M&E systems, separate from government processes, that do not build public sector capacity, and are therefore unsustainable (CDF Secretariat 2003). Borrowers view donors' proposals on monitoring as elaborate and complex, and the indicators as unwieldy, difficult to substantiate, and conforming more to donors' reporting requirements than to what the country needs to manage national service delivery. Ideally, the Bank and other donors would rely on country-led data and monitoring systems for tracking implementation progress. The PRSP advocates this approach, and calls for strengthening borrower systems for data collection, where necessary. Yet OED data (July 1997–April 2003) suggest that only 23 percent of all evaluated projects included institutional arrangements for capacity building for M&E and received feedback from the project-designed M&E system. Only 32 percent of evaluated projects used borrower-provided M&E data from the system designed at project appraisal.<sup>5</sup>

**Managing implementation progress:** Country and sector managers carry out biannual reviews of their portfolios of lending operations through Country Portfolio Performance Reviews (CPPR) and project Mid-Term Reviews (MTR). QAG rates the quality of MTRs to be 84 percent satisfactory or better. But recent QAG evidence suggests that task teams

**Figure 2.3: Deterioration in the Quality of Reporting**



Source: QAG data.

have expressed dissatisfaction with the quality of guidance during implementation.<sup>6</sup> This has led to a recently initiated management effort to review and enhance guidance to task teams working on investment loans.

**Low quality of supporting documentation:** This has been a persistent problem in reporting, especially for the PSR. QAG assessments of supervision suggest that the PSR is not relevant to operational staff needs, and that there is little incentive to use it effectively. They suggest that improvements in progress monitoring and reporting would not require fundamental changes to the PSR format or to the supporting information technology architecture, but rather stronger incentives and managerial emphasis on high-quality and candid reporting throughout the project cycle. Management is simplifying PAD templates and the log-frame, customizing sections in the PAD for instruments such as Adaptable Program Loans and Learning and Innovation Loans, and introducing project concept notes and reviews earlier in the project cycle.<sup>7</sup> These initiatives will streamline investment-lending processes, but will not address the weaknesses noted above in monitoring outcomes of adjustment loans and supervision reporting.<sup>8</sup>

### *Nonlending Instruments*

There has been an improving trend in the overall intellectual quality of economic and sector work (ESW).<sup>9</sup> The monitoring framework for ESW tasks remains relatively weak. Measurable outcomes and performance indicators are rarely set, even for large tasks. Quality assurance mechanisms can be used more effectively to foster greater clarity on outcomes in the design stages, and thereby enhance the evaluability of ESW.

ESW and analytical and advisory services (AAA) are key instruments for supporting the Bank's country strategies, policy dialogue, and project design. Nonlending products range from country diagnostic and advisory reports, to process tasks, informal policy notes, and events such as workshops and conferences. The Bank has undertaken two phases of reform under the Fixing ESW program that have

helped to define and establish a monitoring system for ESW, but the outcome orientation for ESW tasks and related monitoring remains weak.<sup>10</sup>

**Setting clear outcomes and monitorable indicators:** Regional guidelines indicate that the Concept Paper is expected to set out the outcomes and objectives of every ESW task that costs over \$50,000.<sup>11</sup> Unlike the PAD/MRP in lending, Concept Papers for ESW lack specific development objectives or outcomes, measures of success, or interim indicators or benchmarks (QAG 2003b) (box 2.1). The recent QAG assessment of ESW tasks relied on surveys of task team leaders rather than Concept Papers to

#### **Box 2.1: OED's Suggestions for Enhancing the Evaluability of ESW/AAA**

The Concept Paper for individual or clusters of nonlending tasks should specify the audience, outcomes being sought, and the most relevant outcome indicators that may be used to judge progress and achievements. A task may have one or more of the following outcomes:

- Respond to specific client requests for information.
- Provide policy advice and influence policymaking.
- Diagnose the nature of the problem.
- Determine the needed sequence of reforms.
- Inform policymakers on cross-country experience, sequencing, or implementation of reforms.
- Facilitate consensus building among policymakers and stakeholders.
- Facilitate specific lending operations or programs.
- Remain engaged in a country that is not an active borrower or has a poor policy environment.

The desired outcomes should determine the nature, number, and timing of deliverables as well as the leading and intermediate indicators for determining progress and achievements.

A monitoring framework for nonlending activities is more important than ever as management seeks to place all AAA tasks in a comprehensive, country-level perspective to determine their achievements and contributions to country-level outcomes. Concept Papers should establish outcomes and related indicators for individual ESW tasks or for clusters of tasks.

identify development objectives and corresponding success indicators, since in ESW task planning, development objectives are typically not well specified. This lack of outcome objectives and related indicators undermines the evaluability of tasks on completion. Setting of outcome objectives is even less common for the small tasks that accounted for 40 percent of all ESW deliveries in FY02: only two Regions (Middle East and North Africa and South Asia) require concept notes or reviews for all small tasks.

**Quality assurance at entry:** Quality assurance for nonlending tasks consists of peer and internal reviews by management during design and decision draft/event stages. Each Region has guidelines on process requirements for small and large nonlending tasks, including quality assurance. In FY02, QAG noted that peer reviews are rigorous and effective during task entry (QAG 2003b). The peer review process could also help to foster greater clarity in the Concept Paper on task outcomes and related indicators to enhance their evaluability.

**Monitoring and reporting on implementation progress:** The monitoring of ESW involves two aspects: (a) monitoring of outputs, time, and costs through the SAP system, and (b) real time monitoring by country and sector units. All Regions and a few sector units systematically monitor SAP data for large ESW tasks. QAG reviews have reported inaccuracies and inconsistencies in SAP reporting. For real time monitoring, management needs to track progress against the objectives and outcome indicators set out in the Concept Paper. QAG has also advocated creating in-country advisory panels and introducing annual reviews of performance of nonlending tasks into the Country Portfolio and Performance Review (CPPR) process. In November 2002, management proposed a new, flag-based system for monitoring the ESW portfolio to provide timely signals to managers about tasks at risk. The indicators for determining risk are being fine-tuned, and the system has not yet been introduced.

QAG has proposed an assessment of randomly selected ESW tasks with a country-focused approach, summarized in box 2.2. This approach will require the Country Assistance Strategy to articulate

### **Box 2.2: QAG's Country-level Approach to Assessing ESW Quality**

QAG is piloting a program to assess the quality and influence of ESW at the country level. In this pilot, QAG chose a selected, representative sample of all ESW tasks during the Country Assistance Strategy period. The assessment reviews quality of individual tasks, uses borrower inputs to assess likely impact, and considers the task in light of the proposed CAS objectives and its contribution to the country program. QAG has undertaken three such country assessments and rated the overall AAA program in all three countries as satisfactory. QAG expects to provide an aggregate report of its country-level findings to the Bank's Board in FY04.

interim benchmarks and indicators of success for the ESW program. The new approach can enhance the results focus at the country level and should ensure comprehensive coverage of all ESW tasks—small and large, formal and informal, either individually or in clusters.<sup>12</sup> But as proposed, the QAG review will remain focused on the quality of inputs rather than on outcomes or achievements of ESW activities.<sup>13</sup>

### ***Trust Funds***

Management has initiated several reforms to enhance monitoring and reporting on the use of trust funds, as well as on their development contribution. The reforms to enhance reporting on development gains from trust funds are in their infancy, and it is too early to comment on their effectiveness. Performance indicators for trust fund-assisted programs should go beyond the measurement of inputs to capture program outputs and outcomes. OED recommends that the Bank link monitoring and reporting on the use and performance of trust funds with the development activities they support.

Trust funds are a financing mechanism that supports a range of Bank operations, including research, global programs, knowledge services, and the design and implementation of nonlending and lending tasks. The issues in monitoring and evaluation of these activities are equally relevant, even when these activities are financed by trust funds. At the same time, in 2001 the Bank initiated a reform process to strengthen the monitoring, reporting, and evaluation of the development gains specifically attributable to trust funds. Operational Policy/Bank Procedure 14.40, which guides the Bank's trust fund activities, is being revised. The revised Operational Policy sets out the policy framework, including arrangements for monitoring, reporting, document retention, and evaluation. The Trust Fund Operations (TFO) unit is now responsible for policy development, clearances, and allocation of new funds. It supports reporting, dissemination, and oversight. While the primary responsibility for control and ensuring appropriate use of trust funds rests with task teams and relevant vice-presidential units, the Bank has set up a Trust Fund Quality Assurance and Compliance Unit (TQC) to verify the effectiveness of Regional quality control procedures and oversee compliance with trust fund policies and donor agreements.

**Setting clear outcomes and monitorable indicators and quality assurance at entry:** The Initiating Brief for a Trust Fund (IBTF) clarifies the objectives and outcomes of a trust fund. An FY02 review of trust fund activities indicates that the IBTF should incorporate relevant monitoring indicators at the time of approval, and the review process should focus on enhancing evaluability.<sup>14</sup>

**Tracking and reporting on implementation:** Management has initiated several improvements to enhance the monitoring and reporting of development gains for trust fund activities in order to ensure comprehensive reviews of the portfolio. The initiatives include tool kits and learning programs for task team leaders and managers, an accreditation program on trust fund management, and SAP-based status (or supervision) reports for major trust fund programs such as Policy and Human Resources Development, the Japan Social Development Fund, and the Netherlands–World Bank Partnerships Program.<sup>15</sup> These ini-

tiatives are too recent to have fully taken effect. The learning and accreditation pilot program was effective, and now all staff responsible for using trust funds must receive accreditation.<sup>16</sup> There appear to be several inadequacies in SAP-based data and information (financial, performance reporting) on trust funds.<sup>17</sup> Systems for recipient-executed trust funds need strengthening. Given the lack of monitorable indicators in the IBTF, the inadequacies in SAP reporting, and the few standard performance measures for trust fund programs, vice-presidential units (VPUs) still do not have the ability to properly monitor the use of trust funds and their impact on Bank activities. In FY02, TQC recommended that management develop a systematic classification to describe the trust fund portfolio as it relates to the strategic and operational activities the funds support.

The Bank may wish to enhance the monitoring of the development impact of trust-funded activities, and link the monitoring and reporting on the use and performance of trust funds with the development activity they support (box 3.1). This would provide a more comprehensive view of the activity's achievements, highlight the factors behind them, and serve as a useful building block for the Bank's results initiative, which seeks to determine the contribution of individual activities to country-level outcomes.

## SECTOR LEVEL

The framework for monitoring outcomes of the Bank's sector strategy and programs is evolving. Sector Strategy Papers (SSPs) lack measurable, time-bound performance targets that would serve as the basis for M&E. The framework for reporting on the implementation of sector strategies continues to be in flux. It is not clear which SSPs will have progress reports, what form they might take, and when they would be done.

**Setting clear outcomes and monitorable indicators.** The Millennium Development Goals (MDGs) provide the broad framework for setting sector and country goals for development agencies, including the Bank, but they have distinct limitations as operational indicators of outcomes, as summarized in box 2.3.

### Box 2.3: Some Limitations of the MDGs as Monitorable Targets

Limitations of the MDGs as outcome targets include the following:

**Not based on outcomes:** In the education sector, the MDG indicators focus on enrollment and completion rates, and not on improved learning achievement.

**No clear definition:** There is no common definition of *safe water* and *sustainable access* in the urban sector.

**No means of measurement:** There are no established measurements for sustainable access to urban areas or markets.

**Not specific:** On environmental sustainability, the MDG calls for “integration of sustainable development into country policies and programs” without specifying what integration would mean and the level and intensity of integration needed. The MDG about slums is expressed in absolute terms (“improving the lives of 100 million slum dwellers”) and does not address the growing proportion of the population in slums.

**Data not available:** Data on sanitation coverage and secure tenure are available for only half of developing countries.

Source: OED 2002g, p. 32.

The SSP is the Bank’s main instrument for linking the MDGs and other development targets to a set of outcomes for the Bank, with supporting monitorable indicators. SSPs are expected to establish the basis for evaluation through a matrix that specifies a system for monitoring and evaluating their implementation, and that includes indicators, baselines, and targets over a specified time frame. Following a self-evaluation of SSPs in 2001 that noted weaknesses in M&E, SSPs were required to provide a set of monitorable indicators of Bank and client performance. But SSPs continue to be weak in establishing such measurable performance targets and in defining monitoring schedules and responsibilities.

**Quality assurance:** Quality assurance for sector strategies and their implementation is evolving. In FY03, QAG started assessing the activities of Sector

Boards to identify systemic strengths and weaknesses and to provide a framework for greater accountability and learning. QAG will use the results to develop indicators of Sector Board performance, covering strategy in Bank operations, quality support services and knowledge management, human resources and learning, external partnerships, and Sector Board governance. Such assessments should include an explicit focus on the role of Sector Boards in monitoring progress of Bank operations at the instrument and country level in meeting SSP outcomes and MDGs.

**Monitoring, reporting, and reviewing implementation progress:** In 2001, management proposed a succinct biennial report to the Board detailing progress on SSP implementation that would build on monitoring indicators proposed in the SSP. These periodic implementation reviews were expected to identify and address systemic issues such as weak CAS linkages, impact of budget constraints, and gaps in Bank Group coordination.

Since then, management has provided formal SSP implementation reviews on energy and environment and on gender. These interim reports on SSP implementation are provided to the Board for information and are not discussed. The progress reports have generally not been produced on a biannual basis and have varied in quality, with no standard format or scope. In addition, management has prepared four topical briefings in the form of presentations. These are prospective rather than retrospective, and only two have commented on SSP implementation. An update to the 2001 SSP stocktaking exercise was scheduled for FY03, but is yet to be completed. The 2003 update was expected to include the evaluability of SSPs and the extent to which SSPs have adopted frameworks for M&E.

Network anchor and sector units are responsible for monitoring sector strategy implementation and the Bank’s operations. Most of their monitoring is currently centered on lending operations, with less emphasis on nonlending activities and on tracking the consistency between sector operations and country strategies and their mutual relevance. The indicators used for monitoring relate to portfolio quality, risk, and SAP-based information on timing, costs, and deliverables, but not to progress against out-

comes and related intermediate indicators.<sup>18</sup> Institutional arrangements for monitoring sector operations vary among networks. In the Private Sector Development and Infrastructure Network, the Financial Sector Network, and the Human Development Network, the responsibility for monitoring sector operations largely rests with sector units, while in the Poverty Reduction and Economic Management Network, the anchor unit plays a significant role. However, there appears to be no correlation between the quality of a sector's portfolio and the arrangements for monitoring sector operations.

Of the 25 OED sector and thematic evaluations completed since FY98, more than half have recommended strengthening the monitoring of sector outcomes. Some examples are noted in box 2.4. While progress has been made in implementing many of these recommendations, they are not yet complete. A 2001 self-evaluation noted that future SSPs would more systematically incorporate management action responses to OED sector evaluations, offering another way to clarify a selected set of SSP implementation indicators.

## COUNTRY LEVEL

The introduction of pilot results-based Country Assistance Strategies aims to enhance the outcome orientation of Bank assistance and enhance the monitoring of Bank operations at the country level. Most Country Assistance Strategies (CASs) do not systematically set out clear outcomes for the Bank, link Bank interventions to expected outcomes, or define performance indicators. The success of this results framework for CASs hinges on the quality and usefulness of the proposed CAS outcome matrices. CAS Progress Reports are expected to use these performance indicators in assessing implementation and progress against expected CAS outcomes.

**Setting clear outcomes and monitorable indicators:** The CAS is the Bank's tool for articulating its assistance program and outcomes at the country level, typically over a three-year period. The CAS is intended to tie project outcomes to country outcomes and to the higher-order objectives that the

### Box 2.4: OED's Recommendations on Monitoring Sector Progress

Several OED sector and thematic evaluations have reported on inadequacies in the sectoral monitoring framework and have made recommendations to strengthen it.

1. The *1998 Financial Sector Review* (OED 1998) presented a set of performance indicators to measure economic performance of countries and progress in improving the global financial and trade system. The Bank's financial sector strategy does not include any of these performance measures to assess its interventions.
2. A *1999 Review of Health, Nutrition and Population* (OED 1999) pointed out severe deficiencies in monitoring. This led to the development of a recommended group of core intermediate indicators for the sector. But the 2002 ARDE notes that the recommended indicators—such as prevalence of underweight children below five years, proportion of women with pre-natal care, contraceptive prevalence rates, etc.—have still not been fully applied in health projects.
3. The 2000 evaluation of the *Bank's Poverty Reduction Strategy* (OED 2000) called for enhancing monitoring through SSPs by systematically linking sector-specific and intermediate development objectives with progress indicators in poverty reduction. Management was expected to develop a methodology for evaluating the poverty orientation of SSPs and pilot test the methodology. This has yet to be fully implemented.
4. The *Review of the Urban Sector* (OED 2003d) recommended revising the SSP to set priorities and provide explicit targets that would link the sector's key instruments to the main elements of the strategy and to urban poverty alleviation.

project may support (such as sector impact, poverty reduction, and the like). In FY03, an internal review found that the main challenge for CASs is to articulate the links between desired results, appropriate mix of products and services, and types of interventions. CASs are often weak in setting out program goals. For example, only 60 percent of CASs carried out in FY00 and the first half of FY01 contain core targets on poverty reduction or MDGs, and as many

as 50 percent do not include reasonably quantifiable and time-bound indicators. Only a fifth of those that have performance indicators (10 percent of total CASs) distinguish between targets for country and Bank performance (OPCS data).

Management is piloting a results-focused monitoring and self-evaluation framework for CASs to enhance outcome orientation, performance reporting, and accountability (box 2.5). This framework is described in detail in Annex A. By requiring project objectives and outcomes to be linked to CAS outcomes alone, the results-focused CAS provides an opportunity to clarify persistent confusion among project-level task teams about what to monitor. OED endorses the proposed pilot approach to the results-based CASs. The success of the results framework hinges on the quality and usefulness of the proposed

CAS outcome matrices. Operational guidance and tool kits, staff incentives and training, and managerial attention to the quality of results-based CAS matrices are critical.

**Quality assurance at entry:** The Sector Boards, networks, and OPCS all review CASs presented to the Board. For countries where there is an OED Country Assistance Evaluation, the recommendations of the evaluation benefit the in-coming CAS, and are specifically addressed by it.

**Monitoring and review of implementation progress:** The CAS Progress Report (CASPR) updates the executive directors on implementation of the CAS program and on key developments since the completion of the CAS. CASPRs include adequate and candid discussion of shortfalls in achievement of objectives, of progress in poverty reduction, and of portfolio issues. Discussions of the use of triggers in assessing lending and monitoring indicators in evaluating implementation, the nonlending program, and treatment of risks remain inadequate, according to the FY03 internal review. The Country Portfolio Performance Review (CPPR) is a process for reviewing individual lending operations in a country. Management has issued a draft Operational Policy/Bank Procedure and guidelines for the Country Portfolio Review (CPR), which is expected to replace the CPPR. The CPR will be a more dynamic and integrative process that links the CAS with Bank-supported projects, as well as grants, guarantees, and selected advisory and analytical services (box 2.6). The final Operational Policy/Bank Procedure, originally expected in early 2002, is yet to be issued.

### **Box 2.5: Results-focused CASs Designed to Provide an M&E Framework**

The results-oriented framework envisages a set of CAS matrices indicating:

- (a) Key development goals (long-term, strategic) for the country in light of the MDGs
- (b) Outcomes the Bank expects to influence during the CAS, with related intermediate indicators
- (c) Bank interventions proposed to achieve desired outcomes, along with associated monitorable indicators of Bank performance.

To be operationally relevant, the matrix will clearly identify indicators, baselines, and targets; develop monitoring plans; and highlight data sources. To ensure adequate prioritization of outcomes and to avoid creation of large, complex, and unviable matrices, the CAS text would include a summary matrix that consolidates priority outcomes and links it to long-term strategic goals and MDGs.

The Country Assistance Strategy Progress Report (CASPR) would review progress toward desired outcomes and Bank performance, using indicators and targets provided in the CAS matrices. It will highlight adjustments needed to the Bank's program and triggers.

### **SUMMARY**

The monitoring framework is strongest at the project level. Recent initiatives are under way to improve monitoring of trust fund activities and to enhance results orientation at the country level. It is too early to comment on the effectiveness of these new initiatives. Monitoring is weakest for nonlending services and SSPs.

### **Box 2.6: A New and Comprehensive Approach to Country Portfolio Reviews**

The Bank is planning to revamp the CPPR into a more dynamic and comprehensive exercise—the CPR. In the CPR, the Bank and borrower would identify and address systemic issues and learn from experience to improve implementation of the existing portfolio and pipeline of projects and AAA. It aims to reinforce borrower ownership of Bank-financed projects and ensure continued relevance of projects in the portfolio to the country and sector strategy. The Bank would encourage borrowers to take responsibility for the CPR process, together with key implementing agencies and ministries, people affected by Bank projects, and other interested stakeholders and donor agencies.

