



2002 Annual Report on Operations Evaluation

Laurie Effron

Director-General, Operations Evaluation: Robert Picciotto
Director, Operations Evaluation Department: Gregory K. Ingram
Manager, Corporate Evaluations and Methods: Victoria M. Elliot
Task Manager: Laurie Effron
Task Team: Carlos Reyes, Deepa Chakrapani, Gillian Perkins,
Zamir Islamshah



© 2002 The International Bank for Reconstruction and Development / The World Bank
1818 H Street, NW
Washington, DC 20433

All rights reserved.

OPERATIONS EVALUATION DEPARTMENT

ENHANCING DEVELOPMENT EFFECTIVENESS THROUGH EXCELLENCE AND INDEPENDENCE IN EVALUATION

The Operations Evaluation Department (OED) is an independent unit within the World Bank; it reports directly to the Bank's Board of Executive Directors. OED assesses what works, and what does not; how a borrower plans to run and maintain a project; and the lasting contribution of the Bank to a country's overall development. The goals of evaluation are to learn from experience, to provide an objective basis for assessing the results of the Bank's work, and to provide accountability in the achievement of its objectives. It also improves Bank work by identifying and disseminating the lessons learned from experience and by framing recommendations drawn from evaluation findings.

The findings, interpretations, and conclusions expressed here 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 cannot 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 on the part of the World Bank any judgment of the legal status of any territory or the endorsement or acceptance of such boundaries.

Rights and Permissions

The material in this work is copyrighted. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or inclusion in any information storage and retrieval system, without the prior written permission of the World Bank. The World Bank encourages dissemination of its work and will normally grant permission promptly.

For permission to photocopy or reprint, 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.

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.



v	Foreword
	v Background
	vi Evolving Corporate Strategies
	vi Progress in Evaluation, Control, and Risk Management
	vi Remaining Challenges
	vii Recommendations
ix	Abbreviations and Acronyms
1	1. Introduction
3	2. The Control Environment
7	3. Risk Assessment
	7 Lending: Risk Assessment for Projects in the Pipeline and under Implementation
	9 Grants, Trust Funds, and Partnerships
	11 Risk Assessment at the Country Level
	11 Risk Assessment at the Corporate Level
13	4. Control Activities
	13 Bank Policies and Guidelines
	13 Quality Assurance/Enhancement of Lending
	14 Monitoring and Self-Evaluation at the Project Level
	16 Quality Assurance for Economic and Sector Work (ESW)
	16 Control Self-Assessments and Enterprise Risk Management
17	5. Monitoring
	17 Monitoring Projects and ESW at the Corporate Level
	18 Sector and Thematic Monitoring and Evaluation
	19 Country-Level Evaluation
	20 Monitoring and Evaluation in the World Bank Institute (WBI)
	21 Monitoring and Evaluating the Bank's Research
	21 Corporate-Level Performance Measurement
23	6. Information and Communication
	23 Information
	24 New Sector and Thematic Codes
	24 Communication and Disclosure
25	7. OED's Agenda
	25 OED's Outputs: The View from the Committee on Development Effectiveness (CODE)
	26 Recommendations from Previous AROEs
	26 OED Outcomes and Impact
	29 Recommendations for OED
31	8. Conclusions and Recommendations
	31 Main Findings
	32 Recommendations
35	Annexes
	35 Annex 1: OED Review of ICRs
	41 Annex 2: Regional Quality Teams: A Comparison across Regions

- 45 Annex 3: Quality at Entry, Outcomes, and Portfolio at Risk—
A Comparative Analysis of QAG and OED Findings on Lending Operations
- 49 Annex 4: FY02 OED Summary Work Plan and Results Framework
as of March 31, 2002
- 51 Annex 5: OED Surveys
- 59 Annex 6: CAE Tracer Study
- 63 Annex 7: Chairman’s Statement, Committee on Development
Effectiveness (CODE)

67 Endnotes

71 References

Boxes

- 2 1.1 COSO
- 4 2.1 Milestones in the Control Framework for Risk Management
- 5 2.2 Increasing Focus on Poverty
- 8 3.1 Regional Approaches to Risk Assessment
- 10 3.2 Grants, Trust Funds, Partnerships: What They Are,
What Risks They May Involve
- 11 3.3 Risk Assessment Tools at the Country Level
- 12 3.4 Sample Checklist for Country Risk Analysis
- 14 4.1 Conversion and Updating Process of OP/BP: Progress Reporting
- 15 4.2 Efforts to Strengthen M&E Over the Past Decade
- 19 5.1 Strengthening Evaluation Capacity in the Bank and Borrower Countries
- 20 5.2 Learning Achievement of WBI Course Participants
- 21 5.3 Ex ante and Ex post Evaluation of Bank Research
- 26 7.1 OED’s 1997 Renewal Strategy
- 29 7.2 CAE Recommendations and Implementation of Bank Assistance:
A Varied Picture

Figure

- 27 7.1 Quality of OED Evaluations

Tables

- 8 3.1 Quality of Risk Management Processes
- 28 7.1 OED Influence

FOREWORD

The Annual Report on Operations Evaluation (AROE) fulfills a statutory responsibility of the Operations Evaluation Department (OED)—to attest to the progress, status, and prospects of development accountability and organizational learning within the World Bank. This year’s AROE, grounded in a 10-year retrospective of monitoring, evaluation, and risk management systems in the Bank, documents remarkable progress. Further adaptation will be needed to respond to the new demands of the corporate strategy and the Bank’s endorsement of the Millennium Development Goals.

Background

The shift toward an implementation culture dates from an internal Bank review, commonly referred to as the “Wapenhans report,” commissioned by President Preston in 1992. It brought into sharp relief the disappointing performance results documented by OED, and it succeeded in shifting managerial attention from the approval of individual loans to proactive portfolio management. In 1994, accountability for results on the ground was further enhanced by the creation of an independent Inspection Panel equipped to investigate local communities’ allegations of Bank failures to comply with agreed policies.

Further intensification of the managerial focus on development effectiveness has taken place under the leadership of President Wolfensohn. The Executive Directors provided strong support to the

President’s initiatives and contributed to the culture of results by adding a Committee on Development Effectiveness (CODE) to the Board’s oversight structure. A new risk assessment and control framework (COSO) was adopted. A professional corruption and fraud investigation unit was created. A quality assurance and compliance unit to improve safeguard policy performance was introduced within the Environmentally and Socially Sustainable Development Vice Presidency.

Most important, a Quality Assurance Group (QAG) was set up to track operational quality and development risks in real time. OED launched its renewal, which focused on filling evaluation gaps, enhancing organizational learning, and moving evaluation to the higher plane of country and sector programs.

Evolving Corporate Strategies

The “country focus” of the operational matrix, the Comprehensive Development Framework (CDF), the Poverty Reduction Strategy Paper program (PRSP), and the Heavily Indebted Poor Country (HIPC) initiative have put the country “at the center” of Bank operations. Aid coordination and capacity building services have assumed a higher profile in country assistance strategies. Knowledge and partnership have replaced resource transfer and conditionality as favored instruments of operational assistance. The scope and number of institutional partnerships—many with a regional and global reach—have expanded rapidly.

These changes have received broad-based support from the development community. So have the Bank’s efforts to emphasize country performance in lending resource allocations, to formulate explicit sector strategies in consultation with partners, and to recast policies so as to draw clearer distinctions between mandatory “do no harm” safeguards and “do good” socially and environmentally sustainable practices.

A decision matrix weighted toward the Regions has contributed to tensions between the Bank’s country focus and its implementation of more comprehensive and rigorous operational standards. In weak institutional environments and where borrower commitment has been lacking, excessive transaction costs, high operational attrition rates, and risk aversion have hindered development effectiveness. Further efforts to balance the matrix and improve the linkages between country programs and sector strategies would improve the Bank’s development effectiveness.

Progress in Evaluation, Control, and Risk Management

Complex challenges for evaluation, control, and risk management have emerged as the Bank’s strategy evolved toward a more comprehensive approach. In 1997, poverty reduction was incorporated in the Bank’s mission statement and the Strategic Compact channeled incremental resources toward social development, governance reform, and increased participation with the private sector and the civil society. In parallel, CODE endorsed a new evaluation strategy designed by

OED in consultation with management. It sought to respond to an unprecedented transformation in the strategies, products, services, and internal culture of the Bank.

By creating incentives for quality work and measuring progress, QAG and OED contributed to significant improvements in operational outcomes. The quality of economic and sector work also improved after independent and self-evaluation of analytical and advisory services got under way. As the country rather than the project became the unit of account, OED’s country assistance evaluations (CAEs) became a standard evaluation product. New-style OED sector and policy evaluations have informed the Bank’s new sector strategies and the recasting of operational policies. In parallel, development effectiveness has become a focus of the Bank’s research program, and the World Bank Institute (WBI) has embarked on a major expansion of its training programs and collaborative partnerships so as to inform partners about the lessons of development experience.

Remaining Challenges

Major challenges remain despite the substantial progress achieved. Monitoring, evaluation, and risk management capacities must be enhanced further to meet the demands of the corporate strategy and the conclusions of the recent *United Nations Conference on Financing for Development* in Monterrey, Mexico. The consensus emphasizes results at country and global levels, knowledge for development, and partnership. Self-evaluation has yet to move to the “higher plane” of country, sector, and global programs. Most of the Bank’s knowledge programs have yet to embody clear objectives, precise modalities, and explicit fiduciary requirements; and Bank country assistance strategies and collaborative multicountry programs do not delineate precisely the shared goals, distinct accountabilities, and reciprocal obligations of partners.

This means that development effectiveness monitoring and evaluation work within the Bank should converge with similar efforts under way within the development community at large. For all development agencies, monitoring and evaluation remains the weakest link in the risk management chain. The Evaluation Cooperation

Group of the Multilateral Development Banks (MDBs) has achieved considerable progress with respect to retrospective evaluation methods at project level. But little progress has been made in building monitoring and evaluation systems within developing countries. Nor has harmonization of country program evaluation processes and methods been achieved among MDBs, bilateral donors, and U.N. partners.

At the global level, the Bank has endorsed the Millennium Development Goals (MDGs). They commit the development community to specific aggregate targets based on agreed principles of development cooperation, but do not entail specific country and sector programs, nor commitments of resources or reform of developed country policies on debt, aid, trade, and migration. While the MDGs should help to concentrate scarce aid resources on common priorities, the potential mismatch between ends and means involves considerable risks.

To manage these risks, a collaborative approach to performance measurement along with commensurate capacity development efforts will have to be deployed. The choice and implementation of cost-effective and credible approaches to *monitoring inputs, outputs, and outcomes and evaluating the results* of development assistance have become strategically important for the whole development community. The Bank should reach out to partners to promote harmonization and capacity development for performance measurement and management.

Internally, the recent focus at the highest levels of Bank management and the international development community on demonstrating results should help overcome the obstacles that have stood in the way of the results-based management practices recommended by OED. In particular, the time has come for the Bank to develop practical performance measures at a project, country, and corporate level, for internal use, as well as a recast Bank scorecard, to be used as an external reporting vehicle.

Recommendations

Results-based management will be realized only if self-evaluation is strengthened at all levels, if in-

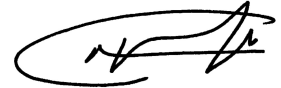
dependent evaluation contributes to improving methods and indicators for country, sector, and global monitoring and evaluation, and reliable data generated by countries are incorporated in measuring results. The agenda for improved monitoring and evaluation includes providing guidance on the overarching evaluation framework and filling gaps in the tracking systems of lending and nonlending services, grants, and partnerships. More transparent Country Assistance Strategy (CAS) and Sector Strategy Paper (SSP) processes should be complemented by rigorous self-evaluation. Against this background, the AROE recommends that actions be taken on four fronts, three by management and one by OED:

- *At the corporate level*, senior management should develop performance measurements that will enable management and the Board to track Bank achievements at project, country, and global levels in relation to the MDGs. It should implement an integrated risk management framework to facilitate the administration of financial, operational, developmental, and reputational risks faced by the Bank. In this context, it should make explicit the strategic risks and rewards of operational policies and corporate strategies.
- *Strengthen the evaluation and control framework for sector strategies, partnerships, and grants*. The operational shift to the higher plane of policy and sector strategies needs to be backed by adequate network accountability for sector performance measurement and management of instruments and services. Management and OED should develop a concise operational policy that provides an overarching evaluation framework. Project-level monitoring should be treated as a fiduciary responsibility to ensure that adequate monitoring is in place before Board approval. The “evaluability” of grant-financed activities should be improved. For all partnerships, clearer distinctions in accountability and more accurate specification of quality assurance, evaluation, and oversight should help to minimize conflicts of interest and increase responsiveness to developing country needs.
- *At the country level, risk assessment and learning should be improved* by developing a com-

prehensive approach that incorporates fiduciary assessments into the Country Assistance Strategy. The results should inform the Country Assistance Strategy. Self-evaluation at the country level needs to be integrated into country programs while CASs need to select verifiable indicators that are solidly grounded in country processes.

- *Finally*, OED will need to improve and disseminate more widely its evaluation methodologies for country evaluations and increase the transparency of the evidentiary basis of its evaluations. It should help Regions and Network Anchors incorporate self-evaluation in country assistance and sector strategies, which would

improve the ability of the Bank to measure its performance without compromising the independence of OED's analyses. It should address the independent evaluation needs of global and regional programs; better monitor and evaluate compliance with safeguards; target more effectively the dissemination of sector and thematic evaluations, and use a wider range of dissemination tools.



ROBERT PICCIOTTO
DIRECTOR-GENERAL
OPERATIONS EVALUATION

ABBREVIATIONS AND ACRONYMS

AAA	Assessment of advisory activities
ACS	Activity completion sheet
APL	Adaptable Program Loan
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 procedure
CAE	Country Assistance Evaluation
CAS	Country Assistance Strategy
CDF	Comprehensive Development Framework
CFAA	Country Financial Accountability Assessment
CODE	Committee on Development Effectiveness
COSO	Committee of Sponsoring Organizations of the Treadway Commission
CPAR	Country Procurement Assessment Report
CPIA	Country Policy and Institutional Assessment
CPPR	Country Portfolio Performance Review
CSA	Control Self-Assessment
CY	Calendar year
DEC	Development Economics Vice Presidency
DECRA	DEC Research Advisory Staff
DECRG	DEC Research Group
DGF	Development Grant Facility
EAP	East Asia and Pacific Region
ECA	Europe and Central Asia Region
ECD	Evaluation capacity development
ED	Executive director
ERM	Enterprise risk management
ES	Executive summary or evaluation summary
ESMAP	Energy Sector Management Assistance Program
ESW	Economic and sector work
FSAP	Financial Sector Assessment Program
FY	Fiscal year
HD	Human development
HIPC	Heavily Indebted Poor Countries
IAD	Internal Audit Department
ICR	Implementation Completion Report
IDA	International Development Association
ILI	Intensive Learning ICR
IMF	International Monetary Fund
ISDS	Integrated safeguards data sheet
LIL	Learning and Innovation Loan
MAR	Management Action Record
MDBs	Multilateral development banks
MDGs	Millennium Development Goals
MMR	Monthly Management Report
M&E	Monitoring and evaluation
OD	Operational Directive

OED	Operations Evaluation Department
OP/BP	Operational Policy/Bank Procedures
OPCS	Operations Policy and Country Services
PAD	Project Appraisal Document
PAR	Portfolio at risk
PATS	Partnership Approval and Tracking System
PCD	Project Concept Document
PER	Public Expenditure Review
PID	Project information document
PIU	Project Implementation Unit
PPAR	Project Performance Assessment Report
PPIAF	Public-Private Infrastructure Advisory Facility
PRSC	Poverty Reduction Support Credit
PRSP	Poverty Reduction Strategy Paper
PSR	Project Supervision Report
QACU	Quality Assurance and Compliance Unit
QAE	Quality at entry
QAG	Quality Assurance Group
QER	Quality Enhancement Review
QET	Quality Enhancement Team
RBM	Results-based management
RSB	Research support budget
SAP	Accounting system (World Bank)
SMART	Safeguard Management and Review Team
SOE	State-owned enterprise
SPRITE	Social policy reform in transition economies
SSP	Sector Strategy Paper
SWAP	Sectorwide approach
S&T	Sector and thematic
TQC	Trust Fund Quality Assurance and Compliance Unit
WBI	World Bank Institute
WDR	World Development Report

1

INTRODUCTION

This year's Annual Report on Operations Evaluation focuses on both risk management and monitoring and evaluation, which are closely related. Risk management involves identifying, assessing, mitigating, and controlling the risks to development objectives and the Bank's reputation. Monitoring and evaluation measure whether intended outputs, outcomes, and impact are being achieved. A results-based management system needs risk management, sound monitoring and evaluation, and reliable performance indicators from client countries.

Over the past ten years, the way the Bank does business has changed substantially to take account of its more demanding external environment. New approaches include matrix management, decentralization, new instruments, greater focus on broad partnerships, and greater transparency. As a result, its risk profile has become more complex and demands on managing risks have become greater. Because OED is concerned with the Bank's development effectiveness, this report focuses on the systems in place to manage the risks to achievement of development effectiveness and the associated reputational risks, and on the monitoring and evaluation systems in place to measure whether the development objectives have been achieved. This report does not deal with financial risks or business operating risks, nor with the reputational risks arising from either of these.