Pending an Operational Policy on the CPR, several country teams, especially in the East Asia and the Pacific and the Africa Regions, are using the CPPR process more strategically, although in different ways. The *2002 Annual Report on Portfolio Performance* reports that in Mongolia, for instance, the country team extended the scope of the CPPR to include not just the project portfolio, but also economic and sector work, fiduciary issues, and harmonization. It was led by the borrower and strongly supported by key donors such as Japan and the Asian Development Bank. The Africa Region model is also participatory. It uses the CPPR in the CAS design process to aid alignment with objectives in the Poverty Reduction Strategy Papers and to foster programmatic approaches.





### 3. Evaluation of Bank Operations



Evaluation determines the extent to which operational programs and activities produce desired results. It attempts to provide attribution and causality. As the main tool for demonstrating results, evaluation is fundamental to the Bank's *Managing for Results* initiative.

Evaluation can be of two kinds: self-evaluation and independent evaluation. Evaluation by the management of programs or activities is self-evaluation. Management may carry out self-evaluation using internal staff or external consultants or contractors. Independent evaluations are conducted by or on behalf of governance bodies that oversee management, such as the Board of Directors. Both self-evaluation and independent evaluation can deal with individual activities and activities at the aggregate levels.

This chapter describes the evaluation framework in the Bank at the instrument, sector, and country levels (summarized in table 3.1). It aims to identify strengths and weaknesses of the evaluation framework, comment on its effectiveness, discuss recent initiatives to improve it, and highlight areas for greater management attention. A description of the roles of units engaged in evaluation activities in the Bank is provided in Annex B.

#### INSTRUMENTS

##### *Lending Operations (Investment and Adjustment Projects)*

The evaluation framework at the project level is well established and reliable. There are two areas for

improvement: (1) the quality of project evaluations can be enhanced by better establishing project outcomes during design and (2) the self- and independent evaluation framework must be adapted to the needs of special purpose lending instruments such as Adjustment Loans and Adaptable Program Loans.

The task team for each project is expected to prepare an Implementation Completion Report (ICR) within six months of project completion. The ICR provides ratings on outcome, sustainability, and institutional development as well as on Bank and borrower performance. Since the ICR is a key pillar for evaluation at the project level, it is important that it be reliable and of adequate quality. To validate ICR reliability, OED conducts a desk review of each ICR and its ratings, producing an evaluation summary for internal use.

About 25 percent of all completed projects are subject to an intensive independent evaluation through OED's Project Performance Assessment Reports (PPARs). A PPAR may be carried out several years after project completion. For instance, in FY02, PPARs were carried out on projects exiting between 1988 and 2001, with an average time lag of 3 years after project completion. PPARs usually involve a mission to the project site. These reports have been disclosed to the public since 2002. Management also

**Table 3.1: Framework for Evaluation in Bank Operations**

Intervention or financing instrument type	Self-evaluation		Independent evaluation (OED)	
	Individual activity	Aggregate level	Individual activity	Aggregate level
<i>Instrument</i>				
Lending operations	Implementation Completion Reports (ICRs)	Lending retrospectives	Evaluation summaries, Project Performance Assessment Reports, Impact Evaluations	Annual Review of Development Effectiveness (ARDE)
Economic and sector work/ analytical and advisory services	Activity Completion Summaries (ACS) for tasks > \$50,000	Regional retrospectives ( <i>ad hoc and rare</i> )	Periodic studies (such as Public Expenditure Reviews, Poverty Assessments)	Sector and thematic studies, implementation reviews, and Country Assistance Evaluations
Trust funds	Implementation Completion Memorandum (ICM) > \$1 million	Annual Review of Trust Fund Portfolio Performance, some donor reports		
Sector strategy		SSP Stocktaking	Sector Implementation Reviews	ARDE
Country program	CAS Completion Reports ( <i>new</i> )	CAS Retrospectives	Country Evaluation Summaries ( <i>new</i> ), CAEs	ARDE, Regional OED Evaluations (CFA, transition economies)
Grants/global programs	Grant Completion Reports (GCR)	DGF evaluations for > \$300,000		OED evaluation of global public programs
<i>Knowledge initiatives</i>				
Knowledge management	For few knowledge sharing activities			Evaluation of Knowledge Sharing (OED 2003a)
DEC research	Research Project Completion Reports	2001 Research Support Budget (RSB) evaluation		
WBI training & capacity building	Individual learning events	WBI evaluations of selected programs		

Source: World Bank data.

carries out periodic self-evaluations of investment lending and adjustment lending that address issues at the level of the individual instrument—quality, compliance with guidelines, risk, and outcome.

As detailed in Annex C, comparison of the ratings from self-evaluation (ICR) and OED summary information indicates that the self-evaluation process is fairly reliable. For about 5 percent of the projects evaluated each year, the rating given in the ICR is changed in the OED review (upgrades plus downgrades). The net rating effect (downgrades minus upgrades) is a 4 percent downgrade between the OED summary and the ICR. These levels have been stable since FY99.

Management rates projects on the achievement of development objectives during supervision and

on completion. The share of downgrades between the final Project Supervision Report (PSR) prepared by the task team before a project exits and OED's review has increased from 8 percent of projects downgraded in FY99–01 to 12 percent downgraded in FY02. On project completion, management prepares an Implementation Completion Report. The ICRs also downgraded the final PSR ratings in a net of 6 percent of projects in FY02. This suggests that the PSR has become less realistic. The net difference in ratings between the ICR and OED's summary has increased marginally, from 5 percent net downgrades during FY99–01 to net 6 percent in FY02. Overall, these patterns in rating changes suggest that Regions are more optimistic about development effectiveness than OED, especially as indicated by

PSR ratings. It also underscores the need for detailed independent assessments of operations through PPARs.

On ICR quality, 94 percent of ICRs reviewed in FY02 were satisfactory. Although the overall quality is high, there is room for improvement in ex-post economic analysis and presentation of evidence, poverty analysis, and the plan for M&E of future operations. Annex C compares management and OED ratings and trends in ICR quality during FY96–02.

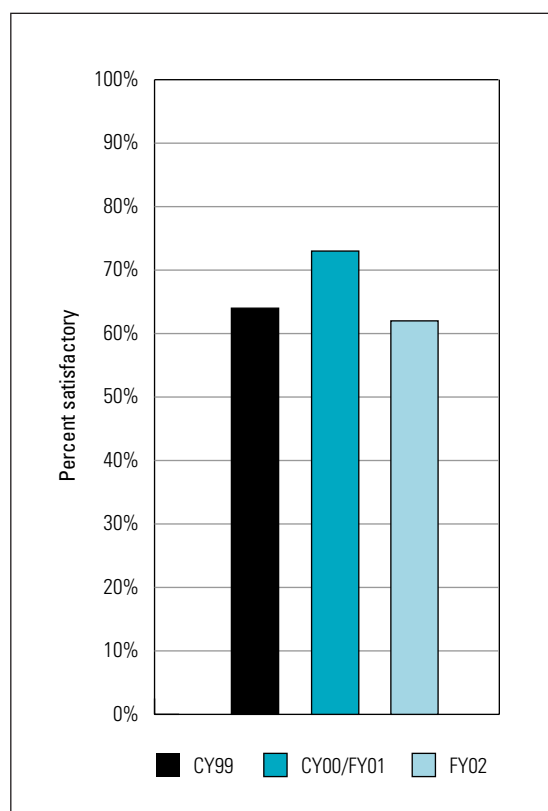
With functioning systems for completion reporting and periodic self- and independent evaluation, the evaluation system for projects is stronger and more comprehensive than the monitoring system. Two constraints to evaluation effectiveness remain to be addressed. First, evaluability or definition of outcomes and arrangements for evaluation need to be sharpened at the time of project design. As shown in figure 3.1, QAG has noted weaknesses in this dimension for about a third of projects reviewed. Second, ICR design has not kept pace with the evolution of lending instruments. The special features of Adjustment Loans, Adaptable Program Loans, and Learning and Innovation Loans need to be reflected in the guidelines for their ICRs (OED 2002a).

### Nonlending Activities

The evaluation framework for nonlending services is not as well established as it is for lending operations. Nonlending programs are a key building block of Bank assistance at the country level, and absence of information on their outcomes could undermine the results orientation of the CAS. Two issues need to be addressed: (a) enhancing evaluability by defining outcomes and related indicators in Concept Papers, and (b) addressing gaps in self-evaluation of individual or clusters of tasks, and for ESW tasks in aggregate.

For nonlending tasks, self-evaluation at the individual activity level for tasks over \$50,000 is reported in Activity Completion Summaries (ACSs), which are meant to be prepared within six months of task completion.<sup>1</sup> The ACS is expected to summarize task objectives and the extent to which they were

**Figure 3.1: Definition of Outcomes and the Quality of Arrangements for Evaluation Vary**



Source: QAG data.

achieved, to describe dissemination processes and activities, and to identify lessons learned.

Since 1999, QAG has reported problems with the content and quality of the ACSs that remain unresolved. In the absence of clearly defined expected outcomes and related indicators in the Concept Paper, it is not surprising that the ACS reports more on activities completed than on the tasks' achievements and outcomes. "Only a handful of ACSs contained findings material to the achievement or lack thereof of success indicators, including client feedback" (QAG 2003b, p. 16). QAG's ratings on quality of ACSs have fallen from 67 percent satisfactory or better in FY00 to 52 percent in FY02. An OED review of a random sample of 20 ACSs for tasks completed in FY02 supports this finding. None of the ACSs used performance indicators for assessing outcomes or reporting on achievements. Further, the ACS format itself requires information only on

task conclusions and recommendations, not on task achievements or lessons learned. Only 2 of 20 ACSs reviewed reported on lessons.

At the aggregate level, except for annual QAG reviews of the quality of ESW tasks, there is little systematic Bankwide evaluation. The roles, processes, and responsibilities governing self-evaluation of ESW are not well established. There is no requirement for the Regional Chief Economist's Office to review outcomes and achievements of ESW on a regular basis. Interviews with Regional staff indicate that ACSs are not actively used to analyze systemic issues in ESW implementation or to capture lessons. Two of six Regions (the Middle East and North Africa and South Asia) have recently reviewed their ESW programs. These retrospectives were ad-hoc; focused on large tasks; took stock of the kinds and numbers of tasks, costs, and timeliness; and assessed them with QAG-like criteria of quality. They did not report on outcomes or contributions to the CAS.

OED does not have a separate instrument for evaluating ESW. But its sector and thematic reviews, CAEs, and other evaluations do systematically assess the nonlending program for outcomes and contributions to Bank strategies.<sup>2</sup> In addition, OED has evaluated specific diagnostic instruments such as Public Expenditure Reviews and Poverty Assessments. A review of Financial Sector Assessment Program is forthcoming; and one on fiduciary instruments is proposed in the FY04–06 OED work program.<sup>3</sup>

There are two broad issues that need to be addressed: first, better evaluability through clearly defining outcomes and supporting indicators in the Concept Paper, and, second, addressing gaps in self-evaluation. On the latter, self-evaluation of individual ESW tasks through ACSs is not yet effective. At the task level (an individual task or a cluster of tasks), the ACS should report on achievements against expected outcomes, as specified in the Concept Paper. At Regional and sector levels, the Regions should undertake a systematic and periodic assessment of the *outcomes* of all ESW tasks (small and large, formal and informal). Sector Boards should build on Regional assessments to draw sectoral lessons on cross-country experience. At the corporate level, management should undertake periodic self-evaluations that

extend beyond quality to the use and trends in outcomes of different products. Completion reports, such as the ACS, could provide useful inputs for aggregate-level reporting of effectiveness.

OED is considering the possibility of reviewing—on a pilot basis—a sample of ACSs, to suggest ways to strengthen completion reporting on large ESW tasks. QAG has also recommended that management review the quality, prevalence, and use of information in the ACS. The purpose would be to improve current tools or to design alternatives for monitoring and reporting on task progress and results.<sup>4</sup>

### *Trust Funds*

The evaluation framework for trust funds is partial. It focuses on compliance with Bank policy and donor conditionality; evaluation of the outcomes of trust fund activities needs to be strengthened. Management may also wish to integrate the M&E framework for activities financed by trust funds with the framework already in place for the core Bank activities the trust funds support.

Task teams using trust funds in excess of \$1 million provide self-evaluation of trust fund activities through an Implementation Completion Memorandum (ICM). Vice-presidential units are responsible for the annual review of the portfolio of activities supported by trust funds. In FY02, there were 497 such trust funds over \$1 million. These amounted to \$1.68 billion and accounted for 57 percent of the total amount of trust funds at the end of FY02. There is no formal mechanism to report on the 4,035 trust funds of less than \$1 million, which amounted to \$1.25 billion and represented 43 percent of the net approved value at the end of FY02.<sup>5</sup>

The information on outcomes and achievements from ICMs and other SAP-based tools is partial. In the absence of an effective framework for monitoring and reporting on trust fund outcomes, as noted in Chapter 2, vice-presidential unit reports are based largely on anecdotal evidence.<sup>6</sup> The Trust Funds Quality Assurance and Compliance Unit (TQC) uses the vice-presidential unit reports to prepare a consolidated annual summary and analysis of the Trust Fund Portfolio. The Trust Fund Portfolio Review

reports on the Bank's experience in using trust funds for program delivery, evaluates the quality of trust fund uses, and ascertains compliance with both the Bank's policies and procedures and donor conditionality. Because of the lack of adequate information from vice-presidential units, the TQC does not comment on achievements of trust fund activities or on the adequacy of ICMs.

The evaluation framework for trust funds currently focuses mostly on compliance and the disbursement of funds. The TQC has recommended that management develop standard performance indicators for the use of trust funds, and that Regional and sector units ensure greater alignment of individual and programmatic trust fund proposals with CAS or sector priorities.<sup>7</sup> This recommendation has been supported by prior OED evaluations, which suggested that the objectives of a trust fund be expressed in terms of its contribution to the outcomes and development effectiveness of the project or activity it supports (OED 2002c,h). This objective should then form the basis for evaluation. Integrating the M&E of trust fund activities with core business operations, as suggested in box 3.1, would have the following benefits:

- Support the results-focused CAS initiative by providing incentives for and enhancing the transparency and comprehensiveness of M&E of core Bank operations.
- Improve Board oversight over the use of trust funds and ensure improved alignment of trust funds with the Bank's country and sector priorities.
- Enhance donors' ability to undertake periodic assessments of the effectiveness and achievements of particular types of trust fund programs.
- Increase efficiency by avoiding creation of parallel structures.

Global trust fund programs accounted for about \$1.68 billion, or 44 percent of all trust fund commitments in FY02 (OED 2002c). Yet regular reporting is required for only some global programs, and even in these cases, the reporting is often ad hoc and is focused more on inputs and activities than on outcomes. In this regard, management may wish to consider setting guidelines for independent evaluations

### **Box 3.1: Integrating Trust Fund M&E with Bank Operations**

There is considerable potential to integrate M&E of the use of trust funds with processes already in place for the activity being supported. Four Regions have started to integrate trust funds into Country Portfolio Performance Reviews (CPPRs).

- The Middle East and North Africa Region reports that discussions of trust fund issues in the CPPR, for example, have helped stem deterioration of the grant portfolio by focusing on implementation problems, reducing operational risks, and improving the effectiveness and quality of trust fund implementation.
- The Latin America and Caribbean Region notes that including trust funds in the CPPR helps stress the importance of trust funds in the Bank's overall country programs.

Where the trust fund facilitates project preparation or complements specific components, management may wish to link the Status Report and ICM to the PAD and/or PSR, as well as other project documentation. The ICR and OED reviews of the projects would then serve as self- and independent evaluations of the trust fund as well. For nonlending activities, management may wish to link trust fund reporting to the Concept Note and/or Decision Draft. As management sets up SAP systems to report on trust fund status and completion, it may be more efficient to integrate Status Reports, when applicable, with existing SAP reports and procedures for projects and analytical and advisory services.

of large trust funds. This would be consistent with the existing practice, which requires external evaluations of Development Grant Facility grants in excess of \$300,000.

### **SECTOR LEVEL**

The Bank has moved from developing sector strategies to implementing them. But there is no framework for self-evaluation of sector outcomes. It is not clear when the SSP would be sufficiently advanced in implementation to warrant an evaluation of outcomes. In the past, OED has provided independent evaluations of Bank experience in numerous sectors,

timed to serve as inputs to the new or updated sector strategy (see, for example, OED 1999, 2002f). In the absence of a monitoring and self-evaluation framework for SSPs, OED plans to provide the Board with Sector Policy and Implementation Reviews (SPIRs) that will independently assess sector issues and strategy implementation.<sup>8</sup> OED and management will need to develop a formal framework and timeline for evaluating the outcomes of sector strategies.

## COUNTRY LEVEL

The independent evaluation framework at the country level comprising OED's Country Assistance Evaluations (CAEs) is well established. The proposed CAS Completion Report (CASCR) responds to a long-standing OED recommendation and closes important gaps in self-evaluation at the country level. It is too early to comment on its effectiveness.

The *Managing for Results* initiative is piloting a framework for self-evaluation at the country level, comprising CASCRs that will be prepared at the end of each CAS period. This responds to a long-standing OED recommendation and seeks to close the gaps in the framework for self-evaluation at the country level (detailed in Annex A). The CASCR will assess outcomes achieved relative to expectations and draw lessons for the next CAS.<sup>9</sup> It will comment on the relevance of programs to the country's development agenda, the quality of the portfolio, measures of development effectiveness, and include insights from clients, where available. A limitation in the first set of CASCRs will be the lack of a results framework in the previous CASs.

As it does at the project level, OED will review and validate the self-assessments of country outcomes in the CAS Completion Reports and provide Country Evaluation Summaries.<sup>10</sup> And just as OED conducts in-depth assessments of project outcomes through PPARs, OED's Country Assistance Evaluations will continue to provide in-depth assessments of the Bank's strategy and assistance programs at the country level. Because the Country Evaluation Summaries will assess the CASCR, they will cover only one CAS cycle (four years).

OED has reviewed one CASCR to date and recommended that the report discuss (a) Bank lending and nonlending assistance delivered since the last CAS as it related to specific CAS objectives and (b) the extent to which Bank operations contributed to the achievement of CAS objectives and outcomes. Because development outcomes may not be realized in the relatively short period covered by a CASCR, OED suggests that the conclusions and lessons for the future presented in the CASCR be clearly linked to the evidence on Bank assistance. OED is working with OPCS and pilot country teams to help refine the structure and methodology for the CASCR, including a performance-rating system for country programs.

OED's Country Assistance Evaluations provide the Board with an independent assessment of the achievements and limitations of the Bank's country program, usually covering a time frame of up to three CAS periods (about 10 years). They are prepared up-stream of, and provide learning inputs for, the CAS. As of June 2003, OED had completed 62 Country Assistance Evaluations, and each year's Country Assistance Evaluations cover about a quarter of the CASs approved in that fiscal year. As CASCRs are mainstreamed, OED expects to increase its resources for Country Evaluation Summaries and reduce the number of CAEs somewhat.

## *Grants and Global Programs*

The evaluation framework for grants and global programs is not fully effective. A recent OED evaluation of global programs highlights weaknesses in the quality of the self-evaluations, especially in reporting on outcomes. There are also gaps in self-evaluation of grants and global programs at the aggregate level. Management has generally endorsed OED's recommendations and plans to create a Global Programs and Partnerships Secretariat to enhance oversight. This new unit should set responsibilities for establishing M&E frameworks for grants and global programs.

Grants are an integral part of Bank activities and are an important complement to its lending and nonlending operations. Of the total grant funds from the administrative budget, amounting to \$191 mil-

lion in FY02, about \$176 million is channeled through the Development Grant Facility (DGF), which then allocates them to about 48 individual programs, managed either by the Bank or by external recipients (OED 2002c).

The program manager for each DGF grant is responsible for preparing a Grant Completion Report (GCR) within six months of the closing date of the final grant Letter of Agreement. The GCR is expected to describe the extent to which program objectives and activities were met, to assess the program's development impact, to identify implementation problems, and to provide lessons learned. The GCR often includes audited financial statements and recipient-prepared evaluation reports. A 2002 OED evaluation of the Global Public Programs of DGF grants (OED 2002c) noted that only half the programs that had disengaged from DGF funding since FY98 provided timely Grant Completion Reports. Some recipients submitted GCRs as late as a year after closing.<sup>11</sup> Additional GCRs have since been submitted; currently, of the 32 DGF programs that have exited since FY98 and for which GCRs are required, 5 have not been received. The 2002 OED evaluation also noted that the quality of GCR reports has been uneven: they focus on achievement of overall objectives and completion of program activities, but are weak in (a) identifying problems in program execution, (b) reporting on achievement of intermediate objectives, and (c) assessing development impact.

GCRs are required only when a program exits from DGF funding; in some cases (Window I grants) this may take several years. The DGF thus requires that grant programs of \$300,000 or more carry out an external evaluation every three years. Since external consultants paid for by the program management conduct most of these evaluations, they are external self-evaluations, not independent evaluations. Six such evaluations were conducted in FY00–01, and similar assessments for all major programs are planned over FY02–04. In many cases, these evaluations are the first for the grant programs.

OED's evaluation of grants and global programs has reported concerns about the quality of the external self-evaluations. Typically, the external evaluations report on implementation of activities, but not

on outcomes and impacts. Only the Consultative Group on International Agricultural Research (CGIAR) has systematically and quantitatively attempted to assess impact. This limits information needed by the Bank to determine funding priorities. OED has called for "standardized evaluation criteria, clear performance indicators, and timely and clear definition of what constitutes independent evaluations of high quality to enable learning, effective prioritization and resource allocation across programs and networks."<sup>12</sup> Management has generally endorsed OED's recommendations and plans to consolidate existing units into a Global Programs and Partnerships Secretariat to provide guidance on and enhance oversight of global programs. The Secretariat will set responsibilities for monitoring and self-evaluation of grants and global programs.

In general, the Bank, as a grantor, has only limited legal responsibility for the use of individual program funds. But where a program is Bank-managed, there is no arms-length relationship between grantor (the Bank) and the program. Hence, the Bank must ensure adequate oversight for such programs. As part of phase two of its evaluation of global programs, OED is reviewing 26 programs that involve the Bank. Following this, OED will determine its role in the evaluation of grants and global programs.

## **KNOWLEDGE INITIATIVES: KNOWLEDGE SHARING, RESEARCH, AND TRAINING**

Evaluation of the Bank's knowledge sharing activities needs strengthening. There is no M&E framework over the Bank's knowledge sharing activities. Evaluation of the Bank's research is sporadic and needs to be formalized.

In 1996, the World Bank committed itself to developing a world-class knowledge management system to improve and expand knowledge sharing among staff and with clients and other development partners, and to help countries strengthen their own capacities for acquiring and using knowledge. There are three components to the Bank's knowledge agenda: (a) network and Regional knowledge management and knowledge sharing activities, (b) research



and data (Development Economics Vice-Presidency, or DEC), and (c) capacity building and training activities (World Bank Institute, or WBI).

**Evaluation of the Bank's knowledge sharing and management activities:** In response to the 1996 commitment, the Bank initiated several knowledge management programs and activities, amounting to approximately \$220 million during FY97–02. These activities include the (a) collection and dissemination of knowledge and best practice by network and Regional units; (b) creation of thematic groups to enhance connectivity among staff around shared areas of work; (c) use of on-demand assistance by advisory services; and (d) knowledge sharing innovations in country programs and projects. The Bank has also created several global knowledge initiatives that leverage technology to enhance knowledge sharing and client capacity.<sup>13</sup>

An OED evaluation, which is the first independent assessment of the gamut of knowledge sharing activities (OED 2003a), notes that M&E has not kept pace with the implementation of the Bank's knowledge sharing initiatives, or with leading industry practices. It reports that the Bank still lacks the framework and measurements needed for assessing the effectiveness of knowledge sharing initiatives and recommends ways to strengthen it.

Management has recently established a Steering Committee at the vice-presidential level to guide and oversee the Bank's knowledge sharing activities. It will set responsibilities for M&E. QAG has also recently initiated an assessment of the activities of Sector Boards, including knowledge management services.

**Evaluation of the Bank's Research:** DEC carries out most of the Bank's research, which is divided into two categories based on the source of funds: (a) Research Support Budget (RSB) activities and (b) non-RSB-supported activities. In FY02, the RSB activities amounted to \$4.2 million, or 29 percent of the total Bank research activities of about \$14.5 million. DEC's share of total Bank research was 83 percent (about \$12.3 million), of which 73 percent was financed by DEC's administrative budget and 27 percent by the RSB.

Oversight of the Bank's research, which is mandated to ensure the quality of Bank research and its

relevance to the Bank's mission, rests with the Bank Chief Economist, with the advice of the Research Committee (for the RSB activities). DEC has carried out self-evaluations of RSB-supported research in a number of dimensions such as relevance, quality, dissemination, and impact. The most recent report on RSB research dates from FY01. As noted in the 2002 AROE, non-RSB activities have never been evaluated systematically, and past efforts to evaluate non-RSB research through client surveys were not systematically followed through. While most of DEC's research is published in professional journals, and is therefore vetted by an external peer review for intellectual quality, such reviews do not necessarily address the impact or applicability of the research to Bank operations and effectiveness.

The Development Economics Vice-Presidency should establish a framework for evaluating all its research (RSB- and non-RSB funded) at three levels: (a) the program level, especially for large research projects; (b) the operational level, for the applicability, utilization, and impact of DEC research in aggregate on Bank operations; and (c) the global level, for the impact of DEC research on the development community at large (research, academia, borrowers, donor agencies). DEC has carried out evaluations at each of these three levels sporadically, but this work should be instituted as a regular and formal feature of its evaluative framework. DEC also undertakes periodic self-evaluation of its cross-support activities through client surveys.

**Evaluation of Training Activities:** In 1997, WBI introduced a four-level evaluation framework for all Bank training activities that ranges from monitoring trainee feedback to assessing outcomes and impacts.<sup>14</sup> WBI now requires that all major learning offerings be evaluated for their perceived usefulness to participants. In FY02, WBI developed and implemented 28 valid and reliable tests to measure enhanced learning, and learning gains have been measured on over 100 courses. In FY02, for the first time, WBI also undertook an impact evaluation for five thematic programs aimed at 1,200 participants in 82 client countries. This impact evaluation represents progress in self-evaluation of training activities. It measures program effectiveness and impact

based on participants' perceptions, but does not validate them with indicators of enhanced learning or knowledge applicability. The impact evaluation also called for greater linkages or coordination of WBI programs with core Bank operations.

WBI's recent shift of its capacity building activities to the country level (starting with five pilot countries) aims to support country teams and borrowers in conducting a capacity needs assessment and using it to develop an action plan for building capacity. WBI's planned capacity building will be reflected in the CAS. Building WBI's work into the CAS is expected to encourage coordination of WBI activities with Bank operations, and to make WBI interventions more strategic and results-focused. The corresponding evaluation strategy is expected to use prospective evaluation methods to collect base-

line data before WBI programs begin, and to set up in-country evaluation infrastructure (local evaluators, intermediate indicators, reliable data collection and monitoring systems).

## SUMMARY

The self- and independent evaluation framework is strongest at the project level. Several recent initiatives are under way to improve the evaluative framework for trust funds, grants, and at the country level, although it is too early to comment on their effectiveness. The evaluation framework is weakest at the sector level, for nonlending operations, and for the Bank's knowledge-based initiatives.



## 4. Corporate Priorities in Monitoring and Evaluation



This chapter uses the strategy and framework proposed by the Managing for Results initiative to discuss the progress in developing M&E processes at the corporate level. The Managing for Results initiative has identified two corporate priorities: (a) helping client countries adopt results-based strategies and (b) increasing the Bank's results orientation.

### HELPING CLIENT COUNTRIES BUILD RESULTS ORIENTATION

The Bank has been active in helping borrowers strengthen their results orientation and evaluation capacity in order to enhance government performance. Yet it has made only limited progress in mainstreaming evaluation capacity building into core Bank operations.

The *Managing for Results* initiative rightly emphasizes results—the achievement of MDGs and the need for well-functioning M&E systems in the Bank and in borrower countries. Bank and borrower M&E are closely related: the Bank relies on country data for measuring its own performance at the project, sector, and country levels. But limitations in country M&E—data, systems, and capable evaluators—have encouraged the Bank and other donors to take an enclave approach to M&E, with little emphasis on the building of sustainable country capacities and systems. This has resulted in a multiplicity of donor M&E burdens on borrowers (CDF Secretariat 2003).

Building governments' own capacity for M&E not only makes the aid process more effective, but it also contributes to improved *government performance*—

the ability of governments to use both domestic and donor funds to achieve development outcomes. Thus, since the 1970s the Bank has been involved in evaluation capacity development (ECD).

OED conducted stocktaking of Bank work to help borrowers build their M&E systems, with a specific focus on systems that measure government performance.<sup>1</sup> The stocktaking is based on an examination of project appraisal documents and other Bank loans and credit documents for projects approved in FY02, and approved Institutional Development Fund grants, Poverty Reduction Strategies Trust Fund grants, and grants made under the Trust Fund for Statistical Capacity Building. Also included is Bank capacity-building work identified by OED in 2002 where that work is still active. In the absence of a management initiative, OED is preparing a database on Bank work in this area.

Based on this stocktaking, OED has found that the Bank is currently providing ECD support at the national level in 21 countries.<sup>2</sup> Over half of these countries (13) have prepared or are preparing a PRSP, reflecting the potential importance of PRSPs for achieving a substantial focus on government performance. Two of the countries are European Union

accession candidates—European Union member states are required to meet certain standards for M&E in government. In about 13 of the countries, the capacity building is designed to support improved public expenditure management in government. In seven countries, civil society is expected to play a substantive role in providing assessments of government performance.<sup>3</sup>

While the Bank has accumulated considerable experience in helping borrowers strengthen their results orientation, it has made only limited progress toward mainstreaming this area of capacity building into Bank operations. OED has identified five impediments to mainstreaming ECD: lack of a clear commitment by the Bank; confusion about the meaning of “results” and “M&E”; inconsistency in the message on M&E in Bank policies and guidelines; lack of knowledge sharing and coordination on M&E; and weak staff skills and lack of resources across the Bank (OED 2002b).<sup>4</sup>

Finally, although Bank networks and sectors typically do not have a systematic approach to helping borrowers strengthen M&E of government performance, there is one promising exception. The Bank’s Global HIV/AIDS M&E Support Team has a detailed work plan to build donor and borrower M&E systems, involving capacity building at the global, Regional, country, and local levels.<sup>5</sup> This team is active in a large number of countries, particularly in Africa.

## RESULTS ORIENTATION IN THE BANK

Management expects to report on Bank performance at the corporate level by transforming the *Annual Review of Portfolio Performance* (ARPP) into an *Operational Performance and Results Review* (OPR). The OPR will initially be constrained by limited information on country-level outcomes pending mainstreaming of the results-based CAS initiative and other improvements to enhance the outcome orientation of lending and nonlending instruments. The OPR will need to evolve over time to become a comprehensive report on the results of all Bank activities. Management has also developed the IDA results measurement framework to report to donors on progress on important development outcomes in IDA-eligible countries.

Internationally accepted developmental goals such as poverty reduction and the MDGs form the foundation for all Bank operations. There are several ways for the Bank to track and report on performance against these broad developmental goals—the Poverty Progress Report, and the recently initiated Global Monitoring Report.<sup>6</sup> The progress or lack thereof in achieving these goals may not be attributable to the Bank alone, but to the development community at large. The challenge for the Bank at the corporate level is to translate the MDGs and other international goals of development into concrete outcomes for Bank activities and to track these outcomes using

**Table 4.1: M&E Framework at the Corporate Level**

Monitoring			Evaluation			
Goals	Defining outcomes & monitorable indicators	Tracking & reporting on implementation progress	Self-evaluation		Independent evaluation	
			<i>Activity</i>	<i>Aggregate</i>	<i>Activity</i>	<i>Aggregate</i>
Corporate (including IDA)	Strategic Directions Papers, IDA Replenishment Agreement, corporate budget documents	Strategy update papers, progress reports on corporate reforms.	Assessment of corporate initiatives (such as the Strategic Compact)	Proposed IDA Results Measurement, <i>Annual Report on Portfolio Performance/ Proposed Operational Performance and Results Review</i>	OED review of corporate initiatives (such as PRSP, HIPC)	ARDE, AROE, OED review of IDA performance

monitorable indicators and targets. The results of the Bank's efforts in relation to the MDGs were reported in OED's 2002 *Annual Review of Development Effectiveness* (ARDE) and are not addressed here.

The M&E framework for the Bank's corporate-level outcomes (including those for IDA) is summarized in table 4.1. The Managing for Results initiative envisages the need for a corporate report on Bank performance. In its implementation plan, management has proposed developing or adapting two instruments to report on Bank performance at the corporate level: first, the IDA Results Measurement System, and, second, the *Annual Review of Portfolio Performance* (ARPP), as it makes the transition to become an *Operational Performance and Results Review* (OPR).<sup>7</sup> This section discusses these two recent self-evaluation initiatives.

### *IDA Results Measurement System*

The IDA13 Replenishment Agreement stipulated that the Bank (a) establish country outcomes as performance measures for the IDA program and (b) put in place an enhanced results measurement system that would link IDA CASs with countries' progress in reaching the development outcomes set out in PRSPs. In response, management has proposed a two-tiered measurement system, as described in box 4.1.

The IDA results measurement system provides a mechanism for reporting to IDA deputies. The selected indicators reflect data availability, consistency with poverty reduction strategies, the MDGs, and other international monitoring activities. They provide a basis for common reporting on the development outcomes in IDA countries in aggregate, and for demonstrating the contribution of IDA-funded activities. Tier I indicators are not directly linked to country-specific targets for progress, institutional capacity and access to resources, or to Bank instruments such as the CAS, project design, and resource allocation to individual IDA countries.

### *Reporting on Bank Performance at the Corporate Level*

Management provides periodic updates on implementation of strategic initiatives such as the self-

#### **Box 4.1: Two-Tiered IDA Results Measurement System**

The first tier of the IDA Results Measurement System aims to assess progress on country outcomes as articulated in PRSPs and in the MDGs. Following a request from the executive directors to include indicators on growth, infrastructure, agricultural productivity, and other aspects of the growth agenda as well as MDGs, management proposed a set of 15 indicators and baseline data to the IDA deputies in April 2003. Management expects to refine and adapt these indicators over the course of the year, and report to the deputies on the updated system and progress against them in Spring 2004.

The second tier will focus on the contribution of IDA programs to country outcomes. The results-based CAS framework is expected to provide the necessary inputs in the medium term, but in the meantime, management expects to use IDA portfolio indicators derived from OED ratings for project outcomes and QAG quality assessments, which are correlated with success in achieving CAS outcomes.

assessment of the Strategic Compact. In most such instances, OED comments on management's updates and provides the Board with an independent assessment of the findings and issues. OED also periodically evaluates strategic corporate initiatives, such as knowledge sharing activities, the CDF, and HIPC (CDF Secretariat 2003; OED 2000, 2002d, 2003a,b).

At the aggregate level, QAG's ARPP reports on portfolio health at the corporate level and OED's ARDE comments on development results and addresses specific management initiatives or issues.<sup>8</sup> The ARPP database excludes Adjustment Loans approved and completed within a single fiscal year. These amounted to \$3.4 billion in FY02 and accounted for more than 23 percent by number and 11 percent by volume of total adjustment lending.<sup>9</sup> The ARPP's assessment of portfolio composition and risk is thus incomplete and needs to be supplemented with information on this sizeable share of lending.<sup>10</sup> An FY00 OPCS survey rated 98 percent of single-tranche adjustment loans to be good or moderate on design, but rated 18 percent to be "inferior" on M&E adequacy, and 29 percent to be "inferior" on risk and sustainability (World Bank 2001b, Annex D, p. 119).

Despite the extensive coverage of individual evaluations, there are only a few mechanisms for reporting on aggregate Bank outcomes and Bank performance at the corporate level. The ARDE provides an assessment of Bankwide country-, sector-, and project-level outcomes, but there is no management-initiated reporting of corporate results or overall Bank performance against outcomes. The proposed OPR will help fill this lacuna; it will address both the quality of Bank strategies and instruments (inputs) and the status of transition to results-based approaches in Bank operations (using emerging data on country outcomes). OED recommendations on the elements of a corporate reporting mechanism are summarized in box 4.2. The first OPR, covering FY03, is expected to be distributed to the Board at the end of calendar year 2003.

Management has proposed that the ARPP evolve to become the OPR. The improved annual reporting

will be based on the results of revised methodology for assessing project-level outcomes and results-based CAS assessments of output and outcome at the country level and project portfolio-level databases.<sup>11</sup>

The OPR will initially be constrained by limited information on country-level outcomes, pending the mainstreaming of the results-based CAS initiative and other management improvements to the outcome orientation of lending and nonlending instruments. It will provide a basis for linking M&E systems at different levels of the Bank. Currently, many network, capacity building, knowledge sharing, and global activities are not aligned with CAS programming, and therefore will not be captured by the aggregate reporting derived from the results-based CAS initiative. It will be necessary for the OPR to evolve to reflect a comprehensive reporting of the results of all Bank activities.

#### **Box 4.2: OED Suggestions on the Elements of an Effective Corporate Report on Results**

An effective corporate report to the Board on the results of all Bank operations should cover the following:

- Achievements of country outcomes as indicated in the CASs in terms of: (a) progress against relevant outcome indicators, (b) alignment between expected CAS outcomes and choice of instruments, and (c) the performance of the mix of instruments deployed in the country
- The adequacy, applicability, and use of individual instruments for Bank interventions: lending (investment and adjustment); nonlending; and grants and trust funds
- Achievements and progress at the sector level in relation to the proposed SSP—reflecting cross-country achievements on a sector basis
- Achievements at the corporate level; aggregation across countries and sectors that includes the contribution of DEC research and WBI training and capacity enhancement to Bank operations
- Contribution of partnerships, global programs, and corporate initiatives such as knowledge sharing to the achievement of corporate outcomes
- Linkages and alignment of M&E systems at different levels of the Bank.

## 5. Self-Evaluation in OED



This chapter reports on the status and results of OED's self-evaluation processes: the annual OED survey of clients and other ongoing self-evaluations. It also notes the progress OED has made in implementing its own recommendations in the 2002 AROE.

### RELATIONSHIP BETWEEN OED AND THE COMMITTEE ON DEVELOPMENT EFFECTIVENESS

OED is independent of Bank management and reports directly to the Board of Directors through the Committee on Development Effectiveness (CODE). OED's independence is assured through its mandate and organizational structure (box 5.1). The work program and budget are approved by the Board upon CODE recommendations, and not by Bank management. OED's most recent work program provides indicators of inputs and deliverables and was discussed by CODE on 2 April 2003.<sup>1</sup>

### RECOMMENDATIONS FROM PREVIOUS AROES

The 2002 AROE highlighted four areas for increased OED attention. Progress in these areas is described below.

**Methodology:** The 2002 AROE recommended that OED refine, elaborate, and clarify its methodology for Country Assistance Evaluations (CAEs). OED has refined the Country Information Form, but efforts to test it as a self-evaluation tool for country strategies have lagged. With the recent introduction

of self-evaluation through CAS Completion Reports, OED has a new opportunity to encourage the adaptation of its methodology for self-evaluation and for the new Country Evaluation Summaries. OED is also reviewing in detail its CAE methodology through the CAE Retrospective (box 5.2).

In line with the 2002 AROE recommendations, OED has enhanced its coverage of compliance with safeguards. All OED project evaluations now contain an annex that reports on the data and methods used, along with a discussion of their strengths and limitations.

**Participation and Consultation:** Last year's AROE recommended that OED use technology (videoconferencing and Web sites) to involve partners better in the early stages of evaluations. During FY03, almost all OED evaluations included broad consultations with external stakeholders. The studies on Global Public Programs, the CDF, and Transition Economies have used Web sites with discussion spaces and videoconferencing facilities to consult with donors (the CDF), staff in country offices, development partners, stakeholders, and external advisers.

The 2002 AROE recommended that OED institute a more transparent process for dealing with differences with management about ratings on project



### Box 5.1: OED's Independence is Assured Through Its Mandate and Organizational Structure

There are four generally accepted criteria for establishing independence of audit and evaluation units: organizational and behavioral independence, avoidance of conflicts of interest, and protection from external influence. OED meets all four criteria.

**Organizational independence:** The head of OED, the Director-General (DGO) of Operations Evaluation, reports directly and exclusively to the Board of Directors or a committee of the Board, and not to management. The Board of Directors appoints the DGO for a five-year term that can be renewed for a similar period. At the end of the tenure, the DGO is ineligible for appointment to the staff of the Bank.

**Behavioral independence:** OED has a long track record of critiquing Bank work, recognizing achievements but also addressing shortcomings and making recommendations for improvement.

**Avoidance of conflict of interest:** OED staff members are precluded from participating in evaluations of work that

they were involved in before joining OED. OED has procedures to protect against official, professional, financial, and familial conflicts. OED staff (other than the DGO) can rotate between OED and management. A comparative study of evaluation processes in multilateral and bilateral donors indicates this is good practice: such staff rotation is “desirable due to the benefit of cross fertilization of experience and knowledge” that builds relevance and substantive understanding in evaluations.

**Protection from external influence:** OED has unrestricted access to all Bank records and staff in conducting its work, and can consult with government officials, the private sector, nongovernmental organizations, the media, and all sources it deems appropriate without prior clearance from Bank management. OED controls the hiring, promotion, and termination of its staff within the framework of the Bank's personnel system, although the networks do clear hiring and promotion decisions for staff mapped to them.

*Source:* North 2003; OED 2003f.

evaluations. OED has made limited progress over the last fiscal year. As OED starts to rate country outcomes and quality of CAS Completion Reports in its Country Evaluation Summaries, there remains a need for a transparent process to deal with differences in project and country-level ratings.