In the early 1990s, the Bank tracked projects and lending amounts as the primary measures of success. Management focused mainly on loan approvals, with less attention and few incentives for identifying and assessing risks, monitoring implementation, or measuring outcomes. By the end of the decade, the Bank had become more focused on measuring the quality of lending. Structures and systems had been put in place to address fiduciary, operational, and safeguard risks of lending, and to assess the quality of some nonlending activities. Both self-evaluation and independent evaluation had improved. The internal culture began to shift toward managing risks and assessing progress toward goals. But an unfinished agenda remains. The Bank has yet to develop a systematic approach to assessing risks at the country level, or to establish links between inputs or

the quality of Bank operations and outcomes at country, sectoral, and global levels.

This year's Annual Report on Operations Evaluation (AROE) follows the mandate established for OED twenty-six years ago to assess, on an annual basis, progress in monitoring and evaluating the development effectiveness of the World Bank. It does not report on the evaluation findings per se. These are the subjects of another OED flagship report: the Annual Review of Development Effectiveness (OED 2001a).

As in last year's report and at the Board's request, this year's AROE examines both risk management and monitoring and evaluation, and uses the COSO internal control framework as the or-

ganizing principle (box 1.1). The report also draws on reports from Operations Policy and Country Services (OPCS) and the Quality Assurance Group (QAG); OED evaluations, ongoing and completed; and interviews with staff in Regions, Network Anchors, Development Economics Vice Presidency (DEC), and World Bank Institute (WBI).

The AROE takes stock of how far the Bank has come over the past ten years in developing both risk management and monitoring and evaluation. In particular, this report examines the extent to which the Bank has adopted the recommendations of recent AROEs, as well as the extent to which OED has adopted recommendations pertaining to its own products and services.

Box 1.1

COSO

The main objectives of using the COSO process are to ensure: (i) Reliability of financial reporting; (ii) Compliance with applicable laws and regulations; and (iii) Efficiency and effectiveness of operations.

The AROE focuses on this third objective and asks: to what extent are the controls in place in the World Bank sufficient to ensure both the efficiency and the development effectiveness of operations?

The five components of COSO are used as an organizing principle in the report:

- *Control environment* establishes the foundation, the structure and incentives, for the internal control system, based on clearly articulated institutional objectives.
- *Risk assessment* involves the identification and analysis of relevant risks to achieving objectives and forms the basis of how risks should be managed.

- *Control activities* are the policies, procedures, and practices that ensure that management's directives are carried out and the institution's mandate is fulfilled.
- *Monitoring* involves external oversight of the internal controls or use of independent methodologies within the process to measure whether controls are adequate and the objectives are being met.
- *Information and communication* support the other four components by communicating the necessary information in a timely, digestible way that enables staff to perform their functions.

Source: COSO 1994, pp. 3–5.

2

THE CONTROL ENVIRONMENT

This chapter reviews the evolution of the mandate, structures, functions, and internal culture of the Bank. All have an impact on the control environment, which establishes the foundation for the internal control system. Many positive changes have occurred in all four areas over the last ten years. Remaining challenges are to improve the control framework for new knowledge activities and partnerships; develop a more coherent framework for evaluating programs and assessing risks; and ensure better monitoring to measure results.

The *mandate, structures, functions, and internal culture* of the Bank have undergone profound transformations in the last decade. The 1992 Wapenhans report commissioned by President Lew Preston sounded a wake-up call for the World Bank as well as other multilateral development banks. It reported a sharp decline in the quality of the portfolio and lending outcomes¹ and recommended that the Bank: (i) emphasize the quality of loans, focusing on quality at entry and risk analysis; (ii) focus more on portfolio performance management; (iii) emphasize a country focus for project implementation issues; (iv) promote borrower commitment and accountability; and (v) strengthen the role of independent evaluation. Over the past decade, the Bank has responded with new initiatives, procedures, structures, and resources (see box 2.1 for Milestones).

The Bank's mandate has, since 1990, become more clearly focused on poverty reduction. The Bank has joined with the rest of the development community in an ambitious venture: helping client countries reach the Millennium Development Goals (MDGs) by the year 2015 (box 2.2). Units throughout the Bank (operational, support, research, training) have begun to realign their work programs to reflect this commitment to explicit development outcomes at the global level. The formulation of the MDGs offers the opportunity—only partially exploited to date—to monitor and evaluate the outcomes of the Bank's assistance in a context of interlocking development partnerships. This is a formidable undertaking because the linkages between individual MDGs and the sectoral and thematic priorities of the Bank are manifold and complex.

The focus on MDGs introduces challenges that include aligning the MDGs at the country

Box 2.1**Milestones in the Control Framework for Risk Management**

- 1992:** Wapenhans report recommended focus on quality at entry, supervision, and quality of portfolio.
- 1993:** Action plan presented to Board to improve development effectiveness (following Wapenhans report).
- 1994:** Inspection Panel becomes operational.
- 1995:** Quality Assurance Group created to monitor quality at entry; develops system for Portfolio at Risk; COSO adopted as the Bank's internal control framework to help assess its controls and make improvements; Control Self Assessments workshops started.
- 1996:** Internal Audit Department charged with assessing internal controls in the Bank.
- 1997:** Quality Assurance Group begins to assess quality of supervision and economic and sector work; Strategic Compact: goals articulated; Bank reorganized, decentralized; matrix management introduced.
- 1998:** Loan Administration Change Initiative introduced; financial management specialists hired; Development Grant Facility established; Quality units established in some Regions and Network Anchors.
- 1999:** Enterprise Risk Management (risk assessment tool) introduced into financial complex; Monitoring and Evaluation Task Force established.
- 2000:** Department of Institutional Integrity merges the former Corruption and Fraud Investigations Unit and the Office of Business Ethics and Integrity; Risk Management Task Force findings; Partnership Council and Business Partnerships and Outreach Group established.
- 2001:** Quality Assurance and Compliance Unit established in ESSD Network for safeguard compliance; Partnership Assessment Tracking System (PATS) put in place; Guidelines for financial assessments and financial monitoring reports revised; Integrated Safeguard Data Sheet, Risk Assessment and Compliance Monitoring Sheet introduced.
- 2002:** Trust Fund Reforms proposed; Trust Fund Quality Assurance and Compliance Unit established; Risk Management Secretariat to Management Committee established.

level with priorities of client country and donors and attributing results to individual partners. Improved aid coordination² and effective use of the Comprehensive Development Framework (CDF) processes and Poverty Reduction Strategy Papers (PRSP) can foster this alignment by putting poverty reduction at the center of the development effort and establishing monitorable indicators to measure progress. This is connected to the drive toward a Knowledge Bank and the increased reliance on partnerships to implement new initiatives, with client countries, with development partners at the regional and global levels, and within the Bank. Both the control environment and the evaluative framework for these activities, as previous AROEs have pointed out,³ remain underdeveloped. OED currently has two evaluations under way that address several of these issues.⁴

Structures and functions have been introduced throughout the Bank to identify risks, to assess

their importance, and to manage them (box 2.1). The list is long and impressive and these initiatives have served to reinforce the risk control framework. A key initiative relating to risk management has been the creation of the Quality Assurance Group (QAG) in 1995. Even as its role has evolved (e.g., toward assessing the quality of analytical and advisory services), QAG's methodology and findings have been internalized throughout operations. In addition, since last year, structures and functions have been introduced to bridge gaps identified in previous AROEs (OED 2001b, pp. 11, 22) on safeguard compliance, trust fund use, and partnerships. These include the Integrated Safeguard Data Sheet, the Trust Fund Reform Program, the Trust Fund Quality Assurance and Compliance Unit, and the Partnership Assessment Tracking System. It is still too early, however, to judge their effectiveness.

The internal culture has changed. Staff and managers recognize the need to identify and as-

Box 2.2**Increasing Focus on Poverty**

The 1990 World Development Report (WDR) on Poverty was followed by a Policy Paper and an Operational Directive that articulated the Bank's mandate to combat poverty. A decade later OED found that this strategy had a significant and positive impact on the Bank's operational work.

At a country level, Country Assistance Strategies (CAS) have improved their poverty focus; the second CAS retrospective found that the percentage of CASs with satisfactory or better poverty focus increased between FY96 and FY99 from 20 to 67 percent. In addition, the Poverty Reduction Strategy Paper (PRSP) introduced in 1999 is to provide the basis for a country-driven, results-oriented framework for assistance. As of end-2001, 52

low-income countries had completed either Interim PRSPs or full PRSPs, and another 7 were under way.

At a sectoral level, the nine sector strategy papers produced between FY97 and 01 had explicit links to poverty reduction.

At a global level, the Bank published a second WDR on Attacking Poverty in 2000. In the past two years, the Bank has been refining its strategic framework to customize its assistance for middle- and low-income countries and for low-income countries under stress. In addition, DEC, working with Networks, has identified a set of indicators to measure progress in reaching the Millennium Development Goals.

sess risks and to put in place mitigation measures. Project appraisal processes and implementation progress reporting reflect risk assessment and compliance. Managers receive regular reports incorporating risk assessments and have responsibility for managing the risks. At the same time, the increased focus on risk assessment (combined with inadequate support systems and mixed ownership of institutional priorities among Regional managers and borrowing governments) has fostered risk-averse behavior. The increased costs of complying with environmental and social safeguard policies are resulting in Bank and borrower avoidance of high-risk, high-reward operations.⁵ A shift to country-level risk management and attention to strengthening fiduciary and safeguard

processes at a country level may help to contain the cost of doing business in individual projects and avoid enclave operations that cannot easily be scaled up. In addition, the tightening of mechanisms for safeguard compliance will have to be matched by adequate systems for measuring outcomes, mediating conflicts, and building borrower capacities to deliver socially and environmentally sustainable results. There is not yet a coherent approach to risk management throughout the institution, although senior management has begun to develop an integrated approach to address remaining significant gaps in the accountability and learning system. Both the gaps and measures to address them are described in Chapters 3–6.

3

RISK ASSESSMENT

Risk assessment is the identification and analysis of risks to achieving objectives, to help determine how they should be mitigated and managed. This chapter finds that risk assessment for lending activities is more advanced than for nonlending activities (such as trust funds, grants, and partnerships), although recent progress has been made in these areas as well. Overall, risk assessment for individual lending and nonlending products and activities is better than for country programs as a whole. At the corporate level, risk management is still at an early stage.

This chapter has four sections. The first examines how risks are assessed for lending operations in the pipeline (before Board approval) and during implementation. A section covering risk assessments for trust funds, grants, and partnerships follows. The third section looks at risk assessment at the country level, and the final section covers risk assessment at the corporate level.

Lending: Risk Assessment for Projects in the Pipeline and under Implementation

For projects in the pipeline, risk assessments ten years ago were addressed through quality control processes, involving peer review and several layers of managerial review. In projects for which economic rates of return were estimated, the impact of development effectiveness risks was expected to be assessed through sensitivity analyses.¹ These continue to be done where feasible, while greater awareness of other, specific risks—fiduciary and

safeguard—has led to a more comprehensive approach. For example, all investment lending must have a financial management assessment, under guidelines revised in FY02, that identifies and categorizes fiduciary risks and proposes mitigating measures to be undertaken prior to or during implementation. Similarly, for safeguards, new systems for ensuring compliance and for improved risk assessment were put in place starting in FY01. The risk assessment includes a safeguard risk review in each Region (box 3.1), with assistance from the Quality Assurance and Compliance Unit (QACU). Network Anchors support quality enhancement on demand through Quality Enhancement Reviews. QAG assessment of these processes in Quality at Entry (QAE) exercises found greatly enhanced quality in fiduciary aspects, and more modest progress on the quality of risk assessments and monitoring and evaluation systems (table 3.1).

Box 3.1**Regional Approaches to Risk Assessment**

Each Region has developed methods for assessing risks. East Asia (EAP) pioneered the use of readiness filters, now used in both the Middle East and North Africa (MNA) and Europe and Central Asia (ECA) Regions, and has started a broader Country Portfolio Performance Review, carried out quarterly, with particular focus on safeguard and fiduciary risks and at-risk projects. EAP carries out Special Operational Reviews to identify issue- and country-based risks, such as child labor. In MNA, the readiness filter aims to reduce portfolio riskiness, improve quality of new projects, and accelerate implementation, and consists of about eight indicators tailored to each

country. In addition, high-risk projects in the lending program have a special review of the Regional Operations Committee. In the Africa Region, risk criteria include size of project, type of project, country, and sector. ECA and the Latin America and Caribbean (LAC) Regions have a regional risk management committee with representatives from within and outside the Region to review all projects that are considered higher risk. LAC also reviews risks of ongoing projects twice a year as part of the portfolio review. South Asia Region, by contrast, is more decentralized, with greater reliance on risk assessment by country managers.

Some types of risks are normally assessed in adjustment operations, such as ownership of policy reform, commitment, and capacity to comply with conditionality. Other risks are less consistently identified and assessed, including the linkage of policy reforms supported under the operation to poverty reduction; acceptance of the reforms by the client country's population; and fiduciary and safeguard risks.² All of these need further attention. The revised guidelines under preparation for adjustment lending should consider how adjustment operations will deal more directly with the high-risk safeguard and fiduciary regimes that should be documented in the relevant CAS.

Monitoring and evaluation (M&E) are integral to a robust risk management system but remain a major weakness in project design. Without good M&E, it is not possible to assess the links between lending assistance and development outcomes and to assert that development objectives have been met. Project risk assessments would be more reliable if they were based on past experience and demonstrable results. Risk assessment could be further enhanced by a better appreciation of the risks at a country level, including political, social, technical, financial, and natural risks, where they are relevant to each operation. This is discussed below.

Table 3.1**Quality of Risk Management Processes**

	Percent satisfactory or better			
	CY97	CY98	CY99	CY00-FY01
Overall quality	82	86	89	94
Fiduciary aspects		77	89	95
Risk assessment and management	72	72	73	78
Selected safeguards				
Environmental assessments		94	92	91
Natural habitats		93	89	75
Forestry		75	100	89
Involuntary resettlement		90	78	87

Source: OAG data.

Note: Safeguard percentages are based on small samples and differences over time are not statistically significant.

Projects under implementation. In 1992, the Wapenhans Report recommended moving away from a project-by-project approach to portfolio management and adopting a country-based approach. This was done with the introduction of Country Portfolio Performance Reviews (CPPRs), which look for countrywide systemic risks to development effectiveness of ongoing projects and measures to address them. CPPRs are still used today and are supplemented by the Portfolio at Risk (PAR) system, introduced in 1996 by QAG. The PAR identifies projects at risk as well as the extent to which actions have been taken on risky projects. These data are now used not only at a country level, but also by Regions, Network Anchors, and at the corporate level. The PAR has thus served to sharpen the institution's awareness of development risk management. On the other hand, the absence of quantitative performance indicators related to the weakness of project M&E lowers the credibility and increases the complexity of the Bank's performance measurement approach.

In turn, the *reliability of the systems for assessing* portfolio at risk may have decreased over time as staff faced incentives to avoid signaling risk.³ QAG is proposing revisions to improve the reliability and realism of assessed outcomes. In addition, monitoring of risk mitigation by the borrower during project implementation has not kept pace with risk assessment during the appraisal process. It was only in mid-FY02 (four years after the Bank introduced guidelines for financial assessments of investment projects prior to Board approval), for example, that the Bank added the requirement to report on the status of financial management in supervision reporting. OED, too, could enhance accountability by more systematic reliance on verifiable performance indicators in its evaluations and more rigorous monitoring of fiduciary and safeguard compliance in Bank operations.

Grants, Trust Funds, and Partnerships

Grants have become increasingly important in the Bank. Although they do not entail financial risks to the balance sheet, they do carry many of the same development and reputational risks as lending operations, as well as other risks (box 3.2). Over

the years, OED has highlighted weaknesses in the control framework and evaluation of grants.⁴ Since 1998, in particular, these aspects have been strengthened. The 1999 operational policy for grants has made the eligibility criteria more transparent. The Development Grant Facility (DGF), also set up in 1998, has improved progress reporting and introduced stricter guidelines for exit criteria and periodic monitoring. In line with OED recommendations contained in a grant process review,⁵ evaluations of major programs have been carried out, which help ensure better accountability in the use of funds, although the degree of independence has varied substantially across these evaluations.⁶ Further improvements are called for in four areas highlighted in the 1998 review. These include use of outside experts to enhance objectivity and strengthen technical expertise in the approval process; strengthening further the arm's-length relationship in programs that are substantially in-house to avoid real or perceived conflict of interest;⁷ strengthening and monitoring exit strategies, to avoid indefinite dependence on Bank funding; and uneven grant completion reporting, financial reporting, and evaluation to strengthen accountability.

Trust Funds. In FY93, OED noted that few trust funds, apart from a few large ones, were subjected to consistent procedures for approvals, reporting, or self-evaluation. Successive OED studies have traced incremental improvements in these processes. But in the wake of internal reviews and audits, the Bank undertook a reform of trust fund management in FY02 to simplify their use and improve controls at all levels. The reform includes: (i) integrating trust fund allocations better into resource planning of the Bank; (ii) simplifying and standardizing criteria for eligibility and replacing ex ante approvals by donors with more thorough and timely ex post reporting; (iii) introducing an administrative fee. The action plan includes strengthening the monitoring and reporting on the use of the trust funds, a completion note, and more comprehensive and detailed reporting to donors. A Trust Fund Quality Assurance and Compliance Unit (TQC) has been set up. Training and a certification procedure will soon be put in place for staff responsible for use of the trust funds. These initiatives are wel-

come and should be monitored in the future by the Internal Audit Department (IAD).

Partnerships have grown dramatically in the past ten years. Reliable figures on the total number of partnerships, the administrative funds involved, or the flow of managed funds through partnerships are not yet captured systematically in the Bank's information systems. In recognition of the risks involved (box 3.2), the control environment has been strengthened. In FY99 the Partnership Council was established to provide advice and share information, and in FY00 the Business Partnership and Outreach Group was reorganized to provide guidance on partnerships with private sector entities. In FY01, explicit criteria for new partnerships were established. In FY02, a pilot Partnership Approval and Tracking System (PATS) was put in place to integrate several processes into one, including grant application, a trust fund initiating brief, approval of senior management through a concept note, and risk assessment for partnerships. The PATS will

also provide centralized information in real time on the Bank's regional and global partnerships.

Some gaps need further attention. Most partnerships are open ended, lack clear performance objectives, and do not delineate precisely the distinct accountabilities and obligations of partners. The lack of distinction between partnership and global programs, or among global, institutional, and regional partnerships, makes it unclear when criteria for global and regional partnerships apply. Until PATS can capture all ongoing partnerships, senior management cannot have a complete picture of aggregate or systemic risks of Bank partnerships. Finally, although Network Anchors are responsible for preparing new programs and overseeing ongoing ones, accountability is diffuse for global strategy formulation and for "quality at entry" and other aspects of quality assurance currently shared among Legal, Resource Mobilization and Cofinancing, OPCS, and Strategy and Resource Management. This fragmentation, in turn, clouds accountability for outcomes.

Box 3.2

Grants, Trust Funds, Partnerships: What They Are, What Risks They May Involve

Grants are resources provided out of Bank income for special programs. They include country-based grants such as the Institutional Development Fund; post-conflict support to particular countries; and grants made in partnership with other donors, such as the Consultative Group on International Agricultural Research.

Trust funds are resources provided by other donors or agencies, entrusted to the Bank for use in a wide range of modalities. They include the Heavily Indebted Poor Countries Trust Fund and the Global Environmental Facility; programmatic trust funds, such as Japan's Policy and Human Resources Development Fund; consultant trust funds; and single-purpose trust funds.

In addition to development effectiveness risks, risks of both grants and trust funds include: dependency, using outside support for activities that would otherwise make claims on Bank or client resources; strategic, arising from approving grants or administering trust funds for purposes that may not align with the Bank's priorities; and cost-effectiveness, arising from the costs of administering smaller grants or trust funds, with diverse

arrangements. In addition, grants may not be limited to situations where lending is not appropriate and no other source of funding is available (subsidiarity principle). For trust funds, fiduciary and reputational risks can arise from a failure to fulfill agreements with donors.

Partnerships are operational relationships with other institutions, entailing shared objectives but distinct accountabilities and reciprocal obligations. They may be partly funded through the DGF or trust funds. They bring important benefits, including opportunities for mutual learning about development assistance, leveraging administrative resources and Bank financing, broadening reach, and facilitating delivery of regional and global public goods. But they also carry significant risks, including development effectiveness, if partners prove ineffective or inefficient in their role; reputational, if partners engage in activities inconsistent with the Bank's mandate; and unfair advantage risk (partners should accept limits to publicity of their partnership with the Bank).

Risk Assessment at the Country Level

Many elements of risk assessment at the country level have been carried out for years. These include credit risk analysis (integral to decisions on lending to IBRD countries) and the use of performance assessments as a guide for allocating IDA resources. More recently, the Bank increased and improved instruments for core diagnostics, of which two are used to determine eligibility for Poverty Reduction Strategy Credits (PRSCs) (box 3.3). The choice of instruments has broadened and currently provides a substantial tool kit for country risk assessment.

Quality management of some of these instruments could be further improved and management is taking actions: CPARs and CFAAs have been mainstreamed as core ESW products, which should provide greater management attention to their quality and delivery, removing gaps in coverage, and their systematic use in preparation of Country Assistance Strategies. Additional improvements should occur as guidelines are refined on the scope and content of these exercises and Bank staff gain more experience and learn from best practices.⁸ Another area that needs attention is potential overlap with other donors, including the IMF, who have relevant diagnostic instruments, in order to avoid creating a burden on client countries.⁹

The CAS, since its inception in FY91, has included a section on risks, although the assessments need further sharpening and coverage. The

CASs do not generally distinguish between country risks and risks to the implementation of the Bank's strategy, and OED found in its Country Assistance Evaluations that CASs often minimize risks or fail to mention important ones (Annual Review of Development Effectiveness [ARDE] 00, pp. 22–23).

Following up on the work of the Middle-Income Countries Task Force, Europe and Central Asia Region (ECA) has begun to integrate public expenditure reviews (PERs), CFAAs, and CPARs into a single assessment, and East Asia and Pacific Region (EAP), specifically in the Philippines, is trying to do this in conjunction with the CAS. In other Regions, welcome integrative efforts are also under way. Such integrated analysis will permit a more comprehensive view of country risk and development challenges. A yet more comprehensive approach is suggested in box 3.4.

Risk Assessment at the Corporate Level

Senior management is putting in place an integrated approach to risk management for Bank activities. In 2000 the Risk Management Task Force recommended that the recently created Management Committee address risks related to the Bank's development effectiveness and reputation. As OED pointed out, this is a critical and appropriate role for senior management. For risk management at the senior level to be effective, the strategic risks of policies and corporate goals need to be de-

Box 3.3

Risk Assessment Tools at the Country Level

Country Procurement Assessment Reports (CPARs) and Country Financial Accountability Assessments (CFAAs) are tools to diagnose procurement systems and financial accountability systems in client countries. CPARs have been carried out for many years as a fiduciary requirement for Bank lending. In 1998 new guidelines broadened them to serve a developmental objective as well. The first CFAAs were in 1998 and as of end-FY01, 22 had been carried out; they have continued to evolve in scope and content since their introduction.

Starting in FY01, the CFAA and CPAR are two of the five core diagnostic economic and sector work (ESW) elements for each borrower. Public Expenditure Reviews are a third core diagnostic

report, which analyze the equity, efficiency, and effectiveness of public spending allocations and management (the other two core diagnostic reports are poverty assessments and country economic memorandum/development policy review).

In addition, the Bank and Fund have a joint program for assessing the vulnerability of a country's financial system. The Financial Sector Assessment Program (FSAP) was started toward the end of FY99 as a response to the Asian crisis, and as of end-December 2001, 18 FSAPs had been completed and 18 were under way in client countries. The results of these assessments often feed into Bank assistance strategies, operational work, and policy dialogue on structural reform.

Box 3.4**Sample Checklist for Country Risk Analysis**

A comprehensive country risk assessment would include a checklist of *potentially relevant* risks, distinguishing between fiduciary (affecting efficacy and efficiency of Bank assistance) and developmental (those that should be addressed for capacity building purposes). For fiduciary risks, the analysis would be incorporated into the CAS, with a proposal for how Bank assistance could be tailored in light of those risks. The depth of the analysis could also be tailored to the Bank's past and potential exposure (largest borrowers). For any given country, not all of the risks listed below would be relevant, and the focus would be only on those risks that related to the development effectiveness of the proposed Bank assistance program.

- **Political:** extent of ownership of reforms, the nature and strength of political opposition; an analysis of winners and losers in a changing policy environment; the possibility of civil unrest.
- **Legislative:** extent to which current legislation could impinge on development programs/projects.
- **Social:** ethnic conflict; labor unrest; child labor issues; increased income inequality.
- **Financial:** creditworthiness.
- **Fiduciary:** public expenditure allocations; how expenditures are monitored, controlled, and reported on; procurement rules and practices.
- **Environmental:** legal and regulatory environment and institutional capacity for enforcement; implications of privatization of mines; industry; and other potentially polluting activities.
- **Natural disasters:** probability of hurricanes, floods, drought, earthquakes, pests, and crop disease that should affect decisions on investments in infrastructure.
- **Economic:** world prices; noncompetitive structures and behavior in suppliers of goods and services.
- **Reputational:** Partner with agencies that could have objectives inconsistent with the Bank's mandate.
- **Evaluation of assistance:** risks identified by analysis of past experience at country level.

financed, priorities established for which control gaps and overlap need attention, and accountabilities clarified throughout the institution for risk management at the country, sector, and project levels. In addition, better information flows in both directions (upwards about risks and downwards about priorities and tolerances) are needed.

The Board, through both the Audit Committee and CODE, will need to provide oversight to

this system and to confirm its request for OED and external auditors to attest to the adequacy of risk management arrangements. In particular, Board oversight bodies should ensure that the control mechanisms (including audit, evaluation, and control) are properly designed, are allowed to operate without interference, and are effectively connected to day-to-day management through appropriate mechanisms.

4

CONTROL ACTIVITIES

This chapter reviews the Bank's control activities, which comprise the policies, procedures, and practices ensuring that management's directives are carried out and the institution's mandate is fulfilled. This chapter reviews Bank guidelines, the processes of quality enhancement for lending and nonlending, monitoring and self-evaluation of projects, and the practices of Control Self Assessments (CSAs) and Enterprise Risk Management. Overall, control activities have been expanded and improved considerably in the past ten years. Management should continue its activities to address remaining weaknesses in project monitoring, control activities for economic and sector work, and accountability for CSA.

Bank Policies and Guidelines

The Bank's operational policies form the basic guidance for Bank staff on implementation of the Bank's development strategy. The process of converting the former Operational Directives (OD) to Operational Policies and Bank Procedures (OP/BP) has been ongoing since 1993, and as of 2001, most had been converted (box 4.1), with the notable exception of particularly complex policies. Of the five still in process (poverty reduction, adjustment lending, indigenous people, involuntary resettlement, and cultural heritage), when last year's AROE was produced, only the OP/BP for involuntary resettlement has been sent to the Board for approval. There remains a need to streamline the public consultation and internal review processes for policy review.

Quality Assurance/Enhancement of Lending

Quality Assurance Group (QAG). Since its creation in 1995, QAG has occupied a central role in quality assurance related to Bank operations. Increasingly, quality assurance has been taken up by Regions and to a lesser extent the Network Anchors, using QAG criteria, data, and techniques for monitoring quality at entry, quality of supervision, and quality of ESW. Quality enhancement reviews (QERs), introduced by QAG to improve the quality of lending operations at an early stage in their design, are now used by most Regions and some Network Anchors. Both Regions and Network Anchors actively follow up trends in QAG indicators. Analysis of QAG findings is in the next chapter.

Box 4.1**Conversion and Updating Process of OP/BP: Progress Reporting**

Although senior management and Executive Directors are regularly provided with information on conversions under way, with expected completion dates, further information is needed to strengthen oversight of the process: (i) how many and which existing ODs and their predecessors, Operational Manual Statements, still need conversion to OP/BPs, and which ODs and their predecessors are no longer valid and will not be converted; (ii) a list of all conversions in process, and their stage of conversion; (iii) a timetable for the conversion of those under way; and (iv) a list of OP/BPs needing an update and a proposed timetable for this.

Quality Assurance Activities in the Regions and Network Anchors. The Regions have taken up the role of quality assurance for lending, both pipeline and portfolio, although there are large differences among the Regions in terms of structures, focus, and functions (see Annex 2). Five Regions (except EAP) initiate QERs for projects in the pipeline for between 5 to 20 percent of projects. All Regions monitor the riskiness of their portfolios (box 3.1); and all have processes for focusing on problem areas, by country and by sector.

Network roles in quality enhancement vary. The Human Development (HD) Network is the most active, initiating them on request for roughly 70 percent of HD projects in the pipeline. All Network Anchors monitor, usually on a quarterly basis, the quality of their respective portfolios and contact sector managers in the Region to discuss problems.

Regardless of the formal accountability for quality, it is clear from interviews with staff in both Regions and Network Anchors that, in practice, Regions bear the major responsibility for quality. Network Anchors provide quality enhancement on demand through cross support, peer review, and QERs. Although they monitor quality on a regular basis, staff in anchors and in sector “quality units” do not “control” quality at entry or quality of the portfolio, but function instead as a support service. The extent to which network management can be held accountable for quality assurance remains unclear.¹

Support activities for safeguard compliance have been strengthened. It was only in 1997 that the concept of safeguard policies was introduced, and ten policies relevant to safeguards were identified. Prior to that, compliance with individual policies was ad hoc and inadequately monitored. In FY01,

the Quality Assurance and Compliance Unit (QACU) was established in the Environment and Socially Sustainable Development Network with substantial resources to ensure appropriate risk assessment and facilitate compliance with guidelines. The Safeguard Management and Review Team (SMART) was set up in 2001 as a safeguard sector board, with representatives of QACU, OPCS, the legal department, and the Infoshop: its main function is to clarify guidelines and review issues arising in Bank policies and practices on safeguards.

In FY02, the Integrated Safeguards Data Sheet (ISDS) was introduced for use by task managers, identifying the safeguards involved in investment projects. The ISDS is to be part of the Project Information Document (PID), and thus available to the public. In addition, for internal use, selected projects also must have a Risk Assessment and Compliance Monitoring sheet. IAD recently audited the ISDS and found that the system is performing adequately but its reliability and efficiency have not yet been tested conclusively.

Monitoring and Self-Evaluation at the Project Level

Monitoring at project level is important for ensuring that management’s directives are carried out and for measuring whether the anticipated inputs, outputs, and outcomes have been achieved. Self-evaluation at the project level is critical to assessing causality between inputs, outputs, and outcomes, and measuring to what extent the institution’s mandate is being fulfilled through lending.

The system of monitoring in projects remains weak, in spite of repeated OED and QAG recommendations and episodic efforts over the last decade to strengthen it (box 4.2). This has been

attributed to a combination of lack of incentives and guidelines for Bank staff, and limited borrower capacity. To date, the Bank has responded with increased levels of training for Bank staff and a program of Evaluation Capacity Development in borrower countries (see Chapter 5, box 5.1), but as yet, there is no coherent framework, set of guidelines, core staff, database, or thematic group that can guide these efforts. To provide an overarching framework for both self-evaluation and independent evaluation, a concise Operational Policy (OP) should be prepared. This OP would serve as the logical and appropriate umbrella for subsidiary documents, including a Bank Procedure (BP) on the “Generation, Dissemination, and Utilization of the OED Findings,” and would also provide the framework for incorporating evaluation and self-evaluation policies and procedures into the upcoming OPs on the mainstays of operational work.

Project evaluation has seen better progress than monitoring, within the context of new guidelines for implementation completion reports (ICRs) issued in 1995. OED reviews the quality of the ICRs in terms of compliance with guidelines and ratings (on outcome, sustainability, and institutional development impact). Out of 286 FY01 ICRs reviewed as of end-February 2002, 92 percent were satisfactory, compared to an historic average of about 95 percent. Rating changes between ICRs and OED’s evaluation summaries have remained relatively constant for outcome in the last few years, with a net rating change (the percentage of ratings downgrades minus upgrades) of around 4 percent (see Annex 1 for details).

In FY00, the Bank reformed the ICR process further, with the introduction of Intensive Learning ICRs (ILIs); about 30 percent of the projects were to have a more in-depth report. In FY00 and

Box 4.2

Efforts to Strengthen M&E Over the Past Decade

The Bank has long recognized weaknesses in project-level M&E and made efforts to address them. The 1992 Wapenhans report pointed out weaknesses and a 1994 review by OED found lax enforcement of existing guidelines and directives and recommended enhanced staff training, recruitment of staff with M&E skills, and a focus on selected sectors. In 1995 OED noted some improvement in the incorporation of key performance indicators in appraisal reports. In 1996, the logical framework (log frame) was introduced, to enhance the development of appropriate monitoring indicators of inputs, outputs, outcome, and impact. A Bankwide Working Group on M&E was established in 1999, which made further recommendations, including developing evaluation capacity in client countries on a pilot basis.

In spite of these initiatives, OED found that monitoring systems for implementation had been put in place in only 60 percent of projects that exited in FY01, the same level as four years ago. The quality of the five components of the monitoring systems (clear objectives; a set of indicators; clear responsibility for data collection and management; capacity building; and feedback from M&E) has improved steadily over the last four years; in operations exiting in FY01, 59 to 78 percent of the components were judged to be high or substantial. QAG found in the

fourth Quality at Entry Assessment, covering operations through FY01, that monitoring implementation progress had improved to 91 percent satisfactory or better, but arrangements for measuring outcomes and evaluating impact remain, at 73 percent satisfactory or better, among the weakest aspects of quality. Perhaps most critical, during implementation, OED rated the use of performance indicators and borrowers’ provision of M&E data as substantial or high in only 40 percent and 36 percent (respectively) of operations exiting in FY01.