**Dissemination:** The survey of key client audiences for sector and thematic evaluations indicates that 37 percent reported having read the report; a further 9 percent reported being aware of the report's findings, although they did not read the report, and the balance, 54 percent, reported not having read the report or being aware of its findings. Of the respondents who had not read the evaluation, 60 percent said they were unaware of it. As recommended in the 2002 AROE, all major OED evaluations now prepare a brief description of the main target audience and dissemination plan at the approach paper stage.<sup>2</sup> But OED must be proactive to better target and engage interested constituencies in the Bank on study findings.

**Self-Evaluation:** The 2002 AROE called for continued emphasis on periodic self-evaluation through surveys. This year, OED has complemented its annual survey with detailed reviews in two areas: a CAE Retrospective (box 5.2) and a review to enhance the use of the Web site as a tool to serve OED's business strategy (box 5.3).

### OED SURVEYS: PARS AND SECTOR AND THEMATIC STUDIES ARE THE FOCUS THIS YEAR<sup>3</sup>

Consistent with its role as independent evaluator of Bank products and services, OED evaluates the quality, relevance, and efficacy of its own products and services each year. This year's surveys targeted a relevant audience—the evaluatees—for five sector and thematic studies issued since last year and task managers on FY02 Project Performance Assessment Reports.<sup>4</sup> The response rate was 38 percent

### **Box 5.2: Forthcoming OED Retrospective on Country Assistance Evaluations**

The CAE is one of OED's major products. To assess its strengths and weaknesses, the 2002 AROE surveyed country teams involved in CAEs completed as of December 31, 2001. Based on these results, OED is undertaking a more detailed assessment of CAE *methodology, utilization, and lessons learned* through a CAE Retrospective that will be presented to CODE in FY04. It will validate and, where appropriate, suggest modifications of measures and standards to ensure quality, reliability, and consistency. Contentious issues such as ratings of outcome, relevance, and effectiveness, as well as the outcome of Bank assistance, will be addressed. The retrospective will also draw lessons from the whole cohort of CAEs on the use of economic and sector work; sequencing of reform; choice of instruments; and lessons on ownership, participation, and partnerships. It will also assess how the CAEs are used by the Board in guiding management, by Bank staff, and by development partners.

(higher than in FY02, when the response rate was 25 percent).

Where the quality of sector and thematic studies was rated positively, most respondents reported that the evaluations provided a good context of sector issues, relevant options for sector strategy and Bank programs, and useful and relevant lessons. Where studies were rated poorly, respondents reported the need to improve and clarify methodological aspects: analysis, evidentiary base, and the link of these elements to the conclusions, and the need to engage in greater consultation with borrowers and Bank staff. PPARs received high ratings on all aspects of quality that reflected perceptions of relevant lessons and good ideas for future project design.

Based on the results of the survey, OED should clarify and better present its methodology, including its evidentiary base for sector and thematic studies. OED should better target and disseminate its sector and thematic as well as its project evaluations to Bank staff, and enhance consultation with its borrowers. Annex D details the survey results.

### **GEOGRAPHIC FOCUS OF OED EVALUATIONS**

The AROE analyzed OED's geographic coverage during the last few fiscal years in order to assess whether the country distribution of its activities is aligned with the pattern of the Bank's lending activities (table 5.1).

### **Box 5.3: Using OED Web Sites to Enhance OED's Impact**

OED has obtained valuable feedback on its Web services from two sources. The 2002 AROE surveyed the use and impact of OED's Web services, including documents, project ratings, and the Help Desk. Of the 320 respondents who used the services, 94 percent found them to be relevant to their work. But of the 44 percent who did not use them, almost half were not aware of the services.

The Bank's Information Solutions Group instituted an independent content review of all vice-presidential-unit Web sites, including OED's, in February 2003. It concluded that as an independent evaluator within the Bank, the "Operations Evaluation Department faces a unique challenge among Bank sites, but meets it well." The site is praised for establishing "its credibility quickly with visitors" and "easy access to the unit's products and a clear explanation of the unit's policies and procedures." The OED Web site on partnerships is noted as best practice. Most of the review's recommendations concern presentation enhancements and minor refinements to navigational elements, and these have already been adopted.

Based on these reviews, OED's focus now is to enhance the use of the internal and external Web sites. It is compiling Web statistics on what information is being used most often and by whom. This will provide baseline data against which to measure future patterns of access and use of Web-based services. The results of this review, along with planned improvements, will be reported and tested through surveys of Bank staff in future AROEs.

The country distribution of OED's work is largely consistent with patterns of operational activities, as measured by both the number of exiting projects and the share of lending. It reflects the need for specific evaluative information and building blocks for larger sector/thematic and Country Assistance Evaluations. Numerous project evaluations were done in Brazil, Bulgaria, China, India, Indonesia, Tanzania, and Vietnam because they were all subjects of CAEs.

Mozambique received intense attention because it was the subject of a planned CAE that has been replaced with the pilot CASCR and CES. Guinea and Madagascar, with seven and eight evaluations respectively, were the only countries where the number of OED evaluations over-represented the exiting projects or share of lending. In Madagascar, a CAE is scheduled for FY05.

**Table 5.1: Geographic Focus of OED Evaluations (FY99–02)**

Country	Number of OED evaluations <sup>a</sup>	Percent share of project exits	Percent share of cumulative lending	Remarks
India	26	4	12	CAE (FY01)
China	22	5	7	CAE (expected in FY04)
Indonesia	18	4	6	CAE (FY99)
Brazil	18	4	6	CAE (FY03)
Mozambique	15	2	< 1	Planned CAE replaced with CES
Ghana	12	2	1	Proportionate
Uganda	11	2	1	CAE (FY00)
Turkey	10	2	4	Proportionate to project exits, share of lending
Bolivia	10	1	< 1	CAE (FY99)
Madagascar	8	1	< 1	Proposed CAE FY05
Bulgaria	8	1	< 1	CAE (FY02)
Guinea	7	1	< 1	
Pakistan	6	2	3	Proportionate to project exits, share of lending
Colombia	6	< 1	2	Proportionate to share of lending
Tanzania	6	1	1	CAE (FY00)
Vietnam	5	< 1	1	CAE (FY02)

a. Includes 283 PPARs and 141 country case studies for sector and thematic evaluations, corporate evaluations, and CAEs over the four-year period.

## 6. Conclusions and Recommendations



The primary objective of monitoring is to inform management decisionmaking. That of evaluation is to enhance accountability and learning. The two elements are complementary because the quality of evaluation can only be as good as the information generated from the monitoring process. In the same way, the quality of evaluation will shape the adequacy of future monitoring.

### CONCLUSIONS

The Bank has been working to establish an M&E system at different levels and for different instruments. By and large, there is a clear and well-established framework for M&E for individual lending operations; it is less developed and less effective for non-lending services and at the sector level. The *Managing for Results* initiative shifts the burden of accountability from the instrument to the country level. This is expected to produce trickle-down effects and incentives for strengthening the framework for M&E of outcomes at the instrument and sector levels. It does not, however, address gaps in M&E of corporate initiatives such as knowledge sharing, Regional capacity building, and global programs.

Three overarching aspects in M&E need attention:

- **Ensuring Evaluability:** For all instruments and at all levels, a priority is to establish clear and realistic expected outcomes and to support monitorable indicators, baseline data, and benchmarks of progress. This is particularly urgent for adjustment lending, nonlending services, sector strategies, and knowledge initiatives. The Bank's results-focused CAS is an attempt to enhance the evaluability of the Bank's country-level assistance. The Bank's quality assurance process at entry should widen its focus to include evaluability.
- **Monitoring and Reporting on Progress:** The monitoring framework for reporting on implementation is well established for investment projects. The current priority should be to improve the quality and realism of reporting. The monitoring framework to track and report on outcomes is less well established for adjustment lending, trust funds, and grants. These latter two have recently been areas of management attention, and it is too early to comment on the effectiveness of the improvements.<sup>1</sup> The framework at the country level is under development. Monitoring and reporting on implementation of sector strategies is ad hoc and needs strengthening.
- **Filling Gaps in Self- and Independent Evaluation:** The self- and independent evaluation framework for projects is well established and fairly effective. The recently introduced CAS Completion Reports will close an important gap in self-evaluation at the country level. Gaps remain in the self-evaluation for nonlending activities, sector strategies, trust funds, grants, and knowledge

activities, and at the corporate level. Independent evaluation of sector strategies and global programs also needs strengthening.

The 2002 AROE suggested the formulation of a comprehensive Operational Policy and accompanying Bank Procedures to formalize an overarching framework for self- and independent evaluation and mandate a results focus for Bank operations. OED and management have been working over the last year and have agreed on a draft Operational Policy that (a) establishes the mandate for self- and independent evaluation; (b) identifies roles and responsibilities at the different levels; and (c) suggests methods for reporting on compliance. The accompanying Bank Procedures statement is still being prepared. It will provide the framework for incorporating evaluation policies and procedures into the Bank's core operational work and include one on the *Generation, Dissemination, and Utilization of the OED Findings*. This work must be expedited and brought to a swift conclusion.

Two important requirements for a stronger results focus are staff capacity and resources. On resources, the *2002 Annual Review of Development Effectiveness* estimated that the Bank's direct costs on country monitoring, formulating and monitoring CASs, and providing global monitoring services amounted to about 2 percent of total direct costs in recent years.<sup>2</sup> A recent strategy update estimates the resource implications of the Managing for Results initiative to be an additional \$5.7 million.<sup>3</sup> As the Bank implements results-focused initiatives, it must examine the adequacy of its review and reporting processes to the Board on the utilization of resources for M&E.

On staff capacity, QAG has proposed that the Bank use a core set of experienced professionals to help units with technical advice to establish clear outcomes, monitorable intermediate indicators, and evaluation systems. Aspects such as knowledge shar-

ing, risk assessment, and M&E must be built in as an essential part of task management at any level. While there is a role for technical advice, the emphasis should be on training and building core competencies in M&E into the job descriptions for all project/task and unit management positions, as well as into performance evaluation systems.

## RECOMMENDATIONS

This report makes the following recommendations to management and OED:

### *Management*

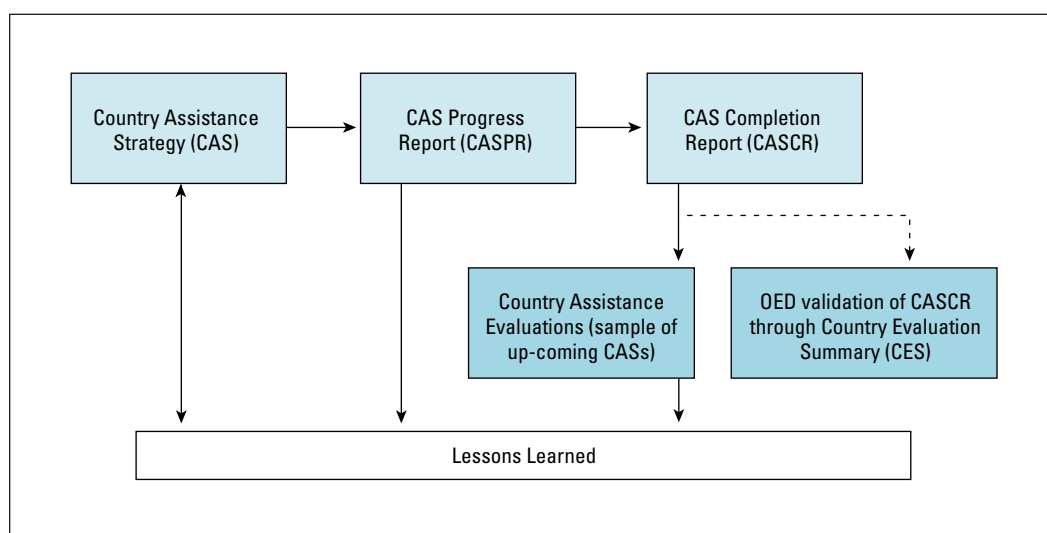
- (a) Issue an Operational Policy/Bank Procedure that sets out the mandate, framework, roles, and responsibilities for self- and independent evaluation.
- (b) Mainstream greater outcome orientation in CASs and a self-evaluation framework for Bank assistance at the country level.
- (c) Strengthen the M&E of the development outcomes of trust-funded activities.
- (d) Strengthen the outcome orientation and self-evaluation of nonlending activities, in aggregate and at the task level, either individually or in clusters.

### *OED*

- (a) Enhance borrower consultations on sector and thematic studies and PPARs.
- (b) Improve internal dissemination of sector and thematic studies through better targeting and more active participation in sector and Regional events.
- (c) Institute a transparent procedure to resolve differences in ratings between OED and management for country- and project-level evaluations.

## Annex A Monitoring and Evaluation Framework at the Country Level

The M&E architecture at the country level is depicted below:



The following table summarizes the purpose, role, and use of the various elements in the M&E framework at the country level.

Document	What is it?	Who prepares it?	When? How often?	Time covered	Ratings
Country Assistance Strategy (CAS)	Describes the Bank's strategy, based on an assessment of country priorities. Indicates the level and composition of assistance to be provided and country's portfolio performance. Main tool for management and Board to review Bank assistance in a country.	Country Team, with government and other stakeholders OPCS reviews	Every 3-4 years	3 years	Outcome matrix
CAS Progress Report (CASPR)	Reports to the Board on implementation of the CAS program and key developments since the last CAS.	Country Team, OPCS reviews	Mid-CAS cycle, delay in CAS	1.5–3 years	No ratings
<b>NEW</b> – CAS Completion Report (CASCR)	Assesses outcomes achieved relative to CAS expectations. Comments on the relevance of Bank programs to the country's development agenda, quality of portfolio, and measures of development effectiveness. Draws lessons for next CAS. Includes insights from clients.	Country Team, OPCS reviews	End of each CAS period (for selected CASs in the pilot stage)	3 years	No ratings in the pilot phase
<b>NEW</b> – Country Evaluation Summary (CES)	Reviews and validates the self-assessment of country outcomes as stated in the CASCR.	<i>OED, for each CASCR</i>	After preparation of CASCR	3 years	Contains ratings
Country Assistance Evaluation (CAE)	Provides the Board an independent assessment of the outcome of Bank assistance (strategy, lending, and nonlending activities) in client countries. Informs the new CAS on outcomes and performance of prior CASs.	<i>OED, for sample of upcoming CASs</i>	Parallel to a sample of upcoming CASs; about 6 annually.	Usually a decade (2-3 CAS periods)	Contains ratings



There are several units in the Bank involved in the M&E of Bank operations that are distinct from the units that focus on financial management and control over the Bank's administrative resources.

The following table highlights the roles and the distinct mandates of the various units in M&E and control activities. OED is the only unit with a mandate to undertake independent evaluation of Bank operations.

**Table B.1: Roles and Mandates, M&E and Quality Control**

	Reporting to the president	Reporting to the Board
	Monitoring and self-evaluation (management)	Independent evaluation (OED)
<b>Bank operations</b>		
* Projects	Quality Assurance and Compliance Unit (QACU) <sup>a</sup> Quality Assurance Group (QAG) * Regional quality teams	OED (ES and PPAR)
* ESW	QAG * Regional quality assurance teams OPCS Network <sup>b</sup>	OED (CAEs and sector/thematic evaluations)
* Sector / theme	QAG (proposed) * Network and sector anchors	OED (sector/thematic evaluations and sector implementation reviews)
* Country	* Regional front offices * Network and sector anchors OPCS Network	OED (country evaluation summaries and CAEs)
External claims of noncompliance	Inspection Panel	

\* Report to the president through vice-presidents.

a. QACU ensures compliance with safeguards for projects at entry and during supervision.

b. OPCS undertakes cross-sector and cross-Regional evaluations at all levels of activity (e.g., CAS Retrospective at country level, SSP Stocktaking at sector level, Adjustment Lending Retrospective at instrument levels).





## Annex C

# OED Review of Rating Changes and Quality of Implementation Completion Reports

Upon project completion, the task leader prepares an Implementation Completion Report (ICR) and Regional management approves it. The ICR marks the transition of a Bank loan from implementation in the portfolio to exit. OED conducts an independent review of each ICR, preparing a summary for internal use that records the validation of the self-ratings and self-assessment based on information provided in the ICR and other evidence.

This annex provides these kinds of information for projects exiting during FY96–02. The first section provides a summary of ICR evaluations conducted by OED during FY96–02. The second describes major trends in ICR quality for the period, and the third shows changes in ratings at different stages of the project cycle.

### NUMBER OF OED EVALUATIONS, FY96–02

Table C.1 summarizes OED’s output of ICR reviews and Project Performance Assessments Reports (PPARs) since FY96. Because of bunching of ICR deliveries to OED toward the end of the fiscal year, the number of ICRs evaluated in some years exceeds the number of ICRs received, as in FY98 and FY02.<sup>1</sup>

### FINDINGS ON ICR QUALITY

OED’s summary ratings provide an independent assessment of the quality and consistency of the content and management ratings in the ICR. OED assesses three dimensions of ICRs: the quality of the

**Table C.1: OED Reviews of ICRs and PPARs since FY96**

	OED evaluations of ICRs, FY96–02						
	FY96	FY97	FY98	FY99	FY00	FY01	FY02
ICRs received	268	397	218	291	307	317	267
Adjustment	43	65	24	45	36	45	28
Investment	224	332	186	236	251	275	239
ICRs evaluated	249	313	284	268	313	288	287
Adjustment	44	45	38	36	31	28	43
Investment	204	268	246	232	245	260	244
PPARs	100	79	71	69	76	73	70
Adjustment	17	13	13	14	14	14	13
Investment	83	66	58	55	62	59	57
Percent of projects assessed by PPARs in that year <sup>a</sup>	40	25	25	26	24	25	24

Source: OED data.

a. The drop in the number of PPARs after FY96 reflects an agreement between OED and the Board to reduce coverage from 40 percent to 25 percent to allow greater OED focus on sector and thematic and country evaluations. These numbers exclude PPARS conducted for GEF projects.

analysis, which includes coverage of important subjects, analytical soundness, adequacy of evidence and lessons learned; information on the sustainability or future operations of the project, including a plan for operation as well as related performance indicators and M&E; and borrower and co-financier inputs. The evolution in quality of ICRs by dimension is summarized in table C.2.

*The overall quality of ICR analysis remains high.* As shown in table C.2, ICR quality ratings since FY99 have remained fairly constant at around 94 percent. Almost all ICRs covered important topics and drew appropriate lessons. These quality ratings reflect better understanding of ICR guidelines and methodology by operational staff.

*The quality of ex-post economic analysis and evidence presented in ICRs can be improved.* For operations where it was relevant, the quality of ex-post analysis and the extent to which evidence presented in ICRs is complete and convincing deteriorated in FY02.

*The poverty analysis component of ICRs has improved, but remains weak.* At only 67 percent satisfactory, this dimension was added in 1997 and has been rated only where relevant, for a relatively small number of ICRs (17 percent in FY00, 18 percent in

FY01, 18 percent in FY02, and 14 percent in FY03's partial sample). As the Bank operationalizes the Managing for Results initiative and the MDGs, OED hopes that there will be increased attention to realistic and measurable poverty dimensions, where applicable, in lending operations.

Forward-looking orientation has improved, although there is room to improve the plan for M&E of future project operations. ICRs are expected to report on a project's "future operations" or the anticipated results after disbursements, in the operational phase, to ensure that planned benefits of the operation materialize and negative externalities, such as environmental impact, are minimized. Although descriptions in general terms of the plan for future operation have improved to 93 percent satisfactory or better, these ratings were made on 59 percent of the ICRs that were reviewed in FY02 (66 percent in FY01); the remaining ICRs were not rated on this dimension because of a lack of information on future operations. On the two other aspects of future operations (performance indicators and M&E plan), ICRs are less satisfactory, both in terms of the proportion of the ICRs that could be rated and the quality of those that were rated. Performance indicators for future operation were rated in 51 percent of all ICRs

**Table C.2: Evolution in ICR Quality**

	Trends in ICR quality on three dimensions (percent satisfactory or better)						
	FY96	FY97	FY98	FY99	FY00	FY01	FY02
<b>1. Analysis</b>							
Coverage of important subjects	91	96	95	95	95	94	95
Ex post economic analysis (if applicable)	71	78	73	88	86	86	82
Internal consistency	92	93	96	94	95	95	95
Evidence complete/convincing	82	84	89	86	87	87	84
Adequacy of lessons learned	92	93	96	90	93	92	92
Poverty analysis	n/a	n/a	n/a	n/a	54	48	67
<b>2. Future operation</b>							
Plan for future operation	82	84	81	89	86	86	93
Performance indicators for operational phase	63	74	72	76	81	78	84
Plan for M&E of future operation	69	63	66	70	73	73	76
<b>3. Borrower/co-financier inputs</b>							
Borrower input to ICR	91	93	97	96	98	98	95
Borrower plan for future project operation	81	84	83	86	89	87	91
Borrower comments on ICR	88	89	94	95	97	97	95
Co-financier comments on ICR (if applicable)	57	73	80	86	92	89	88
Overall quality of ICR	91	95	96	94	96	94	94

Source: OED data.

reviewed in FY02 (56 percent in FY01). Of those that could be rated, the proportion improved from 63 percent satisfactory or better in FY96 to 84 percent in FY02. The plan for M&E could be reviewed in only 36 percent of the ICRs in FY02 (38 percent in FY01), and of these, the proportion rated satisfactory or better improved modestly, from 73 percent to 76 percent.