The M&E Improvement Program was launched by OPCS in FY01 to address some of these weaknesses, in Bank and borrower operations, including work at the national and sector level, as well as at the project level. The program includes efforts to increase the Bank’s own M&E capacity, focused initially in two Regions (AFR and ECA) and two sectors (HD and Community-Driven Development) and efforts to improve evaluation capacity development in borrower capacity in seven pilot countries, focused at a country and sector level, rather than a project level. If these efforts are successful, improvements should be visible in QAG and OED evaluations in the future.

Source: OED and QAG data.

the first half of FY01, only 10 percent of the ICRs were ILIs. OED found no significant difference in quality between the ILIs and more conventional ICRs (see Annex 1). Further improvements in ICR guidelines are needed. For example, they should be tailored to Adjustable Program Loans and Learning and Innovation Loans, as these are new products. In addition, more demanding standards for performance indicators should be introduced in order to enhance the credibility of evaluation ratings.

Quality Assurance for Economic and Sector Work (ESW)

Until FY97, there was little self-evaluation or independent evaluation of ESW; the main forms of quality assurance for ESW were peer reviews and managerial oversight. In July 1997, the Bank introduced Activity Completion Sheets (ACSs) in an effort to initiate systematic self-evaluation, but few ACSs are filled in. In 1998, QAG began to assess formal ESW, based initially on methodology proposed by OED, and later expanded to include dissemination and early indicators of likely impact. Based on a sample of reports, QAG has found an improving trend in overall quality, from 72 percent satisfactory in FY98 to 91 percent in the FY01 cohort.²

In the past, QAG has drawn its sample from a population that excluded informal or “unscheduled” ESW and process activities, such as country monitoring briefs, economic modeling, risk assessments, and work on the Country Policy and Institutional Assessment (CPIA). These activities comprised about one-quarter of total Bank resources on ESW over the FY98-01 period. In addition, substantial analytic work financed through trust funds such as the Energy Sector Management Assistance Program (ESMAP), Public Private Infrastructure Advisory Facility (PPIAF), and the Water and Sanitation Program, was not

part of the ESW population sampled by QAG for quality. These missing categories are expected to be included in new guidelines to be issued by OPCS. In addition, QAG has an ongoing pilot to assess the country-level integration of all analytic and advisory activities in three countries, with a plan to shift in this direction in FY04.

Control Self-Assessments and Enterprise Risk Management

In addition to control activities that focus on specific Bank products, the process of Control Self-Assessment (CSA), introduced into the Bank in 1998, identifies risks within the control environment and cultural factors in business units that can affect other controls, and develop action plans to address the risks. Some 142 CSAs had been carried out Bankwide as of March 2002 (of which 111 were in FY98–00), under the leadership of Controllers. The tool has the potential to strengthen risk management, but its confidentiality to the business unit level has made the accountability chain for managing the risks and monitoring follow up actions unclear. If systematic risks emerged from CSAs, senior management was not able to assess them because of the confidentiality of the process. In addition, coverage by CSA of business units within the institution has been uneven. These drawbacks are expected to be addressed by reporting to senior management on CSA results and by a plan that has been formulated to achieve more representative coverage, as well as broadening the scope of the CSA to include operational risks.

Enterprise Risk Management (ERM) is another control activity introduced into the financial complex in 1999 to perform a more detailed, thorough risk assessment of business processes. It has been piloted in one country office in the South Asia Region in 2001, although results are not yet available on its replicability.

5

MONITORING

Corporate monitoring has been strengthened in the last five years through QAG reviews; OPCS reviews of CASs, ESW, and Sector Strategy Papers; introduction and improved methodology by OED of country evaluations; and strengthened evaluations within DEC and WBI. Nevertheless, aggregate monitoring of inputs, outputs, outcomes, and impacts at a sectoral, country, and global level still faces significant challenges. With increased focus, both internal and external, on Bank performance, clear accountability for self-assessments and improved monitoring at all levels will be required.

Within the COSO framework, monitoring is the use of either independent methods or external oversight to measure whether the internal controls are adequate and the objectives are being met. This is distinct from project-level monitoring, which was discussed in Chapter 4 under control activities. In this chapter, we review the independent methods used within the Bank at an institutional level to determine whether the Bank objectives are being met at a country, sectoral, and corporate level, including training and research activities.

Monitoring Projects and ESW at the Corporate Level

The past five years have seen the introduction of improved monitoring at the corporate level of lending operations. Portfolio performance has been reported for over a decade, but QAG now uses a more systematic and real time methodology (the Portfolio at Risk system), drawing on

country portfolio performance reviews for the country-level perspective. Because QAG findings are presented to CODE and the Board at the same time as OED's ARDE, Board members can see trends in both the active and closed portfolio. In FY01, QAG reported on the quality of ESW for the first time.

As noted in Chapter 2, QAG has been instrumental in improving aggregate monitoring of the lending and nonlending portfolios. For the last five years, it has regularly reviewed and reported on the quality at entry (QAE) of lending operations, the quality of supervision, and the quality of ESW. QAE ratings, which are based on random samples of projects, are aggregated and tracked by management as indicators of trends in the health of the overall portfolio. Are they a reliable leading indicator of the project's eventual outcome? Of the 42 QAE-rated projects that have now closed with outcomes evaluated by OED, 93 per-

cent were satisfactory or better at the time of entry and 81 percent had satisfactory outcomes. Three hundred QAE-rated projects are still under implementation: 86 percent of these projects were satisfactory at entry and 83 percent are currently rated as “nonrisky” in the PAR system. OED ratings on outcome are historically some 10 percentage points lower than final supervision ratings, and although PAR status is not identical to supervision ratings, it is determined primarily by implementation progress rather than by development effectiveness considerations. Thus PAR ratings may also overstate the likely project outcome in terms of development effectiveness (see Annex 3 for discussion). Putting these factors together, QAE ratings are likely to be some 10 percentage points higher than final outcome ratings.

Now that the Regions have enhanced their role in quality assurance, QAG has shifted its monitoring focus to other areas. In FY02, for example, QAG focused its review of supervision on risky projects; did a special review on the quality of fiduciary ESW; and introduced a pilot assessment of advisory activities (AAA) on a country basis and a pilot assessment of sector board contributions to operational quality. This evolution reflects a shift to a more strategic level. Management should now clarify its longer-term vision for QAG’s evolving mandate.

In the last five years, OPCS has carried out a number of corporate-level reviews to examine, from an institutional perspective, the quality of Bank activities, including adjustment lending, ESW, Sector Strategy Papers, and CASs. These serve to review the extent to which these activities have complied with the Bank’s strategy and guidelines, to examine strengths and challenges, and to make recommendations for improvement.

Finally, since its creation in FY94, the Inspection Panel has played a critical role as an independent structure in monitoring compliance of Bank operations with safeguard policies. It was established to respond to complaints from external clients or affected people. It reports directly to the Board.¹ As of end-December 2001, the Inspection Panel had received requests to inspect 25 lending operations, of which twelve were investigated (the remaining were either ineligible or inadmissible claims). The resulting reports have

highlighted important control weaknesses, such as ambiguous policies and overemphasis on lending approvals at the expense of quality.

Sector and Thematic Monitoring and Evaluation

When they were created in 1996, the Networks were not given an explicit mandate to monitor sectoral or thematic inputs, outputs, or outcomes across their respective sectors or themes. Accordingly, few Sector Strategy Papers (SSPs) produced since 1997 have incorporated evaluative work to inform future directions (ARDE 2000); none provided a baseline against which to measure progress in implementing the strategy; and only one of the nine SSPs reviewed by OPCS discussed monitoring responsibilities. With the introduction of the MDGs, the Networks have begun to work with DEC to help define more detailed indicators relevant to their domain. These efforts have been reinforced following the Strategic Forum in January 2002, which placed a high priority on monitoring the Bank’s results on development. A special unit has been set up under the control of the Bank’s senior management to focus on putting a workable system in place, and efforts have been made to identify—together with the Network Anchors—pilot countries and indicators for developing baseline data against which Bank efforts and results can be measured.

It is still too early to assess the results of this initiative. A corporate effort to develop sectoral performance indicators was launched in the early 1990s, but soon fizzled out. The proposals were too complex and the resources allocated for outreach, capacity building, and training were inadequate. Except for information required for tracking the MDGs, most Network Anchors do not yet have easily retrievable data on inputs, outputs, and outcomes in their respective sectors, subsectors, and thematic areas. They do not track quantitative outputs or outcomes of Bank assistance (except for noting percent of outcomes rated at risk or unsatisfactory by QAG and OED) and are far from being in a position to attribute impact on indicators to Bank assistance. The constraints to monitoring and evaluating sectoral and thematic assistance are due not only to the

current system of sector coding, now being revised (Chapter 6), but also to an absence of clear monitoring methods, resources, and accountability.

Country-Level Evaluation

Since the early 1990s, the country has become the focal point of the Bank's strategy formulation. In parallel, the importance of country-level evaluations has increased as well. The introduction of Country Assistance Strategies (CASs) in FY91 was further reinforced by increased decentralization of Bank operations in FY97. But for most of the 1990s, CAS documents were mainly forward looking, and largely silent on evaluating the outcomes of past Bank assistance. In the last three years, steps have been taken to improve evaluation at the country level both within the Bank and in client countries. Aside from nine pilot countries that were selected in FY98 for more intensive CAS self-evaluation, a FY00 CAS retrospective found that only 16 percent had satisfactory self-evaluation. Starting in 1995, OED has carried out some 55 independent evaluations of assistance using the country as the unit of account (Chapter 7 has results of tracer studies of these evaluations). On this basis, it has defined a methodology that, once adopted by Regional managers, should enable country units to carry out systematic self-evaluation, with QAG to check on their quality.²

The impetus to improve the Bank's self-evaluation has been mirrored by attempts to improve monitoring and evaluation in client countries. The Task Force on Evaluation Capacity Development (ECD) recommended in 1994 that ECD become an integral part of the Bank's public sector management strategy in client countries. OED has long sought to help countries improve their capacity for M&E, although until 1997, when it became an integral part of OED's new strategy (see Chapter 7), the efforts were relatively modest and ad hoc. Since 2000 OPCS has provided operational support to the Regions for ECD, focused mainly on seven pilot countries in the M&E Improvement Program. Country-level ECD activities are now under way in about 20 countries. However, many challenges remain (box 5.1).

Other initiatives may reinforce ECD in client countries. In FY97, the Heavily Indebted Poor Countries (HIPC) program was launched, which requires countries to monitor poverty-related expenditures. Few have been able to do so. In FY99, the CDF was piloted in 13 countries, which focused on long-term, participatory development; one of its four main pillars was monitoring results. The PRSP that was introduced in International Development Association (IDA) countries in FY00 also focuses on the identification of monitorable indicators. OED, in its recent review of IDA, argued that all IDA clients should be required to set monitorable

Box 5.1

Strengthening Evaluation Capacity in the Bank and Borrower Countries

OED's first Annual Report on Evaluation Capacity (AREC) Development in the Bank (June 2002) notes that in spite of dramatically increased activity in the past five years throughout the Bank to strengthen M&E both within the Bank and in client countries, a number of challenges remain. These include:

- *The need to clarify definitions of monitoring and evaluation.* Confusion still exists between M&E and monitoring financial spending and national development indicators (monitoring focuses on inputs and outputs, and outcomes, and evaluation establishes causality and attribution and assesses efficiency for the outcomes).
- *Improving the consistency of Bank strategies, policies, and guidelines on country-based M&E.* In the context of PRSPs, the guidelines recommend assessing countries' M&E systems, but do not have a formal instrument for doing this (similar to CFAAs, CPARs) and guidelines on PRSCs do not include any mention of strengthening a county's M&E system as a tool for improving governance and accountability.
- *Creating a fiduciary responsibility* to strengthen both Bank and client country M&E capacities.
- *Better coordination* within the Bank of both ECD efforts and training within the Bank for M&E, through building a data base, forming a thematic group, developing Regional approaches.

performance targets and to produce publicly disclosed reports on performance (OED 2002c, p. 43), and a new initiative to improve results based monitoring of IDA's development impact is under consideration.³ All of these initiatives have highlighted the importance of monitoring and evaluation for accountability and improved governance.

Monitoring and Evaluation in the World Bank Institute (WBI)

Ten years ago, the WBI, then called EDI, monitored its training activities through the use of surveys administered on a voluntary basis by the instructor to trainees at the end of the course; the results were available only to the instructor. In addition, external evaluations were carried out by consultants. OED reviewed WBI's evaluation system in 1996 and recommended that the training surveys become systematic; that the results be reported and used to analyze cross-cutting issues on core courses; that evaluation be expanded to cover ex post and impacts, based on measurable results; that WBI become proactive in enhancing evaluation capacity in partner institutions; and that WBI should maintain close professional links with OED.

In 1997, WBI instituted a more formal system of monitoring and evaluation, which continues to be strengthened today. Four levels of M&E were identified. The first, which is mainly monitoring, is feedback from trainees, now required for all WBI-sponsored training activities and reported quarterly to senior management. The second level

is focused on performance-testing participants before and after the courses; results of Level 2 on WBI core courses in FY99–01 are in box 5.2. This year Level 2 has been enhanced by focusing on the quality of testing and ensuring that all 15 core programs are monitored by testing in at least 2 activities in each core program.

The third and fourth levels of evaluation, outcome and impact monitoring, were launched for the first time in FY02 for five of WBI's thematic programs aimed both at participants in client countries and staff. These evaluations measure changes in behavior of participants and changes in their organizations or institutions. WBI has also continued to do impact evaluations of large activities such as Social Policy Reform in Transition Economies (SPRITE) and the World Links for Development. In the future, the WBI evaluation unit intends to focus on evaluation of client training and capacity enhancement at a country level.

Looking ahead, WBI should further strengthen monitoring and evaluation of training activities. Level 1 monitoring could be better exploited by analyzing results for patterns (by program, by Region, by type of course—distance learning versus face to face). The coverage of Level 2 evaluations should be extended to illuminate differences across types of training and core activities; the results could feed back into strategic decisions on future training. The planned country focus of evaluation of WBI activities should be undertaken, to the extent feasible, in conjunction with OED's country evaluations to take advantage of synergies.

Box 5.2

Learning Achievement of WBI Course Participants

Over the FY99–01 period, WBI delivered 237 core course offerings. On some 60 percent of these, both Level 1 and Level 2 evaluations were carried out. The results from Level 1 show that 85 percent of respondents found the courses useful, with Regional differences (AFR, LAC, and EAP had higher ratings). Level 2 evaluation showed a significant difference in pretest and posttest average scores. Interviews with Regional staff on these course

offerings indicated a need to improve collaboration with WBI. Because the selection of core course offerings that were evaluated was not evenly distributed across core courses or offerings, the positive results of these evaluations are not necessarily representative of all WBI core courses. This is being addressed in FY02.

Source: WBI 2001.

Monitoring and Evaluating the Bank's Research

With the drive to create the Knowledge Bank, the Bank's research has become more central. Assuring its relevance, quality, dissemination, efficiency, and impact has become critical. The majority of the Bank's research is carried out by the Development Economics Vice Presidency (DEC).

DEC's Research Advisory Staff (DECRA) have conducted a variety of assessments over the past decade to evaluate Bank research, as summarized in box 5.3. Nevertheless, some gaps remain. Research funded by the Research Support Budget (RSB) is evaluated in a number of dimensions (relevance, quality, dissemination, impact), but research projects not receiving RSB funds are not subject to the same level of ex ante or ex post review. The bulk of this non-RSB-funded research

is carried out by DEC, in particular by the Research Group (DECRG). Past efforts to evaluate this research have not been systematically followed through. Regular ex post evaluations should be done for the non-RSB-funded research, and the impact of this research should be assessed through regular tracer studies. For transparency, results of the ex post reviews and tracer studies should be reported to the Board on an annual or biennial basis.

Corporate-Level Performance Measurement

Until 1997, little attempt was made to measure the impact of Bank assistance at the corporate level. The implicit "scorecard" focused on the amount of lending and, to a much lesser extent, on the outcomes of individual lending operations. That paradigm began to shift in 1997, with the Strategic

Box 5.3

Ex Ante and Ex Post Evaluation of Bank Research

Oversight of the Bank's research rests with the Bank's Chief Economist, with the advice of the Bankwide Research Committee, whose mandate is to ensure the quality of Bank research and its relevance to the Bank's mission. The Research Committee is comprised of Regional Chief Economists and managers in the Networks, IFC, WBI, and OED. It is assisted in its mandate by a Secretariat, housed in the Research Advisory staff (DECRA).

For the last five years, between \$21 and \$24 million has been spent annually on research in the Bank, representing 2 to 3 percent of the operational budget. Of this, 25–30 percent is funded by the Research Support Budget (RSB), which is allocated year-round based on research proposals. One of the primary responsibilities of the Research Committee is advising on the allocation of the RSB.

For RSB-funded research, there is a thorough ex ante assessment of each project proposal by external and internal reviewers. Ex post evaluation for projects costing more than \$20,000 consists of reviews by outside experts that focus on objectives, design and implementation, dissemination, and results and cost-effectiveness. Within these categories, reviewers are asked to focus on relevance (are the topics critical for policy to client countries) and impact (could the results have influenced Bank programs or country policy). These evaluations

are produced about every two years; one is under preparation for 2002. The projects are not rated, however, and although the merits of doing so have been extensively discussed and decided against, rating would enable aggregation and tracking over time.

DECRA has evaluated the dissemination and impact of research activities throughout the Bank, both RSB- and non-RSB-funded. Dissemination has been assessed through citation counts of Bank research output and Bank journals, readers' surveys of the Policy Research Bulletin, and surveys of Bank operational staff and policymakers. Evaluation of the impact of Bank research is difficult. DECRA efforts have consisted of assessing the influence of Bank research on those outside the Bank, using both quantitative indicators and qualitative approaches, including case studies, and surveying and interviewing Bank staff for internal impact on specific research projects (tracer study), on the use of researchers for cross support, and on general research service to Bank operations. DECRA reports every other year on results of these evaluations, each report having a special focus. For the 2002 report, DECRA surveyed regional research and educational institutions on their development and programs and on capacity building activities supported by the World Bank.

Compact, which proposed the introduction of a table of indicators, which in turn became the Corporate Scorecard. The Scorecard was intended to monitor progress toward goals and development outcomes at the corporate level and was divided into three tiers: (i) final outcome indicators and indicators of country performance, including the international development goals (precursors of the MDGs); (ii) Bank performance, at country and sector levels, summarizing the quality of inputs and subsequent delivery in meeting objectives; and (iii) internal bank measures. The Strategic Compact Assessment found progress in filling in Tiers III and I, but Tier II remained largely empty (World Bank 2001b, Annex 9, p. 16). In addition, the links among the three tiers were not well defined. Measuring the impact of World Bank assistance is yet to be achieved and there is now some effort under way to recast the Scorecard to enable it to reflect Bank performance.

The authorizing environment for performance measurement is shifting. It is becoming a priority at the highest levels of both the Bank and the international community. Initiatives under HIPC since 1997 and the introduction of the CDF and PRSPs more recently have focused on monitoring inputs and outcomes at the country level. Experience with these initiatives has highlighted the difficulties. The Strategic Forum in January 2002 gave new impetus to measuring results, and the need to strengthen this aspect of aid management was a major part of the development consensus reached at the Monterrey Conference on Financing for Development in March 2002.

A new unit was established in February 2002, headed by a Vice President, tasked with putting in place a system for measuring the Bank's development effectiveness. It has focused on carrying out an inventory of ongoing initiatives in performance monitoring; proposed piloting outcome-based CASs and outcome-based Sector Strategy Papers; and linked these efforts to implications of outcome-based approach for lending and nonlending activities. The unit is consulting broadly within and outside the Bank, and in June 2002 co-hosted with multilateral development banks (MDBs) an international roundtable with representatives of the sponsoring MDBs, bilateral donors, other multilateral agencies, clients, and academia on measuring, monitoring, and managing for results.

These changes in the authorizing environment should help the paradigm shift toward the centrality of performance measurement, management, and evaluation for accountability and better governance. These changes present new challenges in methodology, in evaluative skills, and in the distribution of accountability for evaluation. Network Anchors and Regions will need to carry out more self-assessment, incorporating MDGs or MDG proxies into the M&E framework, and central units will have to rely on the Regions and Network Anchors to take on these responsibilities. However, if the efforts in this direction can be sustained, they should result in major progress toward better country, sector, and corporate monitoring and evaluation in the medium term.

6

INFORMATION AND COMMUNICATION

In terms of both information and communication, the Bank is a vastly different institution than it was ten years ago. Information systems have been integrated and shifted to electronic form; new sector and thematic codes have been introduced. Communications have been expanded and modernized; staff training has doubled; and the revised disclosure policy allows the Bank to provide better information to partners. All of these changes, although not without problems, have helped to strengthen the control framework and shift the internal culture of the Bank toward better management.

Ten years ago, information systems were piecemeal and fragmented. Data on projects in the pipeline, in the portfolio, on disbursements, on ESW, on grants, on client countries were stored in some 60 different data systems. Communication was largely paper-based, filing systems were cumbersome. They provided good institutional memory, but inboxes overflowed. New initiatives filtered slowly and unevenly through the institution.

Information

All of this changed dramatically over the past decade. Both SAP and the Business Warehouse were put in place to integrate the many sources of information and to permit, in principle, efficient tracking information relevant to Bank activities and client countries. The automated Dashboard, successor to Compass, is used by business units, particularly in the Regions, as a rough and ready monitoring tool. The use of e-mail and the Intranet

has grown so that all Bank staff are now literally on the same page when they turn on their computers. An Internal Communications Unit was reestablished in 2001.

These changes have—after a period of disruption—made reporting more timely and detailed, and information on Bank processes more readily available across the Bank. Corporate priorities and strategic directions are communicated to staff. Monthly management reports and quarterly business reports provide information on the year's deliverables, budget use, and quality of the portfolio to Regional and Network management teams. These improvements facilitate controls and more consistent risk management. They do not yet capture significant aspects of Bank activity, however. Information on partnerships, for example, is not systematically captured by any system, and the Business Warehouse has no information on informal ESW or other AAA. The

reliance on e-mail has eroded institutional memory, especially about the process and internal debates over preparing strategies and projects.¹

New Sector and Thematic Codes

A constraint to reliable reporting on Bank assistance across sectors and thematic areas is the current system of coding. Key initiatives in thematic areas such as private sector development, environment, and gender are incompletely captured. In addition, the current system is unable to reflect the multisectoral nature of many operations. In September 2001, the Bank began work on a new system that reflects the multisectoral and thematic nature of Bank work and better reflects the corporate priorities. The new coding system will in principle enable the Bank to have better information on its assistance, both lending and nonlending. The introduction of the new system has not been smooth, however. Some Sector Boards are concerned that the rapid recoding of the stock of operations and ESW resulted in a number of mistakes that would need correction before the system was made operational. The new system is more complicated than the former one (operations can be coded in more than one sector and one theme; there are some 57 subsectors and 68 subthemes). It will be vital for staff to receive adequate training and for quality of data entry to be assured.

Communication and Disclosure

In addition to the explosion in the use of the Intranet and e-mail (connecting a decentralized and matrix-managed institution), staff training has expanded. The Strategic Compact goal of doubling resources for training staff as a proportion of administrative costs was largely met, led by Network deliveries.² Last year's AROE recommended that

a comprehensive evaluation framework be developed in conjunction with the New Learning Framework proposed in FY01, and this recommendation remains valid.

DEC and WBI have an important role to play in helping management to elaborate the Bank's new strategic directions and the Regions and Networks to adapt their processes to the MDGs. It is particularly important to communicate both within the Bank and externally that attainment of these goals is dependent not only on the amount of aid but also on global and country policies. DEC and WBI can go further in engaging in an open debate about all that is not known about the linkages between governance, growth, good policies, and poverty alleviation and encouraging further debate and research in these areas.

The Bank has revised its disclosure policy for the first time since 1993. The new policy provides for public release of a much wider range of documents, including adjustment lending documents after Board approval; ICRs; QAG synthesis reports; and a wider range of OED evaluations (Performance Assessment Reports; process evaluations, including the AROE). This will promote greater transparency. Past changes in the Bank's disclosure policy required additional efforts by the Bank to comply with the changes. Progress reports in 1995 and 1997 reported on the implementation of the policy, noting areas of progress and areas where further improvements were necessary, including making environmental assessment reports available in a timely way. The 1999/2000 review of the disclosure policy found that disclosure had further improved, including in the timely release of environmental assessment reports. The greater demands on disclosure will call for further efforts. Both the policy and its implementation should be monitored and periodically adjusted in the future.

7

OED'S AGENDA

This chapter reviews OED's work program and evaluations carried out on its products and services over the past year. OED has continued to emphasize priority areas established in its 1997 renewal strategy: evaluations at the country, sector, and thematic levels and resources devoted to evaluation capacity development and staff training. It needs to broaden its evidentiary base for evaluation; strengthen and disseminate its methodologies, particularly for country evaluations; and ensure better dissemination of its products and services.

For the first part of the 1990s, OED continued to do project-level evaluation on the same basis as it had in the 1980s. It reviewed ICRs (and their predecessors), carried out Project Performance Audits on a portion (40 percent) of closed projects and Impact Evaluations on a smaller proportion, evaluated broader themes and topics, and prepared annual reports on project outcomes and on monitoring and evaluation in the Bank. In 1997, OED adopted a renewal strategy, which refocused its priorities and remains valid today (see box 7.1). As reported in last year's AROE and reflected by this year's AREC, OED continues to give priority to the main objectives established in its strategy, particularly timely country, sector, and thematic evaluations; evaluation capacity development; staff training in M&E; and timely dissemination.

Consistent with its role as independent evaluator of Bank products and services, OED un-

dertakes each year to evaluate the quality, relevance, and efficacy of its own products and services. This year's AROE reports on: progress on recommendations in last year's AROE; surveys of Bank and Executive Directors' staff on OED's recent products and services; and a tracer study on the impact of Country Assistance Evaluations.

OED's Outputs: The View from the Committee on Development Effectiveness (CODE)

OED reports directly to the Board of Directors and CODE. It provides them with indicators of its inputs, deliverables, and timeliness (found here in Annex 3, for the past three fiscal years through the third quarter of FY02). Toward the end of each fiscal year, CODE discusses future directions for OED and outlines areas for OED's attention. At the April 2002 meeting, CODE endorsed OED's proposed work program, agreed

Box 7.1**OED's 1997 Renewal Strategy****Mission:**

To contribute to development effectiveness through excellence and independence in evaluation.

Objectives:

1. Move to a higher evaluation plane, from projects to country, sector, thematic, and global evaluations.
2. Shorten the feedback loop to ensure relevance.
3. Build evaluation capacity within and outside the Bank to promote self-evaluation.
4. Invest in knowledge and partnerships to maintain OED's intellectual leadership and develop alliances as instruments of strategy.
5. Manage for results: evaluations add value if they influence behavior and contribute to the overall success of the Bank's strategy.

with the areas of emphasis, and requested that studies on education and pension reform be advanced. The OED work program has been adjusted to reflect Board comments.

Recommendations from Previous AROEs

Last year's AROE highlighted four areas for attention, consistent with CODE's guidance. Progress in these areas is described below.

Timeliness and relevance of OED's products have been strong. The majority of products produced or under way in FY02 were linked to the preparation of Sector Strategy Papers (e.g., Rural Poverty), to the revision of a Bank policy (e.g., Indigenous Peoples), or to an external review (e.g., Extractive Industries). CAEs continue to be linked to CAS preparation. The remaining major products were carried out on request of the Board (such as Social Funds, Global Programs, HIPC, CDF). Last year's AROE also suggested that OED focus on corporate priorities, particularly the poverty orientation of Bank operations. CAEs have had stronger focus than in the past on poverty reduction, and most major sector and thematic studies have as well (Rural Poverty, Social Funds, Indigenous People, and the Urban Review now under way, for example).

Methodology assessment has been carried out and development and dissemination have seen less progress than is desirable, as recent surveys have shown. For evaluation of adjustment lending, there has been a pilot effort to group assessments of adjustment lending in countries and to focus on country context. OED is simplifying its project evaluation methodology and at the same time

enhancing the coverage of safeguards through changes ongoing in the Project Information Form. At the country level, OED has continued to refine the Country Information Form, as recommended last year, but Regional efforts to test it as a self-evaluation tool for country strategies have lagged.

Participation and communication have been enhanced, and this year's survey results point to the need for further improvements. Evaluation processes have been more participatory and involved broader consultations (e.g., Social Funds surveyed households; the CDF is a joint evaluation with other donors and involves extensive in-country consultations). Advisory Committees of external and Bank staff are being used (e.g., Russia CAE, Global Programs) and entry workshops have become a regular feature of larger evaluations (Transition Study, China CAE). Internal mid-point workshops (Russia CAE) have also been used to get early feedback on OED findings.

Dissemination has improved, but needs continued effort. OED is putting its reports on the Web more quickly than before and has begun providing hot links to related products, but as survey results show, there is further room for improvement in disseminating sector and thematic studies.

OED Outcomes and Impact**Client Surveys: Results**

Self-evaluation has followed last year's recommendations by continuing the annual survey of its products and services. OED has also carried out

a tracer study of CAEs. This section presents the results.

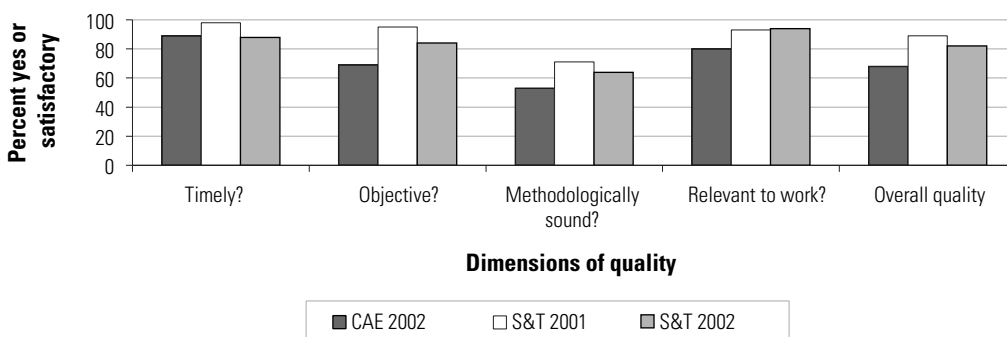
OED carried out four surveys: (i) for advisors and assistants to Executive Directors (ED); (ii) for thematic and sector groups related to four recent OED sector and thematic (S&T) studies; (iii) country teams on OED's CAEs; and (iv) all higher-level staff in operational or research units, asking their views of OED's products and services in general. In total, more than 3,200 staff were surveyed. The response rates ranged from 19 to 42 percent and averaged 25 percent across all surveys, up from 22 percent in last year's surveys. Annex 5 shows detailed responses by survey. Following are the salient messages for OED going forward.

Readership, usefulness, and dissemination. Readership is fairly strong, with 68 percent of respondents to the general survey having read at least one OED report in the past six months and respondents to the ED staff survey reading many OED reports on a regular basis. A high proportion of respondents to these surveys who had read OED reports found them useful to their work. Better dissemination of selected studies is called for, however, as only 35 percent of responding staff had read the recent sector or thematic evaluation relevant to their work. Of the more than 500 respondents to the general survey, over half had used at least one OED service (help desk, Web-based docu-

ments, project ratings) in the last six months, and of these, more than 90 percent thought the services were relevant to their work.

Methodology. Four dimensions of quality were rated on a four-point scale (excellent, good, fair, poor). Figure 7.1 shows the proportion of respondents from each survey that rated each dimension as good or better. Respondents on the S&T evaluations give good marks on most aspects of quality, although one area for improvement is methodology. For CAEs, methodology also received relatively lower ratings from respondents. Almost a third of the respondents rate overall quality as fair or poor, 20 percent think the CAEs have little or no relevance, and only a little over half the respondents think that the CAE is objective and is methodologically sound, although respondents on CAEs carried out prior to FY01 give significantly better ratings on quality than the respondents on CAEs done in FY01 or later (Annex 5). In addition, however, 93 percent of respondents said that if they were to start work on a new country, they would find a CAE to be useful. Even if respondents found the particular CAE on which they were responding had weaknesses, they thought the CAE product itself was a useful one. Finally, respondents were asked to make suggestions for improvement. The two most frequent suggestions were to broaden consultations within the Bank and with external partners.

Figure 7.1 Quality of OED Evaluations



A clear challenge for OED is to continue to strengthen and disseminate its current methodology, particularly in CAEs, and to ensure that both the sources and limitations of the evidence underpinning the evaluations are based are presented explicitly in the reports. The scope for improving evaluation methodology is constrained by a lack of knowledge about the ingredients necessary for rapid growth and poverty reduction and a lack of consensus on the role of the Bank in differing country circumstances (e.g., with respect to conditionality or to social development). In addition, there are different views within the Bank on whether and how to judge performance in relation to objectives.

Influence of OED evaluations on Bank policies and assistance. Respondents were asked to rate on a four-point scale (strongly, somewhat, little, not at all) the extent to which OED evaluations influenced their own views of development priorities in the country or sector and the extent to which the evaluations influenced the Bank's strategy or assistance (see table 7.1). A fairly high proportion of respondents this year think that the evaluations influenced their own thinking and Bank strategy; a somewhat lower proportion think that OED has had an influence on lending and nonlending as-

sistance. This is consistent with the findings from the CAE tracer study discussed below.

Results of the CAE Tracer Study

CAEs have become a major OED product and remain the only consistently produced evaluation at a country level in the Bank. To date some 55 CAEs have been completed. The process has been both challenging and controversial. As a result, OED has worked internally to strengthen, clarify, and disseminate its methodology and has surveyed Bank staff for feedback.