*Borrower and co-financier contributions have been included in the ICRs where they are available.* ICRs are supposed to include the borrower's inputs and specific comments on the assessment, and the borrower's plan for future operation. Many ICRs lack one or more of these components. ICRs included borrowers' inputs and borrower comments about 75 percent and 68 percent of the time, respectively, between FY01 and FY02, and the quality of these comments has remained steady at around 95 percent. The plan for future operation was included in about 46 percent of the ICRs, and of these, the proportion rated satisfactory or better improved from 81 percent in FY96 to 91 percent in FY02. ICRs are also supposed to include co-financiers' comments where applicable. The quality of these comments has improved. Overall, the quality trend for borrower and co-financier contributions suggests that the ICR process has become participatory over time.

### RATINGS CHANGES<sup>3</sup>

At each stage of the project cycle, implementation, completion, and independent evaluation by OED, projects are rated on outcomes—the likelihood of achieving their development objectives.<sup>4</sup>

There are two ways to test the reliability of indicators at each stage: extent to which ratings are changed in subsequent stages and net change in ratings, that is, percentage of ratings that are decreased minus those that are increased. The latter shows the degree of over-optimism (if net change is positive) or underestimation (if net change is negative) of the self-rating.

Table C.3 shows rating changes **on outcome** by exit fiscal year. Outcome rating changes are included here only where the change is from satisfactory (highly satisfactory, satisfactory, marginally satisfac-

tory) to unsatisfactory (marginally unsatisfactory, unsatisfactory, highly unsatisfactory) or vice-versa. Rating changes within “satisfactory” and “unsatisfactory” categories are not included. Table C.3 includes rating changes on outcomes between:

- (a) Last Project Supervision Report (PSR) before operation exits and ICR
- (b) The PSR, ICR, and OED's summary
- (c) The OED summary and the Project Performance Assessment Report (PPAR), if available.

Looking first at changes between the final PSR rating and the ICR rating, there has been a marginal downward trend in the number of total rating and net rating changes in FY02. The net downgrade in FY02 is 6 percent. Next, the total rating changes on outcome between project completion reporting (ICR) and OED's summary has also increased marginally, from 5 percent on average between FY99–01 to 6 percent in FY02. Thus, the PSR and ICR ratings have been less realistic in FY02 compared with FY01.

The rating changes between PSR and OED summary show the degree and direction of differences between outcome ratings during supervision and by OED on completion. This is relevant as it indicates the extent to which PSR ratings reported in the ARPP may be reliable predictors of outcomes. The net rating changes between the PSR and OED summary have increased from 8 percent during FY99–01 to 12 percent in FY02.

Changes in outcome ratings made *within OED*, between OED summary and PPAR, have been large. Between FY96 and FY99, outcome ratings were changed in 14 percent of operations. The total rating changes of 17 percent in FY02 reflect increases in both downgrades (11 percent in FY02 from 8 percent in FY01) and upgrades (6 percent in FY02 from 4 percent in FY01).

Overall, these patterns in rating changes suggest that the Regions are more optimistic about development effectiveness than OED, and that relying on the PSR alone may result in an overly sanguine picture. The changes in ratings between OED summary and PPAR also underscore the need for detailed independent assessments of lending operations through PPARs

**Table C.3: Summary of Outcome Ratings Changes, as Percentage of Total Projects Reviewed at Each Stage, FY96–02 Exits**

	Change			
	PSR to ICR	ICR to ES	PSR to summary	Summary to PPAR
<i>FY96</i>				
Down (a)	8	8	16	10
Up (b)	4	1	4	3
Total (a+b)	12	9	20	13
Net (a-b)	4	7	12	7
<i>FY97</i>				
Down (a)	9	5	9	9
Up (b)	2	3	0	0
Total (a+b)	11	8	9	9
Net (a-b)	7	2	9	9
<i>FY98</i>				
Down (a)	11	4	14	8
Up (b)	2	1	3	7
Total (a+b)	13	5	17	15
Net (a-b)	9	3	11	1
<i>FY99</i>				
Down (a)	7	5	11	6
Up (b)	3	0	2	8
Total (a+b)	10	5	13	14
Net (a-b)	4	5	9	-2
<i>FY00</i>				
Down (a)	6	5	10	10
Up (b)	2	0	2	10
Total (a+b)	8	5	12	20
Net (a-b)	4	5	8	0
<i>FY01</i>				
Down (a)	8	5	10	8
Up (b)	1	1	3	4
Total (a+b)	9	6	13	12
Net (a-b)	7	4	7	4
<i>FY02</i>				
Down (a)	8	6	13	11
Up (b)	2	0	1	6
Total (a+b)	10	6	14	17
Net (a-b)	6	6	12	5

Source: OED data.

Since 1999, OED has undertaken annual surveys of clients to gain their perceptions of the quality and impact of OED evaluations, as background for the *Annual Report on Operations Evaluation*. The intention is to assess the influence of OED evaluations on the Bank's program and activities. This year's survey addressed two kinds of OED products and their respective clients:<sup>1</sup>

- Five sector and thematic (S&T) evaluations (cultural properties, indigenous people, social funds, urban, and water resources): aimed at staff associated with the themes and sectors being evaluated.
- Project Performance Assessment Reports (PPARs): aimed at two sets of clients:
  - (a) Task managers whose projects were approved in FY02 (for their views on the use and learning value added from PPARs in the design of new projects)
  - (b) Task managers and sector directors of projects that were assessed through PPARs in FY02 (for their views as evaluatees).

Some caveats apply to the results of these surveys. First, in addition to the Bank staff groups that were targeted by this year's survey, OED's sector and thematic evaluations and PPARs have several other audiences both inside and outside the Bank, including the Board. Second, each relevant thematic group was surveyed only once, even if it was relevant to more than one evaluation being surveyed. Third, the results of these surveys are *perceptions* of Bank staff rather than a direct measure of actual quality or impact of the OED evaluations. Future AROEs will broaden the evidence base and triangulate data from multiple sources on S&T studies through tracer studies and expert panel reviews. Finally, to the extent

that some of the survey respondents are evaluatees, they may reflect evaluation bias.

The design of surveys, selection criteria for specific OED products to be included in the study, and procedures for sample selection are available on request. The following is a synopsis of the results. This year's results are compared with results from previous years and similar types of respondents, where possible.<sup>2</sup>

### RESPONSE RATES

The surveys were sent electronically to 1,172 staff over a 15-day period in March 2003. The response rates ranged from 33 percent to 43 percent and averaged 38 percent across all surveys (compared with 25 percent for comparable surveys in 2002<sup>3</sup>). The improved response rate may be a result of the use of an external tabulation firm and the resulting perception of greater anonymity, as well as better targeting of respondents.

### READERSHIP

Of the respondents to the PPAR survey, 73 percent have read at least one PPAR. Thirty-seven percent of the thematic group members surveyed about OED's S&T evaluations had read the report. Similar to 2002, a further 9 percent reported familiarity with the report's main findings (figure D.2).

Of the respondents who had not read PPARs or S&T studies, about two-thirds of the group in each case said they were unaware of the reports. Another 28 percent reported lack of time as the reason for not

having read the reports. Given that these reports are available on the Bank’s external site (<http://www-wds.worldbank.org/>) as well as on the OED Intranet site, proactive and better-targeted dissemination by OED is called for. For PPARs and the S&T studies taken together, half the respondents reported having read the whole report. Only about 10 percent reported reading just the executive summary.

### QUALITY OF OED EVALUATIONS

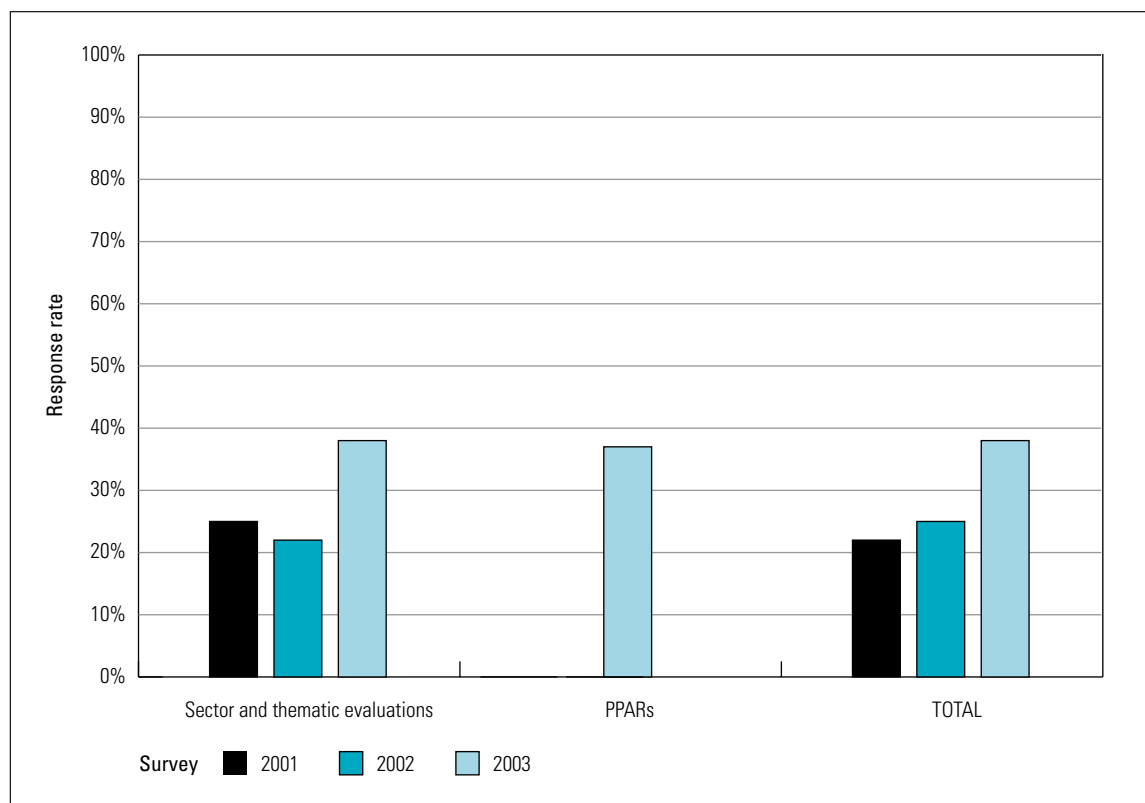
Respondents were asked to comment on several dimensions of the quality of the evaluations: timeliness, objectivity, methodological soundness, relevance to the work, and overall quality; for S&T studies alone, respondents were also asked to comment on the quality of the recommendations.<sup>4</sup> As shown in figure D.3, over 85 percent of respondents reported S&T evaluations were timely and relevant, while only 49 percent opined they were methodological-

ly sound.<sup>5</sup> OED was rated to be objective by 59 percent of respondents, a lower percentage than in previous years.

Where the quality of S&T studies was rated “good or excellent,” most respondents reported that the evaluations provided good context of sector issues, relevant options for sector strategy and Bank programs, and useful and relevant lessons. Where it was rated poorly, they reported the need to improve and clarify methodological aspects: analysis, evidentiary base, and their link to the conclusions, as well as greater consultation with borrowers and Bank staff.

PPARs received high ratings on all aspects of quality, as shown in figure D.4. The positive ratings reflect perceptions of relevant lessons and good ideas for future project design. Although the respondents from the evaluatees group (task managers and sector directors of evaluated projects) rate methodology lower than the respondents from the learning group (task managers who worked on recently approved projects), this difference is not statistically significant.

**Figure D.1: Response Rates**



## INFLUENCE OF OED EVALUATIONS

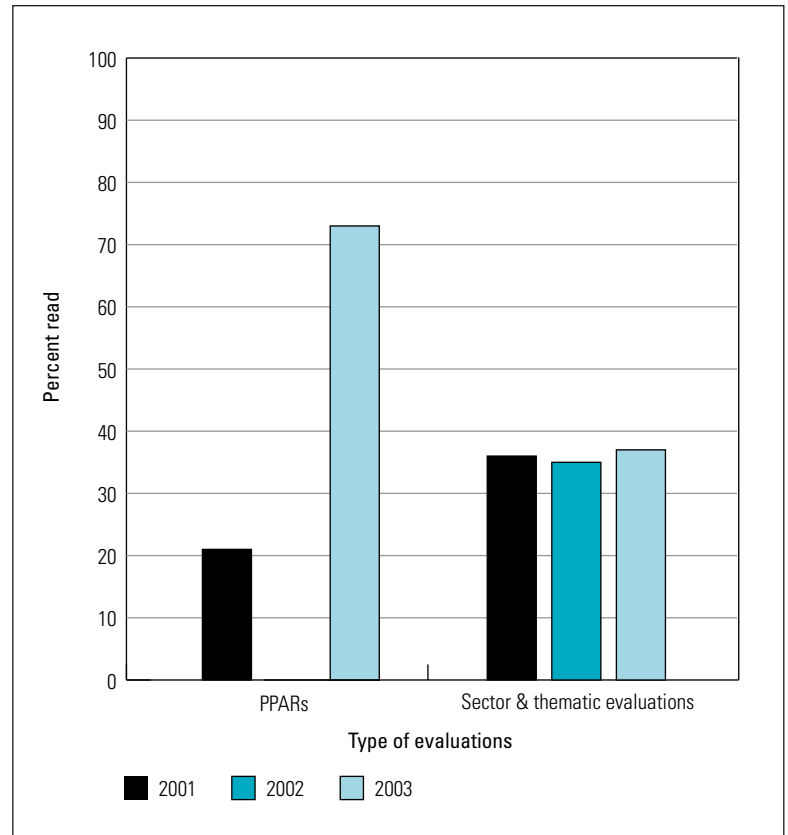
The S&T surveys asked whether the evaluation influenced the respondent's view of development priorities and strategy and whether they influenced the Bank's strategy, lending services, or nonlending services.<sup>6</sup> Among respondents to the S&T survey, over half reported that their own views were influenced either somewhat or strongly by the concerned evaluation, compared with three-quarters in 2002 (figure D.5). The influence of OED on sector strategies, and particularly nonlending services, appears to be high. The PPAR survey respondents were asked if the project evaluations influenced the relevance, design, and implementation arrangements for similar projects or programs. About 70 percent of the respondents reported that the influence was positive (strongly or somewhat) on project relevance and design and 62 percent reported it to be positive on implementation arrangements (figure D.6).

## RESULTS BY DEMOGRAPHIC PROFILE

OED checked for differences in perceptions of OED work by staff level (manager versus non-manager) for S&T evaluations and by location (headquarters versus field-based) for both S&T and PPAR surveys. The findings are below:<sup>7</sup>

- On S&T studies, field-based staff rated OED significantly higher on relevance of the evaluation to their work and on its influence over Bank strategy and their own views of strategy and policy than staff based in headquarters.
- Non-managerial staff rate relevance of S&T studies to their work significantly higher than managerial staff.
- Field-based staff rated PPARs significantly higher on several respects of quality: methodological soundness, relevance to work, overall quality, and

**Figure D.2: Readership of OED Evaluations**



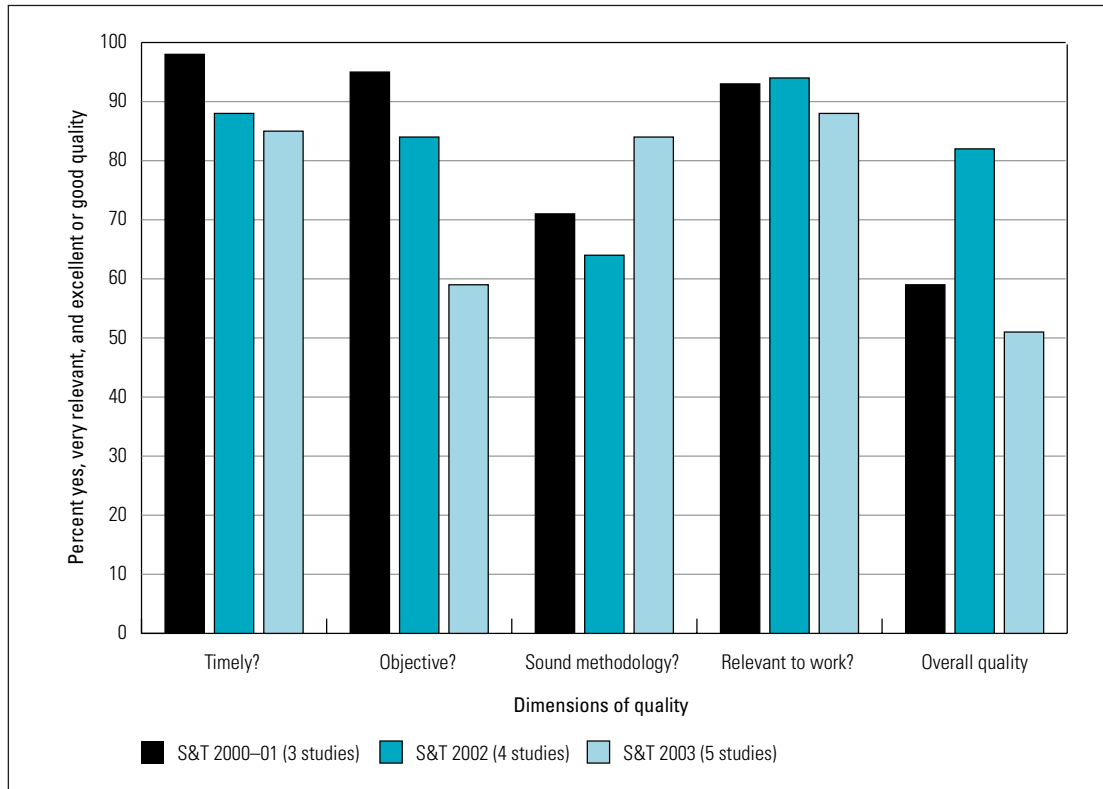
influence on the design and implementation of future lending operations.

## RECOMMENDATIONS FROM RESPONDENTS

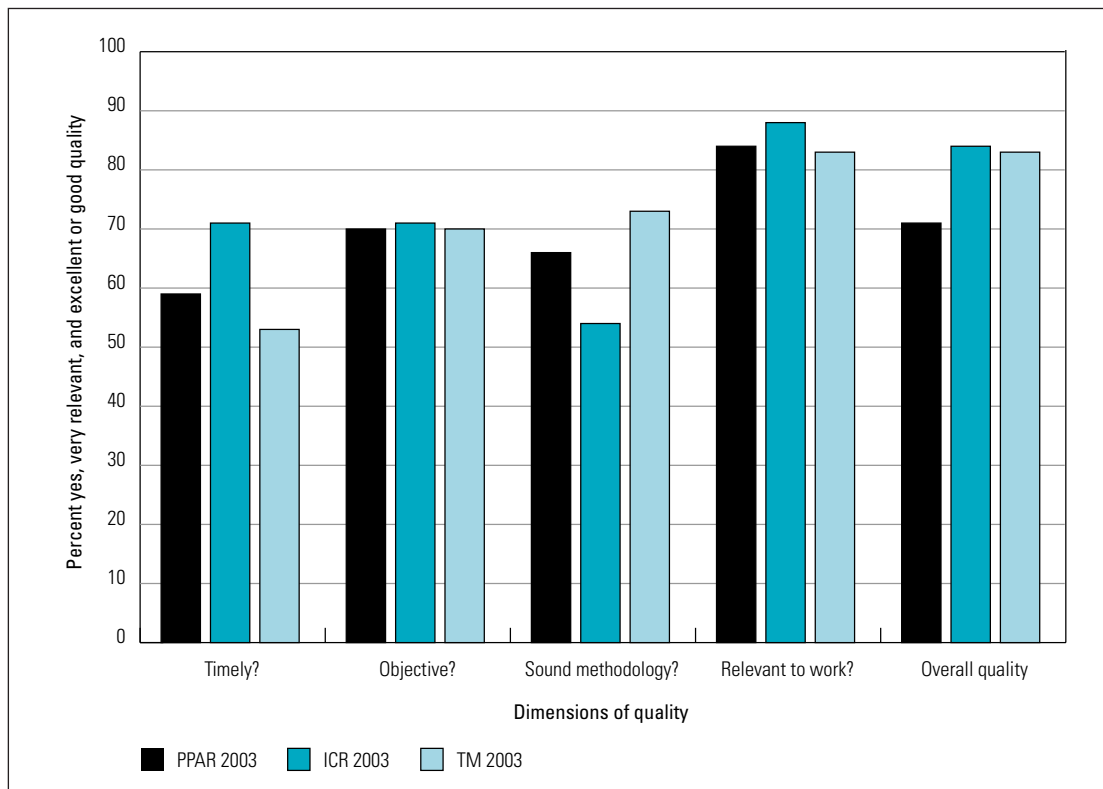
All surveys asked respondents how they would suggest OED improve the quality of its products and the respondents were provided a list of recommendations they could check. As shown in figure D.7, the most frequently checked recommendations among all respondents were for OED to broaden both internal and external consultations, particularly with borrowers, and to widen dissemination.



**Figure D.3: Quality of OED Sector & Thematic Evaluations**



**Figure D.4: Quality of PPARs**



## IMPLICATIONS FOR OED

The results above point to two areas of improvement for OED—dissemination and methodology.

To improve dissemination, OED should:

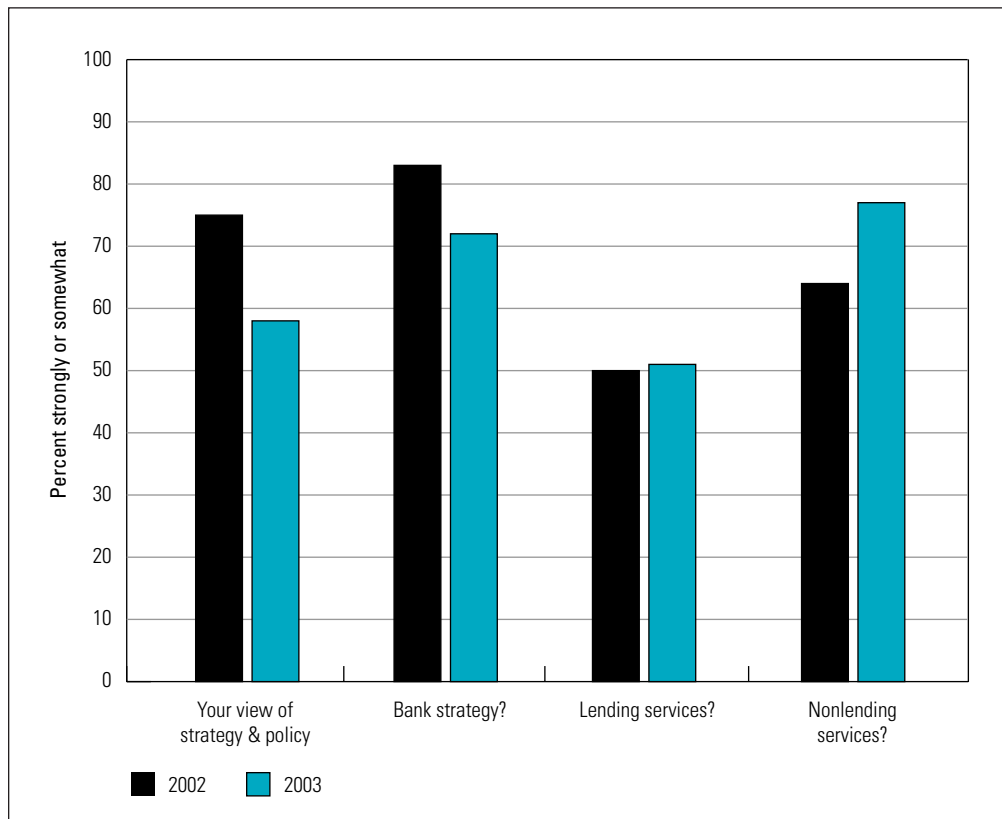
- Engage thematic groups and networks in identifying how OED products can most effectively reach interested staff.
- Enhance the range of dissemination activities such as presentations, brown bag lunches, and

workshops, in addition to document distribution via e-mail.

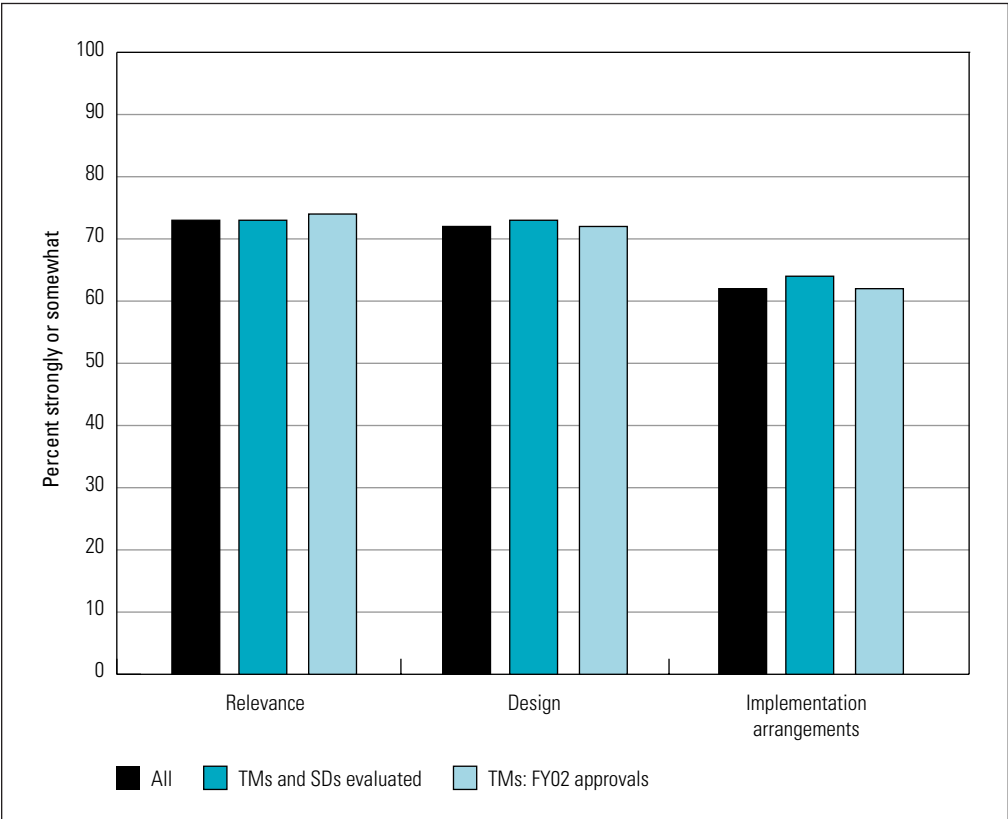
To improve methods, OED should:

- Improve the consultation process within the Bank.
- Improve consultation with stakeholders, especially borrowers.
- Strengthen and clarify methodology for sector and thematic evaluations.

**Figure D.5: Influence of Sector and Thematic Studies**

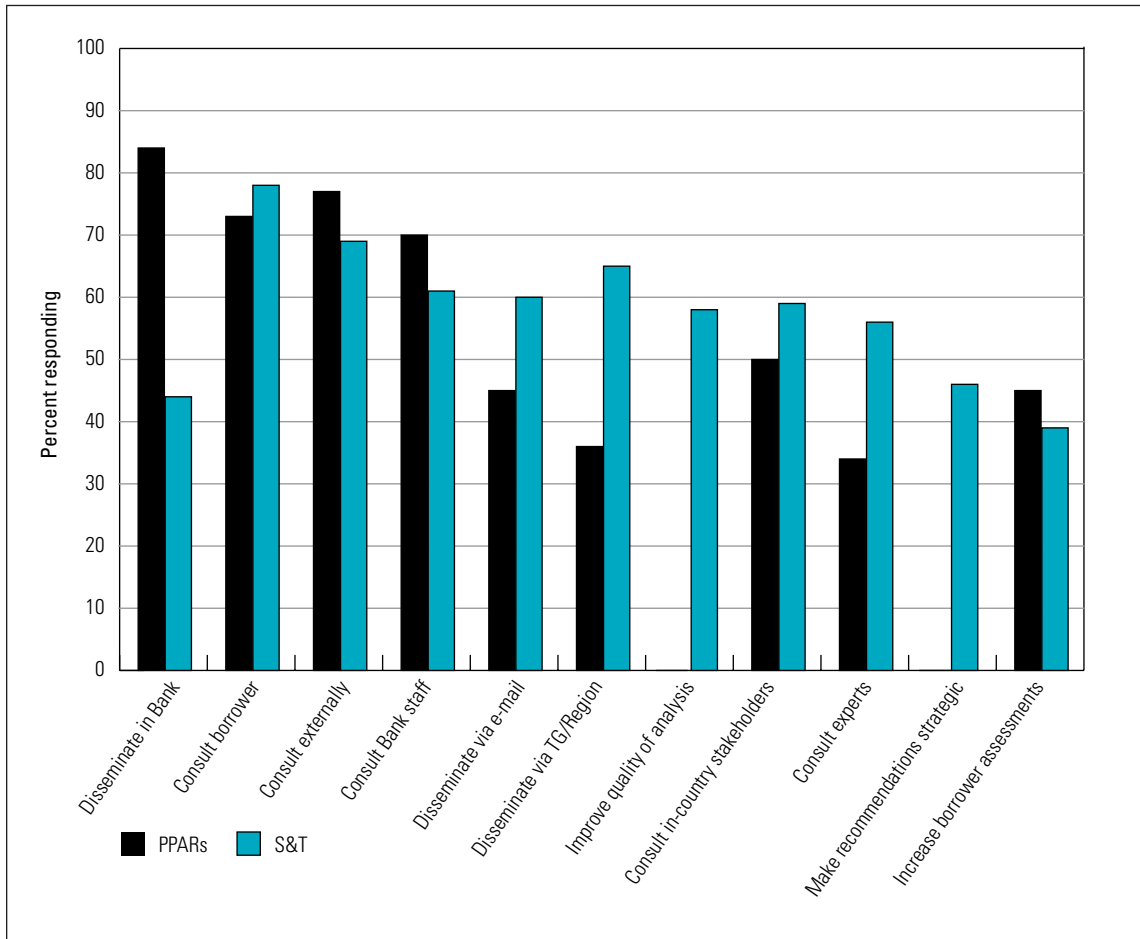


**Figure D.6: Influence of PPARs**



Note: TM = task manager; SD = sector director.

**Figure D.7: Recommendations to OED**





- AAA Analytical and Advisory Services**  
Includes standard economic and sector work, as well as less formal instruments such as Policy Notes, and non-ESW activities including workshops, seminars, conferences, and the like.
- ACS Activity Completion Summary**  
A completion report and a self-evaluation mechanism required for all AAA tasks over \$50,000. It summarizes the extent to which task objectives of each activity were achieved, assesses the process and management of the task, and provides client views based on a survey. Prepared for all tasks at the time of their closing and within 6 months of the delivery to the client.
- APL Adaptable Program Loan**  
An investment-lending instrument that provides phased support for long-term development programs. An APL involves a series of loans that build on the lessons learned from the previous loan(s) in the series. APLs are used when sustained changes in institutions, organizations, or behavior are key to successfully implementing a program. They can be used to support a phased program of sector restructuring, or systemic reform in a qualified sector (power, water, health, education, and natural resource management).
- ARPP Annual Review of Portfolio Performance**  
A QAG report that provides an annual strategic overview to the Board and senior management on the status of Bank lending and nonlending operations, the impact of actions taken to improve them, and the likely evolution of the portfolio in the future.
- CAS Country Assistance Strategy**  
Prepared by the Bank's country team, in collaboration with the government and other stakeholders, the CAS describes the Bank Group's strategy, based on an assessment of priorities in the country. It indicates the level and composition of assistance to be provided based on the strategy and the country's portfolio performance. The CAS is the central vehicle for management and the Board to review and judge the impact of the Bank Group's assistance strategy for country programs and borrowers.
- CAE Country Assistance Evaluation**  
Produced by OED, the CAE assesses and evaluates the Bank's assistance strategy as stated in the CAS, its lending and nonlending activities in client countries. CAEs assess the outcome and sustainability of the Bank strategy as well as of Bank and borrower performance. It is prepared in parallel with the CAS and provides the Board with an independent assessment and informs the new CAS on the outcomes and performance of the prior CAS. Usually, the CAE covers a decade of Bank assistance—the last two or three CAS cycles.

**CASPR Country Assistance Strategy Progress Report**

The CAS Progress Report (CASPR) reports to Executive Directors on implementation of the CAS program and on key developments since the CAS. It is usually prepared mid-term during the CAS cycle or when the preparation of the next CAS is delayed for any reason. It is a process for reviewing the Bank's assistance program at the country level.

**CASCR Country Assistance Strategy Completion Report**

Introduced at the end of each CAS period, the CAS Completion Report is a self-assessment by the country team. It aims to assess outcomes achieved relative to expectations and to draw lessons for the next CAS. It comments on the relevance of programs to the country's development agenda, quality of the portfolio, and measures of development effectiveness, including insights from clients. The CASCR, as it is currently drafted, will serve primarily as a learning tool, as it does not require the country team to self-rate performance against expected CAS outcomes.

**CES Country Evaluation Summary**

This is an OED tool to review and to validate the self-assessment of country outcomes as stated in the CASCR. This will include ratings on country outcomes and Bank performance during the CAS period.

**CPPR Country Portfolio Performance Review**

A process in which the Bank and the Borrower review the implementation process and performance of the Bank's existing loan portfolio, assess the quality of projects entering the portfolio, ensure the continued relevance of projects in the portfolio for sector strategies, and, if necessary, institute a program of measures to strengthen portfolio performance and enhance the devel-

opment impact of projects. The purpose of the CPR is to reinforce Borrower ownership of Bank-financed projects and to develop the design of the CAS. Country teams usually schedule a CPPR every 12 to 18 months for active Borrowers with more than 20 projects under implementation or more than US\$1 billion in outstanding loans.

**CPR Country Portfolio Review**

The CPR is a revamped CPPR process that is more dynamic and comprehensive and includes a review of the Bank's lending as well as its nonlending portfolio. It aims to reinforce Borrower ownership of Bank-financed projects and AAA to ensure continued relevance in the portfolio and to the country and sector strategy. The Bank would encourage Borrowers to take responsibility for the CPR process, together with key implementing agencies and ministries, people affected by Bank projects, and other interested stakeholders and donor agencies.

**ES Evaluation Summary**

Prepared by OED, this internal document is an independent verification of the findings and self-assessment provided in the project Implementation Completion Reports (ICRs). OED independently reviews and validates all the ICRs.

**GCR Grant Completion Report**

The GCR is the completion report prepared when a grant is closed. The GCR is completed within six months after the closing date of the final Grant Letter of Agreement. It describes the extent to which grant objectives were met and the activities were completed; assesses the program's development impact; identifies problems in the program's execution; and provides lessons learned. It usually includes audited financial statements and any external or recipient-prepared evaluation reports.

<b>IBTF</b>	<p><b>Initiating Briefs for a Trust Fund</b></p> <p>A document used to propose a new or substantially revised freestanding trust fund, in order to finance specific activities related to a unit’s work program. This could include country-specific assistance in the unit, policy research, or training sponsored by a central unit or network.</p>		
<b>ICM</b>	<p><b>Implementation Completion Memorandum</b></p> <p>An ICM is the required completion report for all activities funded through trust funds (or grants that are sub-allocations of trust fund programs) over \$1 million. They are completed by all task managers managing trust fund activities within the six months following the closing date and should be submitted to the central Trust Funds Operations unit (TFO).</p>		
<b>ICR</b>	<p><b>Implementation Completion Report</b></p> <p>The ICR is a self-evaluation document that is prepared for every Bank-financed lending operation at the time of project completion or at the end of the disbursement period. The Bank, with input and comments from the Borrower and other co-financiers, if any, prepares the ICR. The ICR provides an assessment of the development impact of the operation and provides feedback to improve future project design and implementation.</p>		<p>progress, and where prescriptive approaches might not work well.</p>
<b>LIL</b>	<p><b>Learning and Innovation Loan</b></p> <p>LILs are loans of \$5 million or less, normally implemented over two to three years, financing small, experimental, risky and/or time-sensitive projects in order to pilot promising initiatives and build consensus around them, or experiment with an approach in order to develop locally based models prior to a larger-scale intervention. LILs are predominantly used in sectors or situations in which behavioral change and stakeholder attitudes are critical to</p>		
<b>MTR</b>			<p><b>Mid-Term Review</b></p> <p>Introduced by the Bank during the course of project implementation (usually at mid-term) in order to encourage efficiency in project supervision to tackle problems proactively. The task teams, with the country management inputs conduct these reviews. Their quality is periodically assessed by QAG in the quality of supervision annual report.</p>
<b>PAD</b>			<p><b>Project Appraisal Document</b></p> <p>PADs are prepared for lending projects by the Bank, based on the Project Concept Document (PCD). Preparation should include examination of technical designs, institutional arrangements, market outlets, economic and financial viability, environmental and social issues, and any other topics needed to ensure a high-quality project at entry, including due consideration to project management issues.</p>
<b>PCD</b>			<p><b>Project Concept Document</b></p> <p>Prepared for all lending projects based on the Project Implementation Plan (PIP), the PCD defines the rationale for a proposed investment operation, the framework for its preparation, and flags issues or areas of special concern to the Bank. It serves as the basis for a Bank decision to assist a Borrower with project preparation.</p>
<b>PIP</b>			<p><b>Project Implementation Plan</b></p> <p>Prepared for lending projects by the Borrower, with the support and assistance of the Bank Task Team, the PIP lays out the framework for project implementation. It serves as the basis for the preparation of the Project Concept Document (PCD) and the Project Appraisal Document (PAD). It is updated throughout the project’s life and serves as the tool for monitoring and reviewing project implementation.</p>



- PSR Project Supervision Report**  
Provides a concise summary of key project information, assessment of key project assumptions, performance, progress, and risks, including the achievement of stated project implementation and development objectives, and issues impacting implementation and remedial actions. It is an internal management and reporting tool and is available as an online form in the SAP system
- PPAR Project Performance Assessment Report**  
Prepared by OED, the PPAR is a detailed project assessment or evaluation prepared—on average—three years after project completion to validate ICR findings, monitor ICR quality, and provide feedback on lessons learned.
- QEA Quality at Entry**  
The Quality at Entry Assessment, prepared by QAG, is an annual review of a random sample of new lending operations in each calendar year soon after their approval. Its purpose is to provide real-time feedback to staff and managers to identify systemic quality issues and to improve the design of future development operations. The assessment of quality is undertaken along eight dimensions: project concept, objectives and approach; technical and economic analyses; environmental aspects; social and stakeholder aspects; financial management aspects; institutional aspects; readiness for implementation; and analysis of risks and sustainability.
- QSA Quality of Supervision Assessment**  
Prepared by the Quality Assurance Group in Bank management every fiscal year, the QSA assesses the Bank’s performance and progress in using supervision as an input to offer real-time feedback to management and staff. It suggests ways to improve the effectiveness and responsiveness of the Bank’s supervision work.
- QER Quality Enhancement Review**  
The QER is a Bank tool that provides on-demand technical assistance to staff and managers working on a project or an analytical work to improve the quality of the design or implementation. A QER is a tripartite partnership between the requesting Region, the relevant sector anchor unit, and QAG. Expert panelists examine project documentation and interview task team leaders and other task team members, peer reviewers, country and sector directors, and other relevant operational staff. QERs can be undertaken for operations under preparation as well as for those under supervision.
- SSP Sector Strategy Paper**  
Produced by the Sector Boards, SSPs are the Bank’s main vehicle to translate its strategic vision and corporate goals in a sector or thematic area into broad goals and operational choices for the sector. They link sector/theme goals to the Bank’s corporate goals, policy framework, partnerships, and comparative advantage. SSPs focus on sectorwide performance, the importance of country institutional and regulatory frameworks, and constraints to efficient implementation of programs. SSPs are also expected to propose an M&E framework to assess their implementation and impact, including indicators, baselines, targets, and timeframe.

Management welcomes the opportunity to discuss the *2003 Annual Report on Operations Evaluation* (AROE) by the Operations Evaluation Department (OED). As requested by the Executive Directors, this year's AROE concentrates on specific areas where OED sees the need for further adaptation and progress on operations evaluation. In retrospect, Management erred when it did not point out the drawbacks of a report that concentrates almost entirely on areas for improvement. The Bank's first DGO argued strongly on the need for perspective in evaluation. In the context of projects, he noted that if the only reports that Executive Directors received were about problems, that would color their image of the Bank's entire portfolio, however well other projects were doing (OED 2003g). The same analogy holds for the overall results framework. Last year's AROE provided some of that overall perspective. It presented a 10-year retrospective on monitoring, evaluation, and risk systems, highlighting significant progress made by the Bank in recent years.<sup>1</sup> Management broadly agrees with the recommendations of this year's AROE to strengthen monitoring and evaluation policy and systems in several specific areas, but it would like to provide some additional context and note that the needed improvements build on a strong foundation.

**Context and History.** The Bank's monitoring and evaluation systems have strengths that compare very favorably with other development agencies. A 1997 stakeholder feedback process undertaken by OED came to the conclusion that OED was the "best evaluation group in the world" (OED 2003g). Manage-

ment's Quality Assurance Group (QAG), set up in 1997, is being taken as a possible model by other international agencies. In particular, the Bank has traditionally had—and continues to have—a comprehensive monitoring and evaluation system for deliveries, quality and results at the project level. Peer assessment through QAG and independent evaluation by OED have produced reliable and objective time-series data on quality at entry, quality of supervision, and development outcomes for Bank operations. This system has been an important source for learning and accountability within the Bank, and the findings show that Management's attention to quality continues to pay off. The Bank has maintained targeted gains in the quality of the lending program and economic and sector work (ESW), and the system provides quick feedback for action when slippages occur. The relative strength of the Bank's monitoring and evaluation systems is reflected in its corporate reporting. Assessment of portfolio performance (through the *Annual Review of Portfolio Performance*), development effectiveness (through the *Annual Review of Development Effectiveness*), and the Bank's monitoring and evaluation systems (through the AROE) provides the Board with a wide range of information and analysis at the aggregate level.