In addition to measuring staff perceptions on CAEs, OED carried out a tracer study this year to assess the extent to which CAE recommendations on a country were incorporated into the subsequent CAS and were reflected in the assistance actually delivered to the client country. The study reviewed 15 CAEs that were completed in FY99 or earlier, to ensure that enough time had passed to permit a review of Bank assistance actually provided.¹ Although determining causality is difficult, the degree of consistency between CAE recommendations, the CAS, and actual assistance is a partial indicator of the impact of CAEs.

Consistency between CAE recommendations and the CAS was strong. In most of the countries re-

Table 7.1

OED Influence (percent somewhat or strongly)

Question	Survey							
	ED staff		General		CAE		S&T	
	2001	2002	2001	2002	2001	2002	2001	2002
Influence your view of strategy and development policy:								
In the sector?	90	90	54	76	—	—	71	75
In the country?	84	85	—	—	—	66	45	—
In the Bank as a whole?	79	88	—	—	—	—	56	—
Influence implementation of projects or programs?	—	67	—	—	—	—	—	—
Influence the Bank's:								
Sector/country strategy?	—	—	—	—	—	71	—	83
Lending services?	—	—	74	—	—	53	—	50
Nonlending services?	—	—	64	—	—	53	—	64

Note: — indicates the question was not asked for that particular survey.

viewed, there was good consistency between the overarching recommendations in the CAE and the subsequent CAS. This was the case in Albania, Indonesia, Kenya, Malawi, Morocco, Philippines, and Yemen and reinforces the message from the surveys, in which 71 percent of respondents thought that CAEs had influenced Bank policies at a country level. In Mozambique, the 1997 CAE came too late to influence the 1997 CAS, although the subsequent 2000 CAS was consistent with the broad CAE recommendations, less so with the sector-specific ones. In Sri Lanka, there has not yet been a full CAS because of the political uncertainty in the country.

CAE recommendations not found in the subsequent CAS include: (i) establishing monitoring indicators, which was considered by country teams to be either unnecessary (the Albania CAE had recommended early warning monitors of crises) or too vague to be useful; (ii) pace of reform, where CAE recommendations were considered unrealistic, such as for privatization in Bangladesh and Ukraine; or the recommendation was untimely, such as addressing devolution in Sri Lanka, where it first had to be pursued in the context of a political solution to the conflict; and (iii) specific lending or ESW activity, as in Sri Lanka where the recommendation was considered to have low relevance.

Consistency between CAE recommendations and implementation was somewhat weaker. Only about half of the number of recommendations tracked across all CAEs were fully or substantially

implemented. The reasons for weak—or a lack of—consistency (box 7.2) include: (i) limited capacity or opportunity in countries, including Bangladesh, Malawi, and Sri Lanka; (ii) pressures to lend, as in Kenya and Morocco;² and (iii) internal Bank incentives or expectations from government and other donors that make selectivity difficult, as in Mozambique. One implication of these findings is that CAE recommendations need to focus on what can realistically be accomplished in the next CAS period, typically two or three years, and which recommendations are for the longer term.

Recommendations for OED

In light of these findings, OED proposes to focus on the following areas:

Dissemination

- Survey results indicate that last year's recommendation about more purposeful targeting of Bank audiences for its products is still valid. OED should engage the Networks in identifying how OED products can most effectively reach interested staff. A wider range of dissemination activities should be used, including live presentations and workshops in addition to paper and Web-based document distribution.
- Each major OED study should state early in its design who the study's target audience is (in addition to CODE and the Board) and how it will be disseminated.

Box 7.2

CAE Recommendations and Implementation of Bank Assistance: A Varied Picture

Implementation of country strategies in Indonesia, the Philippines, and Yemen has been substantially consistent with CAE recommendations. Sometimes the consistency has come about with a lag, as in Albania and Mozambique, where efforts to monitor poverty reduction and improve aid coordination were stepped up some years after the CAE as a result of the PRSP process. In Bangladesh, IDA assistance was limited because of government corruption and low commitment to reform, which was consistent with CAE recommendations, but it meant that other specific recommendations (on state-owned enterprise

[SOE] reform, flood control) could not be pursued. A similar situation existed in Ukraine. In Kenya the Bank moved from a no-lending scenario recommended by OED in FY98 and proposed in the FY99 CAS a program of emergency and adjustment lending. In Morocco, contrary to the FY97 CAE, which proposed greater selectivity in adjustment lending and a focus on fiscal issues, and the FY97 CAS, which proposed one small adjustment loan for the base case and also focused on fiscal issues, over half of the Bank's assistance was for adjustment lending, with fiscal issues still a major concern.

Methodology

OED should:

- Include an annex in each evaluation report stating the methods and data sources used, including their strengths and limitations.
- Evaluate and report on compliance with safeguard policies in each Project Performance Assessment Review.
- Begin to align its evaluation of risk and sustainability to the Bank's overall risk management framework.
- Continue to support the introduction of self-evaluation methods by management for country programs.
- Continue the internal validation and calibration of evaluation standards started in FY02, and use the results to design staff training and guidance to further improve the consistency of treatment.

Participation and Consultation

- Because the staff survey results point to a need for greater consultation and participation with both Bank staff and external partners, OED will need to pay particular attention at early stages of evaluations to ensure that plans are adequate for engaging partners, including better use of technology (video-conferencing).
- In cases of unresolved disagreement between OED and management on individual project evaluations, OED should institute a transparent procedure for discussing reasons for differences.

Self-evaluation

- OED should continue regularly surveying and reporting on the usefulness of OED products to staff and Executive Directors, using the same survey instruments as this year.

8

CONCLUSIONS AND RECOMMENDATIONS

The last ten years have witnessed profound changes in the way the Bank defines its mandate, acknowledges risks to fulfilling that mandate, and is initiating efforts to measure performance in reaching its goals. At the same time, the drive to become a Knowledge Bank has engendered new initiatives and new processes, for which both the control environment and the evaluation framework have yet to be well defined. In 1997, the Strategic Compact highlighted the importance of measuring results and the introduction of the Corporate Scorecard pointed to a strong focus on country, sector, and thematic evaluations.

At that time, OED endorsed these initiatives and recommended moving to a results-based management (RBM) system, where evaluation findings would be linked to resource management. Limited progress has been made on filling in the Scorecard, and still less on moving to RBM. The recent shift in the authorizing environment on the centrality of evaluation for accountability and learning, however, may lead to more rapid progress in both areas.

Main Findings

Control environment. This review of the evolution of the control framework in the Bank over the past decade has found that new structures and functions have been put in place and an altered internal culture has taken root with respect to risk management. One issue that has emerged, however, is increasing risk aversion on the part of both Bank managers and borrowers, who seek to avoid

lending operations subject to heavier controls. This points to the need to scale up risk assessment and seek mitigation measures at a country level.

Risk assessment has been strengthened at the project level through increased fiduciary assessments and safeguard compliance requirements. It is stronger for projects in the pipeline than for projects under implementation, and stronger for investment lending than structural adjustment lending.¹ Monitoring and evaluation need to be strengthened at the project level to focus on results, facilitate learning, and improve the feedback loop to stronger risk assessments for the next generation of assistance. *Risk assessments* for grants, trust funds, and partnerships, by contrast, have seen only recent improvements, and it is still early to assess the effectiveness of the new controls.

Risk assessment at the country level has also been expanded in the last decade to include new and stronger fiduciary reviews. These efforts could

be strengthened if they were combined into a coherent risk assessment framework and reinforced with systematic evaluation of country-level results that could inform future risk assessments. *Risk assessment* at the corporate level is still in the process of being set up; oversight at this level will rest with the Management Committee, but its mandate, priorities, or functions have yet to be defined.

Control activities depend on unambiguous and relevant guidelines, which the Bank has been putting in place. In addition to the information provided to the Board and senior management on the status of conversions currently under way, a comprehensive timetable for policy conversions and updating should be regularly provided to the Board.

Control activities for projects were enhanced by the creation of QAG, whose criteria and methodology have now been adapted for quality assurance by the Regions. Control activities continue to expand to areas of vulnerability, including safeguard compliance and trust funds, but it is too early to judge their effectiveness. Monitoring of project outcomes remains weak, however, in spite of repeated intentions over the years to strengthen it.

Monitoring inputs, outputs, and outcomes and *evaluating* the results of Bank assistance remain a challenge. The recent focus at the highest levels of the international development community on demonstrating results has energized the Bank to strengthen its self-evaluation and monitoring system. Although progress has been made, important gaps remain. More systematic monitoring and evaluation are needed for ESW, grants, and partnerships. At the country, sector, and thematic levels, self-evaluation is largely absent.²

Information and communication improvements have made the Bank a different place from a decade ago and have enabled rapid dissemination of guidelines and best practice; more effective use of control mechanisms; and more timely risk assessments, particularly for the portfolio. The new sector and thematic codes will, in principle, permit better tracking of Bank assistance in line with corporate priorities. DEC and WBI have an important role to play in ensuring that the message of the complexities involved in attaining the MDGs includes the quality of borrowers' governance and global policies.

For *OED's agenda*, the messages from CODE, internal surveys, and a tracer study point to continued challenges to improve and disseminate more widely its evaluation methodologies for country evaluations; increase transparency of the evidentiary basis of evaluations; target better the dissemination of sector and thematic evaluations; and enhance OED's monitoring of the Bank's adoption of OED recommendations.

Recommendations

Results-based management will be realized only when self-evaluation is strengthened at all levels, independent evaluation contributes to improving methods and indicators, and reliable data are generated by countries. The agenda for improved monitoring and evaluation includes providing guidance on the overarching evaluation framework and filling gaps in the tracking systems of lending and nonlending services, grants, and partnerships. More transparent CAS and SSP processes should be complemented by rigorous self-evaluation. Against this background, the AROE recommends that actions be taken on four fronts, three by management and one by OED:

- *At the corporate level*, senior management should design performance measures to track Bank achievements at project, country, and global levels in relation to the Millennium Development Goals. It should implement an integrated risk management framework to facilitate the administration of financial, operational, developmental, and reputational risks faced by the Bank. In this context, it should make explicit the strategic risks and rewards of operational policies and corporate strategies.
- *Strengthen the evaluation and control framework for sector strategies, partnerships, and grants.* The operational shift to the higher plane of policy and sector strategies should be backed by adequate network accountability for sector performance measurement and management of instruments and services. Management and OED should develop a concise operational policy that provides an overarching evaluation framework. Project-level monitoring should be treated as a fiduciary responsibility to ensure that adequate monitoring is in place be-

fore Board approval. The “evaluability” of grant-financed activities should be improved. For all partnerships, clearer distinctions in accountability and more accurate specification of quality assurance, evaluation, and oversight should help to minimize conflicts of interest and increase effectiveness.

- *At the country level, risk assessment and learning should be improved* by developing a comprehensive approach that incorporates the results of fiduciary assessments (such as the CFAAs, CPARs) into the CAS. Self-evaluation at the country level, based on verifiable indicators grounded in country processes, needs to be integrated into country programs.
- *Finally, OED* will need to improve and disseminate more widely its evaluation methodologies for country evaluations and increase the transparency of the evidentiary basis of its evaluations. It should help Regions and Network Anchors incorporate self-evaluation in country assistance and sector strategies, which would improve the ability of the Bank to measure its performance without compromising the independence of OED’s evaluations. It should address the independent evaluation needs of global and regional programs; better monitor and evaluate compliance with safeguards; target more effectively the dissemination of sector and thematic evaluations, and use a wider range of dissemination tools.

ANNEX 1: OED REVIEW OF ICRS

The Implementation Completion Report (ICR) is a self-evaluation instrument required for each closed lending operation and produced by the Regions. An ICR marks the transition from implementation to exit from the Bank's active loan portfolio. OED conducts an independent review of each ICR, called an Evaluation Summary (ES), to validate or adjust the ratings based on the information provided in the ICR and other evidence.

This annex is divided into three sections. The first provides a summary of ICR evaluations conducted by OED between FY96 and FY01. The second describes the major findings on trends in ICR quality for FY96–01 exits. The third shows changes in ratings at different stages of the project cycle for FY96–01 exits.

OED Evaluations, FY96–01

Table A1.1 summarizes OED's output of ICR reviews and Project Performance Assessments reports (PPARs) since FY96. Due to bunching of ICR deliveries toward the end of the fiscal year, the number of ICRs evaluated in some years can exceed the number of ICRs received, as in FY98.¹

Findings on ICR Quality

OED ratings provide an independent assessment of ICR quality and their consistency with ICR guidelines. Three dimensions of ICRs are assessed: (i) the analysis provided, which includes coverage of important subjects, soundness of analysis, adequate and convincing evidence, and adequacy of lessons learned; (ii) the future operation of the proj-

Table A1.1

OED Evaluations of ICRs, FY96–01

	FY96	FY97	FY98	FY99	FY00	FY01
ICRs received	268	397	218	291	307	317
Adjustment	43	65	24	45	36	42
Investment	224	332	186	236	251	275
ICRs evaluated	249	313	284	268	270	286
Adjustment	44	45	38	36	50	27
Investment	204	268	246	232	220	259
PPARs	100	79	71	69	72	73
Adjustment	17	13	13	14	14	14
Investment	83	66	58	55	58	59
Assessment ratio	40	25	25	26	27	26

Note: Figures in table represent number of reports received and evaluated by OED in each FY. 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.

ect, including a plan for future project operation, the inclusion of performance indicators for the project's operational phase, and plans for monitoring and evaluation; and (iii) borrower and cofinancier inputs. The evolution in the quality of ICRs by dimension is in table A1.2.

Quality of ICR analysis is high. Almost all ICRs covered the important topics, were internally consistent, and drew appropriate lessons. The quality of ex post economic analysis for those operations where it was relevant, at 85 percent satisfactory or better, remains one of the weaker aspects of the ICR analysis, although it has shown marked improvement over the last six years. The extent to which the evidence presented in ICRs is complete and convincing remains the weakest aspect, where in spite of some improvement several years ago, it remains at essentially the same level as in FY96 at

83 percent satisfactory or better. This reflects the weak quality of monitoring and evaluation, which makes it more challenging to pre-sent convincing evidence of results. Although the poverty analysis component of ICRs remains weak, at only 65 percent satisfactory, this quality dimension was added only in 1997 and has been rated only where relevant, for a relatively small number of ICRs (in FY00, 46 out of 265 ICRs were rated, and for the partial sample in FY01, 34 out of 169 ICRs, on the poverty dimension). As the Bank internalizes the Millennium Development Goals into its strategy and lending, it is hoped that a results orientation will be reflected through increased attention to the measurable poverty dimension of lending operations.

Forward-looking orientation has improved, although it remains weak relative to the other dimen-

Table A1.2

**Evolution of Quality of ICRs:
Three Dimensions
(percent satisfactory or better)**

FY96	ICRs exiting in following fiscal years				
	FY97	FY98	FY99	FY00	FY01 ^a
1. Analysis					
Coverage of important subjects	91	96	95	95	95
Ex post economic analysis (if applicable)	71	78	73	88	86
Internal consistency	92	93	96	94	95
Evidence complete/convincing	82	84	89	86	87
Adequacy of lessons learned	92	93	96	90	93
Poverty analysis ^a	n/a	n/a	n/a	n/a	54
2. Future operation					
Plan for future operation	82	84	81	89	86
Performance indicators for operational phase	63	74	72	76	81
Plan for monitoring and evaluation of future operation	69	63	66	70	73
3. Borrower/cofinancier inputs					
Borrower input to ICR	91	93	97	96	98
Borrower plan for future project operation	81	84	83	86	89
Borrower comments on ICR	88	89	94	95	97
Cofinancier comments on ICR (if applicable)	57	73	80	86	92
Overall quality of ICR	91	95	96	94	96

* Partial sample.

sions. ICRs are expected to report on a project's anticipated results on the ground in the operational phase, after disbursements. It is important to continue monitoring after disbursements, to ensure that the 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 92 percent satisfactory or better, these ratings were made on only about 70 percent of the ICRs that were reviewed; the remaining ICRs could not be rated on this dimension because information was not available on the plan for future operation. The other aspects of future operation of the project are less satisfactory, both in terms of the proportion of the ICRs that could be rated and the percent satisfactory of those that were rated. Performance indicators for future operation were rated in some 55 percent of all ICRs reviewed. Of those that could be rated, the proportion improved from 63 percent satisfactory or better in FY96 to 85 percent in FY01. The plan for monitoring and evaluation could be reviewed in only 42 percent of the ICRs over the FY96–01 period, and of these, the proportion of satisfactory or better improved modestly from 69 percent to 75 percent.² There is room for considerable improvement in identifying performance indicators and planning for monitoring and evaluation in the operational phase of Bank lending.

Continued improvement in borrower contributions to the ICRs where such contributions were available. ICRs are supposed to include contributions from borrowers, which encompass the borrower's inputs to the assessment, specific comments on the ICR, and the borrower's plan for future operation. Many ICRs lack one or more of these components. ICRs included borrowers' comments about 65 percent of the time throughout the period, and the quality of these comments has improved and now stands at a strong 95 percent. The plan for future operation was included in about 55 percent of the ICRs, and of these, the proportion rated satisfactory or better improved from 81 percent in FY96 to 91 percent in FY01.

ICRs are also supposed to have cofinanciers' comments, and this has increasingly been the case; in addition, the quality of these comments

has improved. Overall, the quality trend for borrowers' and cofinanciers' contributions suggests that the ICR process has become more participatory over time.

The overall quality of ICRs is high. As shown in table A1.2, ICR quality ratings over the last four completed exit fiscal years have remained fairly constant at around 95 percent. The preliminary results for FY01 exits indicate a slight decline, to 92 percent. The satisfactory ICR quality ratings reflect the internalization of the ICR methodology by operational staff. Since FY98 the ICR template has become part of the project document system for all completed investment operations.

Intensive Learning ICRs (ILIs) were introduced to focus operational efforts to enhance the learning of the ICR process. Although ILIs were expected to be 30 percent of the total ICRs, only 10 percent of the 271 FY00 ICRs evaluated were ILIs. The ILI share remains the same for the partial FY01 sample. No significant differences are found in ICR quality ratings by ICR type.

Ratings Changes³

At each stage of the project cycle, from implementation through completion and independent evaluation by OED, projects are rated in terms of their likely achievement of relevant project objectives (outcome). At completion and evaluation, projects are also rated on institutional development impact and sustainability. One indicator of the robustness of the ratings at each stage is the extent to which ratings are changed in subsequent stages. A second indicator of reliability is the net change in ratings—that is, the percentage of ratings that are decreased minus those that are increased. This shows the degree of overoptimism (if the net change is positive) or underrating (if the net change is negative) of the self-rating.

Table A1.3 shows these rating changes by exit fiscal year. They include rating changes on outcomes between: (a) the last Project Supervision Report (PSR) before the operation exits the portfolio, and the ICR, carried out by operational staff in the Region; (b) the PSR, ICR, and OED's ES; and (c) the ES and the Project Performance Assessment Report (PPAR), if one is carried out by OED. In any given year, about 25 percent of closed operations will be assessed.

Table A 1.3**Summary of Ratings Changes, as Percentage of Total Projects Reviewed at Each Stage, FY96–01 Exits**

Change	Exit FY																							
	FY96				FY97				FY98				FY99				FY00				FY01			
	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
Outcome	a	b	a+b	a-b																				
PSR to ICR	8	4	12	4	9	2	11	7	11	2	13	9	7	3	10	4	6	2	8	4	8	1	9	7
ICR to ES	8	1	9	7	5	3	8	2	4	1	5	3	5	0	5	5	5	0	5	5	5	1	6	4
PSR to ES	16	4	20	12	9	0	9	9	14	3	17	11	11	2	13	9	10	2	12	8	13	1	14	12
ES to PPAR	10	3	13	7	9	0	9	9	8	7	15	1	6	8	14	-2	10	10	20	0	n/a	n/a	n/a	n/a

Note: The data for FY01 exits represent a partial sample and reflect the processing of all ICRs through February 2002. In FY01 only two PPARs were done on operations exiting that year.

Outcome rating changes are included here only where the change is from satisfactory (highly satisfactory, satisfactory, marginally satisfactory) to unsatisfactory (marginally unsatisfactory, unsatisfactory, highly unsatisfactory) or the reverse. That is, rating changes within “satisfactory” and “unsatisfactory” categories are not included. Looking first only at changes between the final PSR and the ICR, there has been a slight downward trend in the number of total rating changes, but no discernible trend in net changes, which have been consistently positive. This suggests that even at the time of final supervision of operations, Bank staff tend to overestimate development outcome. Next, total rating changes on outcome are fewer between Regions and OED (ICR to ES) than within the Regions (PSR to ICR). Net changes have been either close to those within operations or somewhat lower and show no consistent trend over time. Thus, although OED has changed outcome ratings at the ES stage from ICRs in a net downward direction every year since FY96, its net changes have been more modest than those within the Regions.

The rating change between PSR and ES is shown here to indicate the degree and direction of differences between supervision ratings on outcome and OED ratings. This is relevant for providing an indication of the extent to which PSR ratings may be reliable predictors of outcomes. Given the magnitude of the changes in the same direction (downward) at both the PSR to ICR stage and ICR to ES stage, it is not surprising that the largest total and net outcome rating changes have generally occurred between PSR and ES.

The net rating changes are consistently downward by around 10 percent over the period.

In addition, outcome rating changes made *within OED*—that is, between ES and PPAR—have been quite large. Between FY96 and FY99, outcome ratings were changed in about 14 percent of operations assessed in more depth by OED. In most of these years, the net change was downward, suggesting that a more intense assessment of operations by OED results in a more rigorous rating than the ES, which is based on a desk review of the ICR.

The rating changes in FY99 and FY00 show a different pattern on net rating changes from previous years, but they are based on a relatively small number of assessments carried out to date by OED, and this pattern may not hold in the future. In FY99, for example, the net change in outcome ratings between ES and PPARs is based on 36 PPARs, and in FY00, it is based on only 23, compared to some 60 per year in FY96–98. In the future, the number of assessments for operations in these exit years will increase as OED carries out additional PPARs. The relatively large share of rating changes for the 60 or so assessments carried out each year by OED between FY96 and FY99 raises questions about the robustness of the ES process alone to verify outcomes. It suggests that in-depth assessments are needed to verify results, underlining the importance of OED’s PPAR program. This conclusion is strengthened by a review of other ratings—institutional development impact, sustainability, and Bank and borrower performance. It shows a similar pattern to outcome

rating changes between the Regions and OED and within OED.

For example, the net rating changes on institutional development impact between ICRs and ESs in recent years (FY00 and 01) have been positive (at 1 and 5 percent, respectively), suggesting that the Regions are more upbeat on this dimension than OED. Total rating changes are generally higher within OED's own ratings (ES to PPAR) than between the Regions' ratings and OED's (ICR to ES), and net changes have generally been downward, as with outcome rating changes. On sustainability, as another example, the rating scale changed in FY00, making comparisons with previous years difficult, but for the FY96–99 period, net changes were consistently over 12 percent

each year, and at the PPAR stage OED changed ratings downward at least 10 percent of the time. In FY00 and 01, after the change in rating scale, the pattern is similar to outcome ratings, with a net rating change of some 4 percent, in a downward direction.

Overall, these rating changes suggest that the Regions are more optimistic about development effectiveness than OED and that the PSR/ICR/ES process may result in an overly sanguine picture of development effectiveness. This emphasizes the need for the independent assessments of operations carried out by OED and suggests that further efforts to carry out impact assessments are needed as proposed under OED's FY03–05 work program.

ANNEX 2: REGIONAL QUALITY TEAMS

A COMPARISON ACROSS REGIONS

Regions are at the center of risk management in the Bank. To fulfill their mandate with respect to lending activities, most Regions have set up quality units or teams, whose function it is to ensure that risks are adequately identified and assessed, quality at entry is strong, the active portfolio is monitored, and management is informed of issues in a timely way. In the context of this year's AROE, OED interacted with the Regional quality teams to get an overview of their structure and functions. There are also substantial differences in the size and scope of these activities.

The attached matrix captures similarities and differences. LAC has the most long-standing quality unit, since 1996; AFR and ECA set up their units in 1999; and the MNA quality unit is the most recent, established formally in FY02. LAC covers up to 20 percent of loans through Quality Enhancement Reviews (QERs); ECA, as much as one-half to two-thirds (this latter figure includes those initiated by Network Anchors). AFR covers about 5 percent, which translates into a relatively high number of QERs because of the large number of credits/loans processed in that Region. In addition, the AFR quality unit is expanding the use of QERs to cover supervision activities for selected projects. MNA covered about 10 percent in FY02 through QERs, while SAR covered 5 percent, and EAP did not initiate QERs.

The Regional quality units do not have a formal clearance function. In practice, however, they may be heavily involved because it is difficult for an operation to be channeled to senior management for approval unless the quality unit's comments have been taken into account. In ECA, for example, the quality unit reviews most Project Appraisal Documents (PADs), but they systematically review and clear "informally" *all* Adjustable Program Loans as

well as all Learning and Innovation Loans. AFR, ECA, and LAC are active in reviewing both Project Concept Documents and PADs.

AFR, ECA, and LAC quality units have adopted criteria for identifying risky projects, relying on self-assessments by country and sector managers, and also using other criteria, such as size, complexity, or country risk factors. ECA has a risk management committee, which includes fiduciary and safeguard staff from within and outside the Region, charged with looking at a range of implementation issues that could have a (negative) impact on development effectiveness. AFR has separate processes for projects that are considered high risk for implementation reasons or from a fiduciary perspective. LAC does a systematic review of pipeline project risks, both reputational and developmental, once a year and of active projects twice a year, as part of the portfolio review. It bases its assessments on inputs from both country and sector units, as well as fiduciary and safeguard teams, civil society, the Regional external affairs unit, and its own analysis. As in ECA, the resulting list of high-risk projects is presented to a panel of experts from within and outside the Region, to validate the risks and provide suggestions for mitigating or managing them.

The MNA quality unit has developed project readiness filters for each country, comprising some seven to nine elements, including factors such as ensuring availability of counterpart funding for first year of project; designation of key project staff; compliance with fiduciary and safeguard aspects; and identification of appropriate monitoring indicators of performance. As in the three Regions discussed above, MNA reviews Project Concept Documents (PCDs) and PADs; it also reviews Project Supervision Reports of all at-risk projects and all ICRs, and has a Monthly Man-

agement Report (MMR), which is disseminated in the Region through the unit's Web site and reviewed monthly by the Regional Management Team. The MMR includes a section highlighting critical operational risks for management attention. This report received high ratings in a survey of MNA staff (see below).

EAP and SAR, the two most decentralized Regions (roughly 50 and 80 percent of operations are decentralized, respectively), have a lighter central oversight function. EAP pioneered the use of the readiness filters and has also broadened the Country Portfolio Performance Review and made it a quarterly process, concentrating on safeguard and fiduciary issues, as well as projects at risk. The CPPR in EAP now also includes projects in the pipeline. In addition, more recently, EAP has concentrated on training staff in a wide range of issues on portfolio and pipeline monitoring and has taken a somewhat different approach to risk assessments: for the past three years it has carried out a Special Operational Review in countries to identify systemic risks that could affect development effectiveness or the reputation of the Bank. In addition, both SAR and EAP have made proactive use of the Investigation Unit in the Department of Institutional Integrity in the Bank, to look into ongoing projects that are considered particularly at risk for corruption.

South Asia had a Regional quality enhancement unit until FY02, when it was disbanded. In its

place, the Region has strengthened oversight in the country units in the field by hiring operational advisors in each of the five country offices. This implies a greater reliance on self-assessments by country directors who ensure that risks are identified, assessed, and managed in both the pipeline and the portfolio. There is central oversight of fiduciary risks by the chief financial officer, with extra resources available for riskier projects. The Regional Management Team meets monthly to review the quality of the portfolio.

The Regional quality advisors have a monthly lunch meeting to share cross-regional experiences. They also interact to varying degrees with OPCS, QAG, and Network staff. One issue being explored by OPCS is the consistency across Regions of criteria for identifying higher-risk projects. To the extent that the criteria differ, aggregation across Regions is not reliable; developing a corporate view of systemic risks in lending is challenging. OPCS plans to have a report on this aspect soon.

MNA's Operations and Country Services Department carried out a survey of its staff for feedback on the quality unit's services. Two of the unit's products—the Monthly Management Report and the Regional Guidelines and Business Processes—were among the highest-rated items in the survey. It would be useful for the other quality units to get similar feedback from task managers and sector and country managers.

Table A2.1 Regional Quality Enhancement Teams (QET), as of March 2002

	AFR	EAP	ECA	LAC	MNA	SAR
Objectives of QET: formal TOR?	Yes, RVP announcement		??	Yes	Yes	Decentralized; no formal quality unit in HQ
Number of HL staff Report directly to Functioning since	5 1999	2 Core Services Unit 1998	3 Core Services Unit 1999	6 Operational Support Unit 1996	2 Country Services Unit July 2001	— — —
Quality assurance for pipeline						
QERs as % of projects past FY?	About 5%	No	Between one-half and two-thirds, systematic in certain sectors	About 20%	About 10%	About 5%
Review PCDs/PADs?	Yes, all PCDs, PADs	No	Yes, both PCD and PAD; all PADs are discussed at weekly CSU meetings	All PCDs; and PADs that go to ROC	All PCDs, PADs	In field
Criteria for high-risk operations?	Yes, country and sector units identify, plus apply other criteria: APLs; PRSCs; size; country; fiduciary	By type of loan; (PRSC, LL, etc.); also those identified by sector unit or Network	Yes, implementation, safeguard, fiduciary	Yes, based on inputs from Bank units, fiduciary and safeguard teams, civil society, the Regional external affairs unit, and own analysis	Yes, part of ROC rules; includes adjustment, country factors, reputational risks; MNA-Q also developing "project readiness filter"	Mainly on safeguards and fiduciary, identified by specialists
Monitor high-risk operations?	Through QER/ROC	ROC targets about 25% of operations for review, including PRSCs, SALs, LILs, APLs, and others as identified above	Risk management committee set up to cover review of high-risk projects	Yes, through annual risk reviews	Yes, through QER, ROC, and the Exceptions and Outliers report of the MMR	Mainly based on fiduciary, safeguard considerations

(Table continues on the following page.)

**Table A2.1 Regional Quality Enhancement Teams (QET),
(continued) as of March 2002**

	AFR	EAP	ECA	LAC	MNA	SAR
Portfolio monitoring	Yes, use PAR; and developing criteria for higher risk: older projects; country instability; use QERs for supervision of selected projects	Countries use CPPRs, plus make use of investigatory unit, INTIU	Yes, use PAR, regular meetings with sector and country managers, reported in MMR	Yes, use PAR	Yes, report through MMR, use PAR flag system, and CPPRs	Yes, use PAR; Operations advisors in field monitor; monthly RMT meeting
Separate process for fiduciary risk?	Yes, separate risk list, monitored by CFO		Risk management committee in Region includes financial and safeguard staff	Yes	Yes, Regional guidelines for FMA and through MMR	Yes, by CFO, extra resources assigned for higher-risk projects
Review ESW	No	No	No	No	A few; has plans to review all ESW systematically starting FY03;	No
Review/track ICR quality	Review ICRs; track ex post quality; do clinics on methodology	No	Reviews only IIs; don't track ex post quality		Review ICRs; track quality ex post, using OED data	??
Monitor own effectiveness?	No	Through OAG data	No	Yes, how???	Yes, survey in 02 on quality unit	No

Note: APL = Adaptable Program Loan

CFO = Chief Financial Officer

CPPR = Country Portfolio Performance Review

CSU = Core Services Unit

FMA = Financial Management Assessments

ILI = Intensive Learning ICR

INTIU = Institutional Integrity and Investigative Unit

LLI = Learning and Innovation Loan

MMR = Monthly Management Report

PAR = Portfolio at Risk

PCD/PAD = Project Concept Document; Project Appraisal Document

QER = Quality Enhancement Review

RMT = Regional Management Team

ROC = Regional Operations Committee

ANNEX 3: QUALITY AT ENTRY, OUTCOMES, AND PORTFOLIO AT RISK
 A COMPARATIVE ANALYSIS OF QAG AND OED FINDINGS ON LENDING OPERATIONS

Several years of QAG data are now available on quality at entry (QAE) ratings, and enough time has passed since the first Quality at Entry Assessment (QE1) that a number of the operations assessed by QAG have closed and been assessed by OED. This note examines the extent of agreement between QAG and OED on quality at entry as well as the extent to which the QAE ratings are good predictors of outcome.

OED compared QAG’s QAE ratings, assessed just after Board approval of the operations, and OED’s ratings on quality at entry for the same projects, assessed by OED at the time of exit. These projects are a partial sample, consisting of 46 closed projects. Table 3.1 shows the results of the comparison for all four QAE exercises (CY97; CY98; CY99; and CY00-June 01). If there were perfect agreement between QAG and OED on the

quality at entry of all the projects, the observations would fall in the upper left and lower right quadrants. Table A3.1 below shows that there was agreement on 39 of the 46 projects, or in 85 percent of the cases.

How do QAG’s QAE ratings correlate with outcomes, as rated by OED at the time of exit? These are shown in table A3.2 for the 42 projects whose outcomes were rated by OED at the time of exit.¹ Roughly 17 percent of the projects, or 7 out of 42, have different ratings. This is similar to the level of accord on quality at entry found above.

Another way to look at these figures is to ask what is the net difference in rating between QAG’s QAE and OED at exit? That is, of the 42 projects now closed and rated by both QAG and OED, what percent was rated by QAG as satisfactory at entry and what percent had satisfactory

Table A3.1

Quality at Entry: QAG and OED Ratings, for Closed Projects

	OED: quality at entry	
	Satisfactory	Unsatisfactory
QAG: Satisfactory at entry	39	4
QAG: Unsatisfactory at entry	3	—

Table A3.2

QAG’s QAE and OED Ratings on Outcome, for Closed Projects

	OED: outcome at exit		Total
	Satisfactory	Unsatisfactory	
QAG: Satisfactory at entry	33	6	39
QAG: Unsatisfactory at entry	1	2	3
Total	34	8	42

outcomes rated by OED? Table A3.3 shows these figures by year of QAG’s QEA and for the total sample. Thus, of the 42 operations that have been assessed by QAG for QAE and are now closed and rated on outcome, QAG found that 93 percent had satisfactory quality at entry, while 81 percent of these projects had satisfactory outcomes. Thus, the net change (decrease) in ratings between QAG at entry and OED’s outcome ratings at exit is some 12 percentage points.