**Recent Progress.** In recent years, we have diversified our products and services to better meet client demand, and have increasingly moved with our partners toward a focus on the country as the unit of account in development. In recognition of this movement—and to sustain and build on historic achievements in monitoring and evaluation—Man-

agement adopted late last year, after discussion with Executive Directors, an implementation action plan on better managing for results. This action plan is part of a broader effort to scale up the impact of development assistance, and is closely related to the ongoing program to modernize and simplify the Bank's operational products and to harmonize operational policies, procedures, and practices with those of other donors, to the extent possible around borrowing country systems. The action plan focuses on three areas: providing greater support to countries to strengthen their capacity to manage for results; enhancing the focus on results in Bank strategies, instruments, incentives, and reporting systems; and fostering a global partnership to harmonize results-based approaches and coordinate support to countries in managing for results. During the first year of implementation, progress has been made in all three of these areas.

**Strengthening Country Capacity to Manage for Results.** The Bank has provided guidance to countries that want to strengthen the focus on results in their Poverty Reduction Strategies (PRSs), with emphasis on realistic target-setting, sustainable monitoring and evaluation systems, and increased participation in the monitoring process. We have also supported middle-income countries interested in introducing results-based approaches to public sector management and expenditure systems. This work draws heavily on the findings of Public Expenditure Reviews, Country Procurement Assessment Reports, and Country Financial Accountability Assessments. Through the Monitoring and Evaluation Improvement Program, the Bank has provided demand-driven support to countries to strengthen results-oriented monitoring and evaluation (M&E) systems. The Bank is also stepping up support for country efforts to improve the statistical systems they need for monitoring development results, most directly by developing a new lending program that supports a multidonor, sectorwide approach to statistical capacity building.

**Enhancing the Focus on Results in Bank Strategies, Instruments, and Reporting.** Progress has been made in each of these areas.

- **Strategies.** As a central element of this work, and reflecting our increased emphasis on country outcomes, Management is piloting a results-based Country Assistance Strategy (CAS) that is built on a subset of country-owned objectives and a results framework that better links these objectives to Bank programming. It also calls for preparation of a CAS Completion Report by country teams to draw more fully the lessons of experience and to fill a recognized gap in our evaluation of country programs. The Completion Report is reviewed by OED, and this review is shared with Executive Directors. Thus far during the pilot phase, three CASs that incorporate the results approach have been prepared (Sri Lanka, Cameroon, Mozambique) and many more are under preparation in FY04. Country teams have begun mainstreaming the approach.
- **Instruments.** At the same time, Management is strengthening the outcome orientation of lending operations, particularly through clearer articulation of objectives and streamlined monitoring systems. Work is under way to integrate the monitoring and evaluation architecture from the CAS down through the various lending products and analytic services offered by the Bank, and to enhance linkages to each country's own results framework. In terms of investment lending, the Project Concept Note and new Project Appraisal Document (PAD) formats, in use as of July 1, are designed to facilitate the selection of outcome-oriented objectives, definition of relevant output and outcome indicators, establishment of baseline data, and identification of responsibilities for data collection and updating. The new Project Status Report, now under preparation and scheduled for adoption at the beginning of the next fiscal year, will be equally outcome-oriented, following the results framework established in the PAD. Proposed improvements in M&E frameworks for adjustment operations were included in the draft operational policy that was recently submitted to Executive Directors. An initial discussion took place at the Board's Committee on Development Effectiveness.

- **Reporting.** Progress has been made in gaining broad acceptance among donors and borrowers for an enhanced IDA results measurement system, and in piloting the system in advance of IDA14 discussions. The Bank has also experimented with introducing results reporting into the strategy and budget processes through preparation of revised unit compacts. One of the key changes, highlighted during this year's discussion of the *Annual Review of Portfolio Performance* (ARPP), is Management's plan to convert the ARPP to an *Operational Performance and Results Review*. However, because this report will need to build on the accumulation of results-based CASs and projects and programs with stronger results orientation, it will not achieve its full potential until we have run a full CAS cycle and have a significant number of completion reports drawing on results-based CASs. The ARPP estimated that the transition to the new report would take four to five years.

**Global Partnership on Managing for Results.** The June 2002 Roundtable on Results was an initial opportunity for representatives of the multilateral development banks (MDBs), the Development Assistance Committee (DAC) of the Organisation for Economic Development and Co-operation, and the United Nations to exchange information on results-based approaches to development, and to establish a community of practice dedicated to harmonizing and coordinating support to countries in their quest for development results. Later in the year, international consultations on the IDA results measurement system and continued work on MDG monitoring brought critical measurement and reporting issues to the fore. These issues were the focus of a conference organized by the Bank on Strengthening Statistics for Measuring Development Outcomes, which helped move toward more systematic identification of core data gaps, clarification of reporting accountabilities, and coordination of international support for statistical capacity building. This work will serve as an important input to the Second International Roundtable on Managing for Development Results, to be held in Marrakech on February 4-5, 2004. The Bank is working with the other MDBs and with the newly established DAC Joint Venture on

managing for development results to reach consensus on core principles and an initial global action plan for 2004 to be endorsed in Marrakech.

**The Medium-Term Vision.** As with the earlier work to enhance quality, it will take three to five years to fully implement the main elements of the results agenda within the Bank. Furthermore, the quality and reliability of our systems depends critically on the capacity of countries themselves to measure and monitor progress towards key development outcomes—and this will remain a long-term challenge. Within the Bank, becoming a more results-oriented agency will require consistent management attention and appropriate incentives. This is the same formula that worked in achieving the quality targets. Underpinning this change in behavior are the systems and processes for articulating expected results, tracking progress during implementation and assessing achievement of results upon completion. This information is captured in results-based monitoring and evaluation systems that begin with ex ante strategic planning for results and end with ex-post evaluation of outcomes for learning, feedback and accountability. That system is under development.

**The Results Reporting System—the New Architecture.** The results agenda builds towards a Bankwide reporting system on our contribution to development results, drawing on three information sources: country programs, global programs and cross-cutting sector programs. Ongoing work to streamline and strengthen results-based monitoring and evaluation of lending products and analytic services, as well as development of the results-based CAS will serve as the foundation for results reporting within country programs. At the same time, our support for global programs has expanded rapidly in recent years, but our ability to assess their quality, monitor progress and evaluate results has not kept pace. Development of a common format for monitoring and evaluation of global programs is a new challenge, and one that we view as an essential complement to the oversight of country programs. The new Global Programs and Partnership Council and Secretariat have taken on this challenge. New global programs and partnerships

will have M&E frameworks that build on those used already by the Development Grant Facility. At the same time, the Secretariat will work to systematically improve the M&E framework for these activities. Finally, the ability to track and report on results across sectors and themes is not only a management tool for implementation of sector strategies, but is also the most compelling and widely understood way to report on our progress and performance to the outside world. It is consistent with reporting on MDGs, and is a further development of the work undertaken for the IDA results measurement system and for seven pilot areas last year, where results frameworks were prepared on the basis of existing sector strategies and core outcome indicators were identified for sectorwide monitoring. This work will be expanded in the future as the foundation for results reporting by sectors and themes, a key input for the annual *Sector Strategy Implementation Update*. The upcoming progress report on the results agenda will lay out in more detail the results reporting architecture.

**Progress Report on the Results Agenda.** Early next year, Management will report to the Executive Directors on first-year progress in implementing the results agenda. This report will incorporate the experience of the pilot Monitoring and Evaluation Improvement Program adopted in 2000, which has been integrated into the broader results agenda. Design and implementation of the results agenda has been informed by OED findings, and the 2003 AROE analysis and recommendations confirm the importance of continued implementation of the results agenda, while providing useful additional inputs to the process.

## II. AROE RECOMMENDATIONS

The AROE sets out four specific recommendations for Management consideration: (a) issuing an Operational Policy/Bank Procedure (OP/BP) that sets out the mandate, framework, roles, and responsibilities for self-evaluation and independent evaluation; (b) mainstreaming greater outcome orientation in CASs

and a self-evaluation framework for Bank assistance at the country level; (c) strengthening the monitoring and evaluation of the development outcomes of trust-funded activities; and (d) strengthening the outcome orientation and self-evaluation of nonlending activities, at the aggregate and the task level, either individually or in clusters.

**Management Views.** Management generally agrees with all of these recommendations. The question of appropriate timing and sequencing remains. Management's first priority is related to OED's second recommendation: improving the results orientation of CASs and developing an effective CAS Completion Report. This is the linchpin for improving the relevance and effectiveness of our support within the country business model. It is the logical vehicle for organizing Bank support to help borrowing countries in managing for results, and it will be the backbone for corporate monitoring and reporting on results. The results chain from outcome goals to individual inputs will strengthen the outcome orientation of Bank outputs, notably nonlending services. The CAS Completion Report will also serve to synthesize self-evaluation of nonlending services, taking a country perspective on an entire program and drawing on the results chain set out for each ESW product. In this context, QAG's recent decision to supplement its evaluation of individual ESW products with a country-based approach to assessing the quality of ESW will give the Bank another self-evaluation tool for nonlending services. By looking at an entire time slice of ESW, QAG will be better able to assess the contribution of ESW tasks, individually and as a whole, to results. The good practice guidelines for monitorable action plans to be included in diagnostic ESW that Sector Boards will have in place by the end of the fiscal year will also strengthen the outcome orientation of nonlending services, making later self- and independent evaluation easier. Management agrees with OED that it will be useful to develop an operational policy on the roles of, and responsibilities for, self-evaluation and independent evaluation, and it is working with OED to produce such a policy, which will be informed by the

findings of CODE's upcoming review of the mandate and functions of OED.

**Trust-Funded Activities.** Work is under way to address OED's recommendation relating to the monitoring and evaluation of the development outcomes of trust-funded activities. The highest priority, in Management's view, is for trust-funded global programs and partnerships. As mentioned above, the new Global Programs and Partnerships Council and Secretariat are addressing this issue—through requirements in place for new programs and through the continued improvement in the design of M&E tools for these activities. Many other trust-funded activities fall under existing M&E requirements—for example, trust funds that support project-related activities—for which the quality of M&E is being addressed through work already mentioned. Still other trust funds support single activities, often small in scale, for which the M&E framework needs to be scaled to the size of the activity. A substantial share of these activities are now administered through a new system that requires an evaluation framework at the outset and self-evaluation at completion. Ongoing work related to the automation of trust fund administration, planned for completion within two years, is taking on many of the remaining M&E issues with regard to trust funds. The attached matrix provides a more detailed response to each of OED's recommendations, including this recommendation with regard to trust funds.

### III. CONCLUSION

The Bank continues to make progress in implementing its multiyear strategic framework, anchored in the MDGs and the consensus on scaling up the impact of development assistance (World Bank 2003). For the Bank, scaling up requires an unwavering focus on results, building on lessons from earlier successes with quality. In its first year of implementation, the results agenda has made progress on many fronts.

**Challenges Ahead.** While much has been accomplished in laying out the basic design of the results framework and piloting new approaches, clearly much more remains to be done. The 2003 AROE provides important insights and timely inputs for the work ahead. Last year, Management committed to work on the difficult task of moving up the self-assessment ladder toward stronger measures of Bank performance or as recommended by OED, "verifiable performance indicators for progress against poverty" (OED 2002i). Management emphasized that this task would not be easy and would take time, especially since no objective outcome or impact measures of Bank performance are available and—given the many factors that affect development outcomes in borrowing countries—subjective judgment is difficult. Progress on the IDA results framework shows that it may be possible to reach consensus on a results system that measures development outcomes at the aggregate level, while measuring the Bank's performance on subjective but realistic and widely accepted criteria.

## MATRIX: MANAGEMENT RESPONSE TO 2003 ANNUAL REPORT ON OPERATIONS EVALUATION

Recommendation	Management Response
1. Issue an OP/BP that sets out the mandate, framework, roles, and responsibilities for self- and independent evaluation.	<ul style="list-style-type: none"> <li>During the development of the results agenda, Management decided, in consultation with OED, to write an umbrella OP for monitoring and evaluation in the Bank, covering both self- and independent evaluation. This work will involve retiring outdated statements, notably OD10.70, <i>Project Monitoring and Evaluation</i>. As requested by Executive Directors, Management will submit the draft OP to Executive Directors following CODE's consideration of the review of the independent evaluation function currently taking place.</li> </ul>
2. Mainstream greater outcome orientation in CASSs and a self-evaluation framework for Bank assistance at the country level.	<ul style="list-style-type: none"> <li>Management accepts this recommendation. It is the foundation of the results-based CAS that is now under development. Management has undertaken a pilot exercise to produce a minimum of five CAS Completion Reports and results-based CASSs. OED is reviewing each CAS Completion Report and reporting to Executive Directors. At the end of the pilot phase, Management will review progress and report to Executive Directors on the lessons and next steps. That progress report is planned for the end of FY04.</li> </ul>
3. Strengthen the monitoring and evaluation of the development outcomes of trust-funded activities.	<ul style="list-style-type: none"> <li>Management accepts this recommendation in principle. For Management, the highest priority is trust-funded global programs and partnerships. The newly created Global Programs and Partnerships Council and Secretariat requires M&amp;E for new activities, building on the DGF approach, requiring annual progress reporting, financial reporting, and periodic independent evaluation; the Secretariat will continue to work to refine M&amp;E approaches for these activities. Other trust funds contribute to a large number of activities, many of which already fall under existing M&amp;E frameworks. Some trust funds, including consultant trust funds, are used for Bank operational activities, notably project-related activities, for which monitoring and evaluation are required. As the OED report notes, any activity that draws more than \$1 million from trust funds is required to produce a self-evaluation. Some, like the PHRD and Bank-Netherlands Partnership Program (BNPP), have produced detailed evaluations. Management has introduced the SAP-based Trust Fund Status Report (TFSTAR), including a self-evaluation section, and it is required for major programmatic trust funded activities, such as individual activities funded by the PHRD program, the BNPP, the Trust Fund for Statistical Capacity Building, and the Japan Social Development Fund. Other programs, including the Institutional Development Fund, also use TFSTAR, and Management is considering extending its use to other trust funds. TFSTAR now covers a substantial share of trust funds under \$1 million. Ongoing work related to the automation of trust fund administration, planned for completion within two years, is taking on many of the remaining M&amp;E issues with regard to trust funds. Many trust-funded activities are partnerships and fall under the partnership evaluation rules. Again, as OED notes, Development Grant Facility (DGF) activities, many of which are cofinanced by trust funds, are required to produce evaluations. The Trust Fund Operations unit will take the lead in reporting to Executive Directors on progress in systematically and cost-effectively strengthening the results framework for trust-funded activities.</li> </ul>
4. Strengthen the outcome orientation and self-evaluation of nonlending activities, in aggregate level, and at the task level, either individually or in clusters.	<ul style="list-style-type: none"> <li>The outcome orientation and self-evaluation of non-lending activities—basically ESW—will be strengthened in the near future in several ways. Two measures are planned to strengthen the outcome orientation of ESW. First, this is one of the important aims of results-based CASSs. Through their results chain construction, leading from each outcome goal to supporting activities, they will clearly establish the outcome orientation for country ESW. Second, by the end of the fiscal year, Sector Boards will have in place good practice guidelines for monitorable action plans to be included in diagnostic ESW, improving the evaluability of diagnostic ESW. These good practice guidelines spell out the recommended coverage, content, and methodology that best address problems most often encountered in the relevant sectoral or thematic area. It is then the role of the Sector Director and, in some cases, of the Regional Chief Economist to ensure that good practice guidelines are appropriately used, notably to achieve high-quality concept papers and decision drafts, with accountability reinforced by evaluation. Two other measures will strengthen self-evaluation of ESW. First, QAG reviews of ESW at the task level will be supplemented by the mainstreaming of the recent pilot country-level evaluations of ESW programs. By evaluating multi-year programs of country ESW, QAG will be better able to assess the contribution of ESW—individual studies, sets of studies, and the overall program—to results. Second, CAS Completion Reports will look specifically at the contribution of ESW activities to the achievement of the results set out for the country program, drawing on the results chain for each ESW activity established in the CAS. In addition Management is reviewing the format of the ESW Activity Completion Summary, with the goal of making this tool less cumbersome for team leaders to use, while at the same time improving its results focus.</li> </ul>

**Key Findings and Recommendations of the AROE.**

In response to the Board's request that the AROE provide more information about areas where progress was most needed, the 2003 AROE recognizes the increased attention in the Bank to focusing on results and strengthening monitoring and evaluation since last year. It highlights three broad areas requiring greater improvement and management attention:

- The need to enhance outcome orientation and evaluability of activities in the design stage
- The need to strengthen the quality and realism of reporting on implementation progress and outcomes, based on measurable indicators set out during the design stage
- The need to close gaps in self- and independent evaluation.

The AROE makes four recommendations for management's consideration:

- Issue an Operational Policy/Business Procedures (OP/BP) that sets out the mandate, framework, roles, and responsibilities for self- and independent evaluation.
- Mainstream greater outcome orientation in CASs and self-evaluation framework for Bank assistance at the country level.
- Strengthen the monitoring and evaluation of development outcomes of trust-funded activities.
- Strengthen the outcome orientation and self-evaluation of nonlending activities, in aggregate level, and at the task level, either individually, or in clusters.

Management generally agrees with the recommendations but notes that questions of appropriate tim-

ing, sequencing, and tradeoffs remain. Highest priority has been given to filling gaps in the monitoring and evaluation framework for country and sector strategies, and strengthening implementation of project level monitoring and evaluation systems. Efforts to strengthen monitoring and evaluation for trust-funded activities will require more time to become fully operational.

The AROE includes three recommendations to improve OED's work:

- Enhance borrower consultations on sector and thematic studies and PPARs.
- Improve internal dissemination of sector and thematic studies through better targeting and more active participation in sector and Regional events.
- Institute a transparent procedure to resolve differences in ratings between OED and management for country- and project-level evaluations.

**Main Outcomes of the Committee's Discussion.**

The committee welcomed the AROE's focus on the Bank's monitoring and evaluation system and generally supported the recommendations of the AROE. Members felt the report was of high quality and identified appropriate priorities for management for the coming years. Members welcomed the broad agreement of management with the findings of the AROE. It also welcomed the fact that OED was subjected to self-evaluation.

The key outcomes of the committee's deliberations included the following. CODE:

- Stressed the importance of ensuring the evaluability of operations at the design stage
- Agreed that the Bank's M&E framework should cover all Bank operations, including sector strate-



- gies, adjustment lending, and nonlending activities
- Agreed that the new OP/BP for M&E should reflect the results of CODE'S review of the evaluation function and DGO mandate
- Supported management's medium-term plan to convert the *Annual Review of Portfolio Performance* (ARPP) into an *Operational Performance and Results Review* (OPR) to better reflect implementation of the agenda on managing for results
- Expressed concern about the adequacy of management's response for strengthening the evaluation for ESW
- Asked management to continue to update the Board on progress in implementing the agenda and action plan on Managing for Results
- Expressed a preference for OED to continue preparing an annual report on operations evaluation.

**Highlights of Views Expressed.** Speakers commented on the *need to enhance outcome orientation and evaluability of activities* in the design stage of Bank operations, and on *weaknesses in the Bank's monitoring and evaluation framework*. They also expressed views about *building evaluation capacity in client countries and in the Bank*; the *need for a comprehensive/OP/BP on evaluation*; the *plan to convert the ARPP to an Operations Performance and Results Review (OPR)*; and a *biennial versus an annual report on operations evaluation*.

**Outcome Orientation and Evaluability.** The committee agreed with the AROE's assertion that it is critical to establish clear and realistic expected outcomes, and to develop monitorable indicators, baseline data, and benchmarks of progress at the design stage, especially for adjustment lending, nonlending services, sector strategies, and knowledge initiatives. Members believe that improving the evaluability of Bank operations is of overriding importance and consider it an important element of quality at entry that affects the usefulness of subsequent evaluation exercises.

In response to concern about the OED finding that task teams appear to be reducing their use of advice from Quality Enhancement Reviews and peer reviews, management said it was following up on a number of QAG recommendations on this important issue. In particular, it is looking at the role of sector managers in the Quality Enhancement and

peer review processes.

**Weaknesses in the Bank's Monitoring and Evaluation Framework.** The committee expressed concern about the areas of weakness in the Bank's monitoring and evaluation framework identified in the AROE, namely *sector strategies, adjustment lending, nonlending services (ESW), trust funds, and grants*. They agreed that coverage of M&E should be expanded to cover these areas. Members took note of management's concern that the effort to expand and improve M&E coverage could require additional resources but agreed with OED that this is more a matter of greater efficiency in the use of existing resources rather than the need for additional resources. They stressed the need to prioritize. The view was expressed that monitoring should be an inherent part of project implementation but care should be taken so that the opportunity costs are not excessive.

**Sector strategies.** CODE members said that improving the framework for reporting on the implementation of sector strategies must be a priority for OED and management. The view was expressed that the problem might be one of accountability within the matrix system. CODE members welcomed management's plans to present an update on the implementation of sector strategies. They also stressed the importance of having meaningful outcomes and relevant monitorable indicators.

In response to the accountability issue, management said that the annual reporting on sector outcomes would be on the basis of the operational programs implemented by the Regions as well as global programs implemented by network anchors. Discussions will go beyond the narrow context of the network anchor work programs to incorporate Bank-wide activity. Management will look at aggregates and together with the Board, make judgments about performance and whether resources need to be allocated differently.

**Adjustment lending.** Members noted OED's finding that there is no explicit requirement to set outcomes for adjustment lending. They stressed the importance of correcting this weakness, given the trend towards more adjustment lending in the Bank. Members noted that it is critically important that the

upcoming OP/BP on adjustment lending set some very clear and definable standards for M&E and looked forward to a more detailed discussion on the issue when the draft OP/BP is discussed.

**Nonlending activities.** Several speakers questioned management's proposal to use the new CAS completion reports (CASCRs) to monitor the implementation of nonlending activities in light of the expected brevity of these reports. They expressed doubt that the CASCRs would provide the scope for the in-depth assessment of nonlending activities that would be required. Thus they expressed concern on the adequacy of the management response to this recommendation and suggested that CODE might examine this issue in the course of its upcoming review of the evaluation mandate.

**Evaluation Capacity Development.** The committee noted with concern OED's finding that the Bank has made only limited progress in mainstreaming evaluation capacity building into core operations. Members stressed the importance of understanding better the reasons for the limited impact of donor efforts to build country evaluation capacity to date. They noted that this issue is especially important given the focus on PRSPs and the role of the Bank in helping countries adopt results-based strategies. They urged management to set clear goals on the nature and extent of evaluation capacity that may be feasibly developed in individual or groups of countries and to use this information as a basis for developing collaborative and participatory programs that meet country needs and demands.

Management said that the Bank has strengthened assistance to countries that want to make their PRSPs more results-focused with emphasis on realistic target-setting, sustainable monitoring and evaluation systems, and increased participation in the monitoring process. The Bank has also supported non-PRSP countries to introduce results-based approaches to

public sector management and expenditure systems. The Bank has provided demand-driven support to countries to strengthen results-oriented M&E systems through the M&E Improvement Program.

**Need for an OP/BP on M&E.** The committee agreed that work on a comprehensive operational policy on monitoring and evaluation of Bank operations should be expedited and agreed that this could be achieved by broadening OD 13.60 on Monitoring and Evaluation. Management assured the committee that there would be adequate consultations to ensure that the new OP reflects the results of the upcoming CODE review of the evaluation function and mandate of the DGO.

**Conversion of the Annual Report on Portfolio Performance (ARPP) to Operational Performance and Results Review (OPR).** The committee welcomed the planned conversion of the ARPP to a corporate report on the outcomes of Bank strategies and operations (OPR) to better reflect corporate implementation of the results agenda. They asked management to provide more details on the scope, content, and timing of the OPR, perhaps in a technical briefing, and encouraged management to make the OPR a comprehensive report on the results of all Bank activities. They also suggested that client countries should be consulted once the outline of the OPR has been agreed.

**Annual vs. Biennial AROE.** Several members did not support OED's proposal to make the AROE a biennial report. They noted that the AROE is a key tool for monitoring the implementation of the Bank's corporate strategy and should remain an annual report until the Bank has completed its transition to a results-based monitoring and evaluation framework. A member suggested that OED prepare a comprehensive report on a biennial basis and a more selective report in the alternate years on progress in M&E on one or more strategic topics.

---

Chander Mohan Vasudev,  
*Acting Chairman*



## Endnotes

### Chapter 1

1. This AROE does not address risk management aspects—fiduciary, safeguard, reputational, and the like—which are tracked by the Risk Management Secretariat.

### Chapter 2

1. Monitoring at the corporate level is discussed in Chapter 4.
2. Management does not agree that there is evidence to single out adjustment loans with regard to monitoring against outcomes or with regard to supervision. Because of the existence of an agreed policy matrix, monitoring of adjustment loans against agreed actions is strong. Overall, the measured quality of adjustment lending remains high. While not normally measurable during the life of an adjustment loan, OED research from the mid-1990s found that poverty was reduced and social conditions improved in countries with well-implemented adjustment programs (see OED 1995).
3. An internal management survey of task team leaders for investment loans conducted in September 2002.
4. Bank Procedure 10.00 requires that in the project identification stages, the Borrower prepare a PIP, which will provide a clear statement of the project (i) objectives, including draft financing plan; (ii) organization and management; (iii) an initial, time-bound implementation schedule for each project component; (iv) preliminary procurement and Management Information System plans; and (v) technical assistance and training needs for improving capacity. To be effective, the initial PIP is to be continuously updated as the project takes shape—during preparation, appraisal, and implementation.
5. Management notes that investment projects are just one way—and often not the best way—to support capacity building for the enhancement of country M&E. It can go against the objective of keeping projects simple and streamlined, runs the danger of creating a proliferation of unsustainable project-specific M&E systems, and can get in the way of harmonization. Management's implementation action plan for better measuring, monitoring and managing for development results lays out management's plans for helping countries build M&E capacity, notably in the context of supporting the improvement of overall public expenditure management.
6. QAG noted that 80 percent of task teams reported that Regional front offices did not provide timely or constructive inputs; 73 percent were not satisfied with guidance of sector managers.
7. The Project Concept Note will focus on strategic issues; situate the project in the CAS context; address development, safeguard, and fiduciary risks upfront; and provide the basis for a managerial decision on whether or not to go ahead with project preparation.
8. Management notes that the new Project Concept Note and PAD templates are in use as of July 1. By the end of the calendar year, the accompanying simplification of the PSR will be in place. The objective of the new PSR is to create a simple and straightforward document to facilitate regular discussions between task teams and their managers on project progress, using the agreed outcome indicators as the centerpiece of that discussion. Management also plans to issue its new policy on adjustment lending by the end of the year. That policy provides the opportunity to clarify policy and good practice on monitoring the implementation of adjustment loans.
9. In FY02, QAG rated the overall quality of ESW at 94 percent satisfactory or better. QAG has also reported significant improvements in the quality of Bank inputs and processes (83 percent satisfactory), reflecting effective management attention to ESW quality (QAG 2003b).
10. The first phase in FY01 focused on improvements in ESW recording and monitoring systems. The second,

in FY02, focused on the strategic role of ESW, including introduction of core diagnostics.

11. Small tasks that are not required to prepare formal concept papers usually prepare an Activity Initiation Sheet (AIS), which is expected to state objectives and outcomes.
12. In its FY02 ESW assessment, QAG reviewed 91 “Other ESW tasks” (amounting to \$6.5 million) parallel to the main sample. These tasks included diagnostics, updates of technical papers, conferences, workshops, country dialogue, and consultations that are integral to the Bank’s country-level program (QAG 2003b).
13. Except to the limited extent that borrower comments are used to determine ratings on *likely impact*.
14. The trust fund coordinator, sector/country management, Trust Funds Operation, the Legal Vice-Presidency, Accounting and Trust Funds Division, and Loans Department clear IBTFs.
15. Other initiatives are systems-based control and reporting of financial charges and Web-based updates of financial statements and financial information to donors.
16. The pilot results may not be representative of Bank staff overall, because participation was voluntary and consisted of “motivated staff that wanted immediate accreditation” (WBI 2003).
17. Issues in financial reporting using SAP data are being addressed through a trust fund Portal Project. The Audit Report Compliance System (ARCS), for instance, includes trust fund audit reports, but the FY02 portfolio review raises four unresolved issues. First, ARCS does not allow reports to be retrieved by trust fund number. Second, ARCS uses the Bank fiscal year, which usually differs from those in client countries. Third, ARCS does not automatically accept changes made in the SAP/Loan Accounting System. Last, SAP has to be adapted to provide periodic monitoring reports to meet TFO and donor needs.
18. The Human Development Network and Private Sector Development and Infrastructure Network systematically review outcomes of closed lending operations for lessons.
2. In 1997, OED undertook a review of ESW that led to the *Fixing ESW* reform process.
3. OED Work Program and Budget (FY04) and Indicative Plan (FY05–FY06), March 2003.
4. Management has moved steadily from a task-by-task basis to a country program basis, linked to the country’s own development objectives, in its approach to assisting developing countries. It is moving its self-evaluation framework in the same direction. The CAS Completion Report will look at the contribution of Bank instruments, including ESW, and may be the best form of self-assessment of ESW.
5. Data source: Trust Fund Operations Unit estimates. The grant amount as indicated refers to the total approved amount, less Bank administrative fees. As of end FY02, the grant amount was \$2.93 billion.
6. In FY04, the TQC is planning to review the basis for vice-presidential unit assessments and quality assurance mechanisms for the same.
7. The purpose of standard performance indicators is to allow independent and verifiable assessments; provide a framework for selecting appropriate trust fund instruments, depending on intended impact; and ensure that external funds are used appropriately.
8. OED Work Program and Budget (FY04) and Indicative Plan (FY05–06), March 18, 2003.
9. The CASCR will be discussed with government and other stakeholders during the CAS preparation process and will be included in the up-stream review package for the new CAS.
10. The final CASCR will be included as an annex to the new CAS.
11. The DGF did not comply with the requirement to report on GCRs in its annual review for FY02 to the Board (OED 2002c, p. 47).
12. OED 2002c, p. 48. Phase II of the global programs evaluation will elaborate on quality, timeliness, and independence of evaluations and recommend ways to focus them on development effectiveness.
13. The discussion of Global Programs and Grants above includes these global knowledge initiatives.
14. Level I focuses on monitoring feedback from trainees. Level 2 focuses on measuring participants’ knowledge/skills at the start and the end of the courses. The third and fourth levels of evaluation relate to outcome and impact measurement. The evaluation framework covers training conducted by WBI and the Learning Board.

### Chapter 3

1. The 1998 OPCS reporting requirements called for ACS within six months of delivery to the client. A more complete document that includes a client survey is required for tasks over \$50,000.

## Chapter 4

1. This review included Bank lending and formal diagnostic work focused on strengthening country systems for monitoring and evaluating government performance. It does not include efforts that are restricted to strengthening national statistical systems (which might focus on the MDGs and other development indicators) or to strengthening financial management information systems.
2. This approach extends the stocktaking undertaken in 2002, discussed in OED 2002b. Nine are in Africa (Ethiopia, Ghana, Madagascar, Malawi, Mauritius, Mozambique, Sierra Leone, Tanzania, and Uganda). Five are in Latin America and the Caribbean (Bolivia, Brazil, Chile, Colombia, and Honduras). Four are in Europe and Central Asia (Albania, Kyrgyz, Poland, and Romania). The others are in East Asia and the Pacific (the Philippines), the Middle East and North Africa (Egypt), and South Asia (Bangladesh).
3. WBI is also supporting the involvement of civil society in making assessments of government performance in several countries through mechanisms such as household surveys. This work focuses on the quality of governance and the extent of corruption.
4. Management would like to note that supporting countries in their development of results-based public expenditure systems and M&E capacity is one of the three main elements of its action plan for enhancing results orientation. Support for essential statistical capacity development, as discussed at the June 4-5 international technical meeting on Improving Statistics for Measuring Development Outcomes co-hosted by the Bank, the work via the results-based CAS to reflect in country support the findings of diagnostic tools with respect to country results-based management systems, and, for PRSP countries, continued monitoring to identify key country needs for follow-up, are all part of that action plan. However, realistically, it will take time and continued attention for developing countries to put in place strong results-focused systems, judging from the experience of industrial countries.
5. Information from the Global HIV/AIDS M&E Support Team, February 2003.
6. All these reports comment on the performance of the Bank as well as other donor institutions, the borrowers, and the development community at large.
7. A third instrument is corporate strategy and budget documents. Management has piloted "unit compacts" that provide a basis for administrative units to define their outcomes and link unit activities, to the extent possible, to outcomes. It aims to provide greater comparability across similar units (for example, networks) and facilitate senior management discussions on optimal allocation of resources. There is currently no attempt to link stated outcomes to budget allocations or aggregate across units. The success of this initiative will depend on the extent to which administrative units use the Compact as a management tool for programming and reporting.
8. The 2002 ARDE comments on the implications of the MDGs, the 2001 ARDE reported on instrument choice, and the 1999 ARDE focused on the Comprehensive Development Framework.
9. Single-tranche loans accounted for 19 percent by number and 15 percent by volume of Adjustment Loan disbursements in FY01, and 16 percent by number and 34 percent by volume in FY00 (World Bank data).
10. Management would like to note that the exclusion from the ARPP of Adjustment Loans approved and completed within a single year is a definitional choice and not a blind spot. The portfolio for the purpose of the ARPP is specifically defined as projects still under execution. Management has other forms of self-evaluation for loans approved and completed within a single year, notably QAG Quality at Entry reviews and adjustment lending retrospectives.
11. QAG is adding results focus as an explicit criterion for assessing operational quality at entry and supervision in the next round of quality assessments in FY04.

## Chapter 5

1. OED Work Program and Budget (FY04) and Indicative Plan (FY05-FY06), March 2003.
2. This is used within OED to plan dissemination activities.
3. The 2002 AROE surveyed four audiences: (i) advisers and assistants to executive directors; (ii) thematic group members on sector/thematic studies; (iii) country teams on CAEs; and (iv) higher-level staff on various OED products and services. OED surveyed more than 750 Bank staff on specific products, and 2,250 Bank staff not covered by targeted product surveys. Given comprehensive coverage of the latter, OED expects to undertake such a survey biennially. Staff of executive directors were not surveyed this year because of large turnover.
4. OED 2003c,d, 2002e,f, 2001a. The PPAR survey was targeted to: task managers whose projects were approved in FY02 and task managers and sector directors at the time of the ICR for the projects assessed by OED in FY02.

## Chapter 6

1. Management notes that it has given a great deal of attention to adjustment lending, including a major retrospective and an issues note discussed with executive directors, consultations with clients and stakeholders—including face-to-face consultations in several continents, and extensive work on a draft Operational Policy, to be discussed with the executive directors in the second half of calendar year 2003.
2. Management notes that the 2 percent of total cost figure cited by OED contains much that is not M&E. One problem with past initiatives to improve M&E was their lack of funding or inadequate funding. Management has made an attempt to carefully estimate the cost of the additional work required to move forward the results agenda (including M&E) in FY04 and to fully fund that work.
3. An additional \$1 million targeted for the Global Monitoring Report has not been considered.

### Annex C

1. Figures in table C.1 represent number of reports received and evaluated by OED in each fiscal year. The annual distribution of these numbers before FY01 ignores late incoming reports that were included in the previous year's cohort. With the automatization of the ICR review and evaluation process, this practice has now been eliminated.
2. Projects may have had indicators or plans for M&E, but ICRs did not reflect it.
3. The *Annual Report on Portfolio Performance* uses the term “net disconnect” to describe differences in rating changes between operations and OED. This term is defined as the difference between the percentage of projects rated as unsatisfactory by OED and the percentage rated by the Regions in the final PSR as unsat-

isfactory in achieving their development objectives. This annex deals with both total and net rating changes along the entire spectrum from PSR to PPAR. The term “net disconnect” applies to only one of the four instances for “rating change.”

4. On completion, projects are also rated on institutional development impact and sustainability.

### Annex D

1. Last year, OED surveyed more than 750 Bank staff on specific OED products and services, with a response rate of about 25 percent. Based on the results, OED launched detailed reviews of two aspects of its work: Country Assistance Evaluations and Web-based services. OED also surveyed 2,250 Bank staff in general (not covered by OED surveys) on the quality and influence of OED products and services—the response rate was 25 percent. Due to survey fatigue, OED did not conduct the general survey this year. The survey of the staff of executive directors was not conducted because of large turnover.
2. In 2000–01, the number of responses to some of the surveys was low. Comparisons of 2003 survey results with those of 2000–01 are valid only for certain questions in some surveys; only these are presented.
3. The response rate in 1999 for S&T evaluations was 46 percent (n = 59).
4. Respondents rated evaluations on overall quality on a 4-point scale: excellent, good, fair, and poor.
5. Decline in methodological soundness and objectivity are significant statistically at 95 percent confidence.
6. Respondents ranked influence on a four-point scale: strongly, somewhat, little, or not at all.
7. Statistical significance was tested at 95 percent confidence.

## Bibliography

- CDF Secretariat. 2003. *Toward Country-led Development: A Multipartner Evaluation of the Comprehensive Development Framework*. Washington, D.C.: World Bank.
- Kusek, Jody, and Ray Rist. 2001. "Building a Performance-Based Monitoring and Evaluation System: The Challenges Facing Developing Countries." OED, Washington, D.C.
- North, V. Haven. 2003. "Evaluation in Bilateral and Multilateral Development Assistance Organizations and Danida's Evaluation Program Set-up." Annex 2 in *A Review of Evaluation in Danida*. Copenhagen: Ministry of Foreign Affairs.
- OED (Operations Evaluation Department). Forthcoming. "Review of the Poverty Reduction Strategy Paper Initiative." Washington, D.C.
- . 2003a. *Sharing Knowledge: Innovations and Remaining Challenges*. OED Study Series. Washington, D.C.: World Bank.
- . 2003b. *Debt Relief for the Poorest: An OED Review of the HIPC Initiative*. OED Study Series. Washington, D.C.: World Bank.
- . 2003c. *Implementation of OD 4.20 on Indigenous Peoples: An Evaluation of Results*. Report No. 25754. Washington, D.C.
- . 2003d. *Improving the Lives of the Poor through Investment in Cities: An Update on the Performance of the World Bank's Urban Portfolio*. OED Study Series. Washington, D.C.: World Bank.
- . 2003e. *Extractive Industries and Sustainable Development: An Evaluation of World Bank Group Experience, Annex C*, "World Bank Experience." Washington, D.C.: World Bank.
- . 2003f. "Independence of OED." OED Reach. Washington, D.C.
- . 2003g. *World Bank Operations Evaluation Department: The First 30 Years*. Washington, D.C.: World Bank.
- . 2002a. *2002 Annual Report on Operations Evaluation*. Report No. 24322. Washington, D.C.
- . 2002b. *2002 Annual Report on Evaluation Capacity Development*. Washington, D.C.
- . 2002c. *The World Bank's Approach to Global Programs: An Independent Evaluation, Phase I Report*. Washington, D.C.
- . 2002d. *IDA's Partnership for Poverty Reduction: An Independent Evaluation of Fiscal Years 1994–2000*. OED Study Series. Washington, D.C.: World Bank.
- . 2002e. *Social Funds: Assessing Effectiveness*. OED Study Series. Washington, D.C.: World Bank.
- . 2002f. *Bridging Troubled Waters: Assessing the Water Resources Strategy since 1993*. OED Study Series. Washington, D.C.: World Bank.



- . 2002g. *2002 Annual Review of Development Effectiveness: Achieving Outcomes—The Millennium Challenge*. OED Study Series. Washington, D.C.: World Bank.
- . 2002h. “Grant Programs: Improving Their Performance.” OED Précis 224. Washington, D.C.
- . 2002i. *2000–2001 Annual Report on Operations Evaluation*. Washington, D.C.
- . 2001a. *Cultural Properties in Policy and Practice: A Review of World Bank Experience*. Washington, D.C.
- . 2000. *Poverty Reduction in the 1990s: An Evaluation of Strategy and Performance*. OED Study Series. Washington, D.C.: World Bank.
- . 1999. *Investing in Health: Development Effectiveness in the Health, Nutrition, and Population Sector*. OED Study Series. Washington, D.C.: World Bank.
- . 1998. *Financial Sector Reform: A Review of World Bank Assistance*. OED Study Series. Washington, D.C.: World Bank.
- . 1995. *The Social Impact of Adjustment Operations: An Overview*. OED Study Series. Washington, D.C.: World Bank.
- QAG (Quality Assurance Group, World Bank). 2003a. “Quality at Entry in FY02 (QEA5): A QAG Assessment.” Washington, D.C.
- . 2003b. “Quality of ESW in FY02: A QAG Assessment.” Washington, D.C.
- . 2002. *Annual Report on Portfolio Performance: Fiscal Year 2002*. Washington, D.C.
- WBI (World Bank Institute). 2003. “Learning Brief: Trust Funds Learning and Accreditation Program Pilot Formative Evaluation.” Washington, D.C.
- . 2002. “Measuring, Monitoring and Managing for Development Results: An Approach to M&E in WBI Country-Focused Activities.” Washington, D.C.
- World Bank. 2003. *Strategy Update Paper for FY04–06: Implementing the World Bank’s Strategic Framework*. Washington, D.C.
- . 2001a. “Better Measuring, Monitoring and Managing for Development Results.” OPCS, Washington, D.C.
- . 2001b. “Adjustment Lending Retrospective: Final Report.” OPCS, Washington, D.C.
- . 1998. *Assessing Development Effectiveness and the International Finance Corporation*. Washington, D.C.