The analysis of closed projects is a modest proportion of the total number of operations rated by QAG for QAE, as 87 percent of the projects rated by QAG for quality at entry are still active. Table A3.4 shows a comparison of QAE ratings and the latest portfolio at risk status. “At risk” signals that the project has at least three flags and is either considered “potentially risky” or “risky.” Of the 300 operations rated by QAG for QAE and still active, 79 percent have a Portfolio at Risk status consistent with the QAE rating (satisfactory at entry and nonrisky, or unsatisfactory at entry and at risk). The gross change in ratings is thus fairly substantial, at 21 percent.

On a net basis, however, the difference in ratings is much smaller, because projects with both satisfactory and unsatisfactory quality at entry ratings have moved into a different status during

implementation. Table A3.5 shows the net changes in ratings by QEA vintage and for all 300 projects. For QEA1, for example, of the 69 operations rated by QAG and by the Portfolio at Risk, 80 percent were rated satisfactory at entry and 80 percent were considered nonrisky in the most recent Portfolio at Risk, showing no net difference in rating (although they weren’t the same projects: 9 projects were considered satisfactory at entry but are not at risk, and 9 projects were considered unsatisfactory at entry but are not at risk). Overall, the net change in ratings between QAE rating and PAR status for the 300 projects assessed by QAG and still active is 3 percentage points.

There has been a historic net disconnect of some 8 to 10 percentage points between the last supervision rating and OED’s outcome ratings, as reflected in Evaluation Summaries. Thus, although the Portfolio at Risk status is not identical to supervision ratings, it too may overestimate the extent of satisfactory outcomes, as its primary focus is on implementation progress rather than on development effectiveness (except for the one flag that relates to development effectiveness). This suggests that the 3 percentage point net difference between QAG’s satisfactory QAE figures and Portfolio at Risk figures may understate the differences that will emerge in out-

Table A3.3	Rating Differences between QAG Quality at Entry and OED Ratings on Outcome				
	QEA1	QEA2	QEA3	QEA4	Total
QAG: percent satisfactory at entry	90	92	100	100	93
OED: percent satisfactory outcome	70	85	100	100	81

Note: QEA1 covered operations approved in CY97; QEA2 in CY98; QEA3 in CY99; and QEA4 covered operations approved from January 2000 through June 2001.

Table A3.4	QAE and Portfolio at Risk for Active Projects		
	Portfolio at risk		Total
	Nonrisky	At risk	
QAG: Satisfactory at entry	223	36	259
QAG: Unsatisfactory at entry	27	14	41
Total	250	50	300

Table A3.5**QAE and Portfolio at Risk for Active Projects**

	QEA1	QEA2	QEA3	QEA4	Total
QAG: percent satisfactory at entry	80	85	85	93	86
Portfolio at Risk: percent nonrisky	80	77	86	89	83

Note: QEA1 covered operations approved in CY97; QEA2 in CY98 ; QEA3 in CY99 ; and QEA4 covered operations approved from January 2000 through June 2001.

come ratings. Based on the net differences in ratings for the sample of the closed projects (12 percentage points) and the net difference in ratings in the active projects (3 percentage points) it is likely that the percent satisfactory in QAE ratings is some 8 to 10 percentage points higher than the independently validated outcome ratings will prove to be. In part, this difference is due to

unanticipated factors that come into play in the process of implementation.

The overall conclusion is that the trend in outcome ratings is partly due to the upward trend in quality at entry. One shouldn't expect, however, that the recent increases in the percentage of satisfactory quality at entry will be translated into equally satisfactory outcomes.

ANNEX 4: FY02 OED SUMMARY WORK PLAN AND RESULTS FRAMEWORK
AS OF MARCH 31, 2002

Objective and Strategy

OED's mission is to contribute to development effectiveness through excellence and independence in evaluation. Its strategic objectives are to focus its country, sector, and thematic evaluations on a "higher plane" while enhancing project evaluation quality. It aims to manage its work through focusing on results, investing in evaluation capacity within the Bank and among borrowers while establishing a healthy and balanced work environment for all OED staff and members.

Regular program	FY00 Actual	FY01 Actual	FY02 Plan	FY02 YTD	FY02 Proj.	Output indicators
Project evaluations						
<i>ICR Reviews</i>						
Number	281	280	280	197	280	Plan: 75% completed within 60 days from receipt in OED; actual: 71%. Plan: 100% completed within 90 days from receipt in OED; actual: 80%.
\$ Million	1.2	1.1	1.2	0.8	1.2	
<i>Project Performance Assessment Reports (PPARs)</i>						
Number	70	70	70	33	70	Increased efficiency in completion time not yet achieved.
\$ Million	3.1	2.4	3.3	1.7	2.6	
<i>Project Impact Evaluations</i>						
Number	1					Product discontinued in FY01 for budget reasons.
\$ Million	0.1					
Country Assistance Evaluations						
Number	10	8	7	7	9	100% CAEs delivered to CODE ahead of CAS review, where applicable; 7 delivered as targeted.
\$ Million	3.4	3.4	3.4	2.2	3.1	
Sector and thematic evaluations						
Number	6	4	5	2	6	OED input delivered to CODE in advance of SSP review, where applicable; 2 delivered as targeted.
\$ Million	5.5	5.2	4.0	3.3	4.6	
Corporate and process evaluations						
<i>Annual reports</i>						
Number	2	2	2	1	2	ARDE internal and external formal seminars/workshops within three months of completion.
\$ Million	1.1	0.8	0.8	0.6	0.8	
<i>Process evaluations</i>						
Number	1	1	1	0	0	Postponed for delivery in FY03 (HIPC report).
\$ Million	0.5	0.1	0.9	1.4	1.6	
<i>Other</i>						
Number	2	6				
\$ Million	0.3	0.9				

(continued)

ANNEX 4: FY02 OED SUMMARY WORK PLAN AND RESULTS FRAMEWORK
 (CONTINUED)

Regular program	FY00 Actual	FY01 Actual	FY02 Plan	FY02 YTD	FY02 Proj.	Output indicators
Evaluation development						
<i>Evaluation capacity development</i>						
Number	6	7	7	9	9	ECD program with high-intensity support in 9 countries and low-intensity support in 9 additional countries; ECD diagnosis supported in 4 countries; 2 new Bank loans with substantive ECD; support for the establishment of IDEAS, draft report on ECD to CODE.
\$ Million	0.4	0.6	0.7	0.5	0.7	
<i>KM, outreach, and learning</i>						
\$ Million	2.4	3.6	3.3	2.5	3.3	10 KM lessons papers; 29 information syntheses packages; 8 knowledge sharing events; 871 help desk queries; 30 queries disclosure-related; 400 new documents added to the Gateway; 353 contributors to the Gateway; 2 papers on "Evolution of the Evaluation Function in the Bank"; 45 Précis/FTBs; 4 books; 30 Working Papers; 2 Proceedings, Multilingual and external Web site content management. Training: 18 Workshops; 11 BBLs; 2 courses on "Rapid and Economical Evaluation Methods" and "Sampling Methods for Evaluation"; 2 core programs delivered to ACS staff; 120 attendees at Workshops; 145 attendees at BBLs; Delivery of "International Program on Development Evaluation Training."
<i>Methods and staff development</i>						
\$ Million	0.9	0.9	1.3	0.5	1.0	Revised PIF and audit procedures; input to OP/BP and ICR guidelines for adjustment operations; simplified project evaluation instrument; inputs to ICR guidelines for APLs, LILs, and guarantee-financed operations; tested and refined Country Information Form.
Total for regular program	18.9	19.2	19.2	13.4	19.0	

ANNEX 5: OED SURVEYS

Each year OED undertakes an evaluation of the quality, relevance, and efficacy of its own products and services as background for the Annual Report on Operations Evaluation. Since 1999, OED has surveyed clients about the quality and outcomes of its evaluations. The 2001 and 2002 surveys sought views from four sets of client groups:

- (i) Advisors and assistants to Executive Directors (EDs)
- (ii) Members of country teams in countries where a CAE has been undertaken
- (iii) Staff in the sector/thematic area on four sector and thematic (S&T) evaluations

(forestry, gender II, rural development II, and environment)

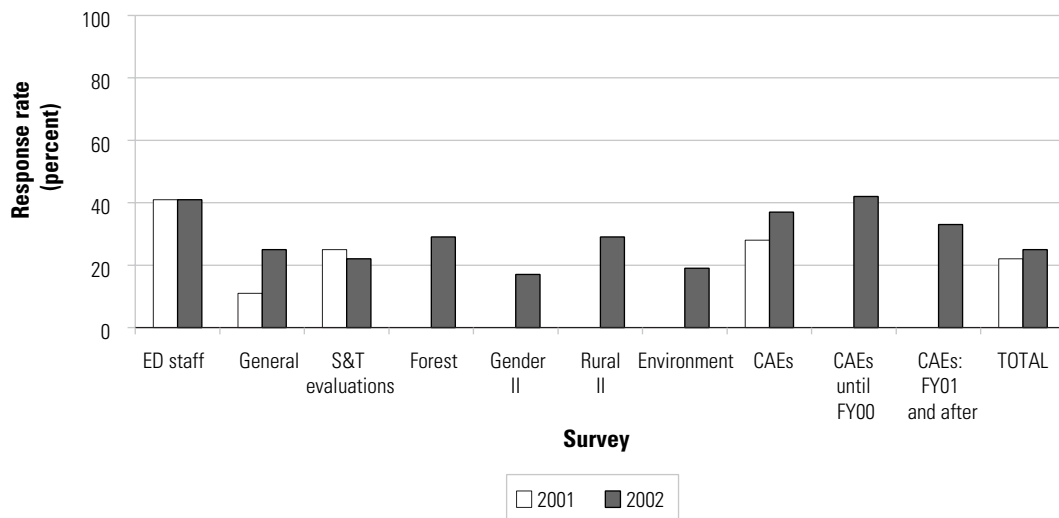
- (iv) Bank staff in Regions, Network Anchors, DEC, and WBI.

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 to results from previous years where appropriate.¹

Response Rates

The surveys were administered to a total of 3,237 staff over a 15-day period in February 2002. The

Figure A5.1 Response Rates



Note: OED surveyed respondents on four S&T evaluations: forest, gender II, rural development II, and environmental sustainability. The results of the survey on CAEs are presented in two sets: CAEs undertaken on or before FY00 and those undertaken since.

response rates ranged from 19 percent to 42 percent and averaged 25 percent across all four surveys compared to 22 percent for the comparable surveys in 2001.² The better response rate may be due to the use of an external tabulation firm and the assurance of anonymity.

Readership and Usefulness³

ED staff and Bank staff were asked about the types of OED evaluations they had read in the past 6 months. The list included country evaluations, sector evaluations, thematic evaluations, project audits, and corporate studies. The results are summarized in figure A5.2. Among ED staff who responded, all had read OED studies, a comforting finding, as the Board is the primary audience for OED reports.

Of the respondents to the general survey of Bank staff, 68 percent had read at least one OED product in the previous six months, of which 47 percent reported having read project audits, 35 percent CAEs, and 24 percent corporate studies. Of the respondents who hadn't read OED reports, 44 percent noted lack of time. Another 40 percent of the respondents noted that they were not aware of them or did not know where to find

them. Several new staff member respondents mentioned their lack of knowledge of OED.

Among Network members surveyed about the S&T studies relevant to their work, 35 percent of respondents reported having read the respective study, similar to findings last year. An additional 9 percent noted familiarity with the main findings of the S&T evaluations. Of respondents who had not read the report, 61 percent said they were not aware of it (compared to 40 percent of respondents in 2001) and a further 9 percent could not obtain a copy of it. Given that all the S&T studies that were the subject of the S&T surveys were available in ImageBank, OED may need to be more proactive in the dissemination of studies. Lack of time was a reason for not reading the reports for another 28 percent (35 percent in 2000–01).

Respondents who read OED evaluations were asked to rank usefulness, on a binary scale of useful or not useful. ED staff respondents that read OED reports continued to find them useful. The same is true of general survey respondents (figure A5.3).

Respondents were asked to comment on the quality of OED evaluations: timeliness, objectivity,

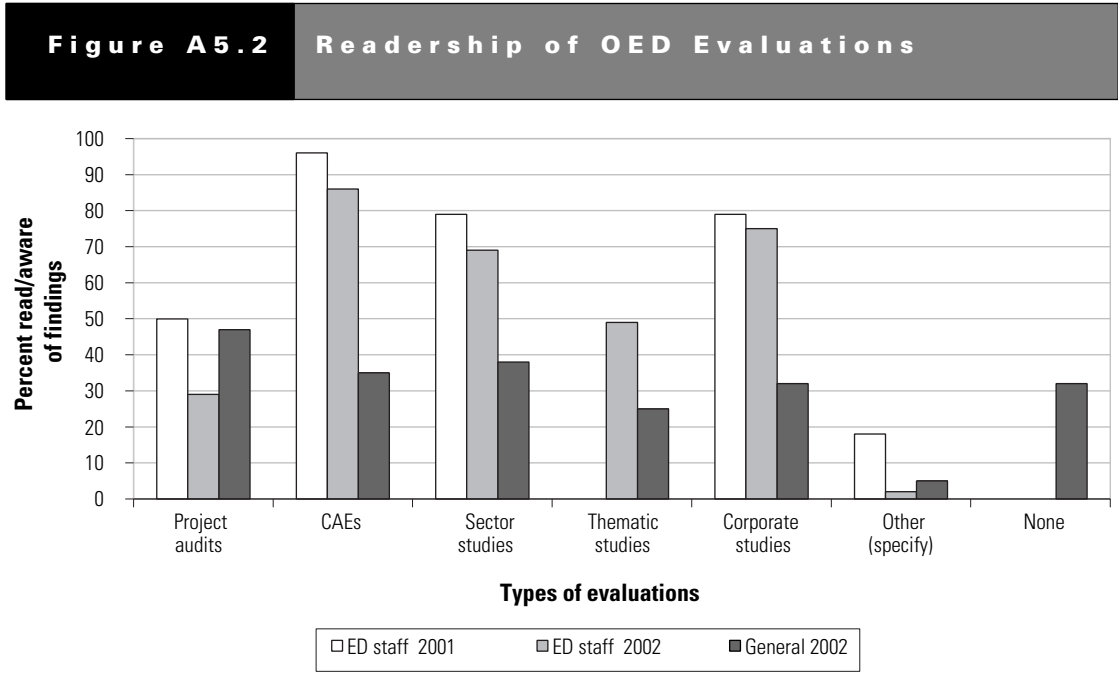
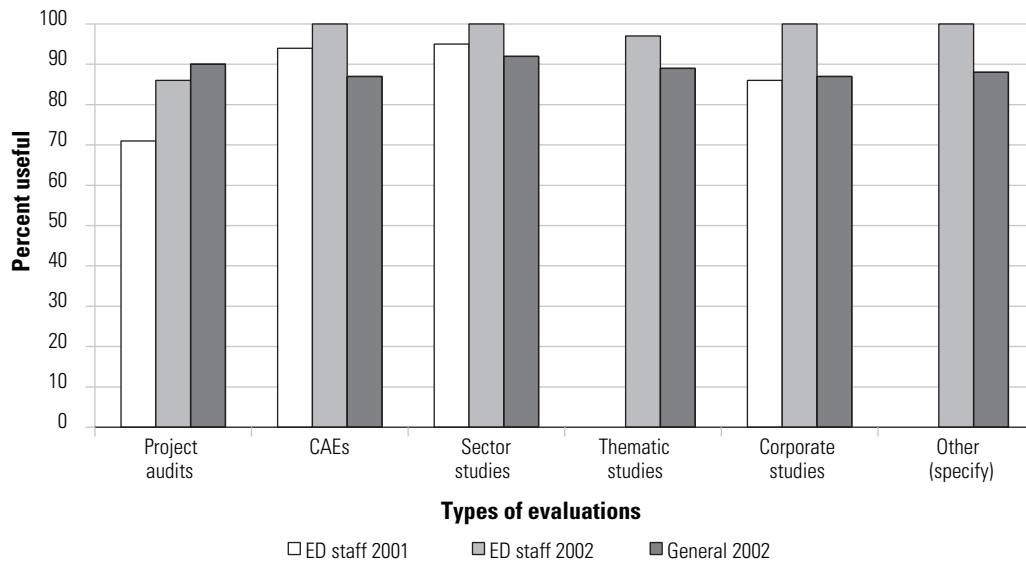


Figure A5.3 Usefulness of OED Evaluations

methodological soundness, and relevance to the work, overall quality, and, in some cases, quality of the recommendations.⁴ ED staff respondents rated OED evaluations very high on all dimensions of quality, as they did last year. Respondents to the general survey also rated OED

products high on most dimensions, but in each case there was a lower percentage of positive responses from the general staff survey than from the ED staff survey (figure A5.4a).

On country evaluations, respondents gave relatively strong marks for CAEs' relevance and time-

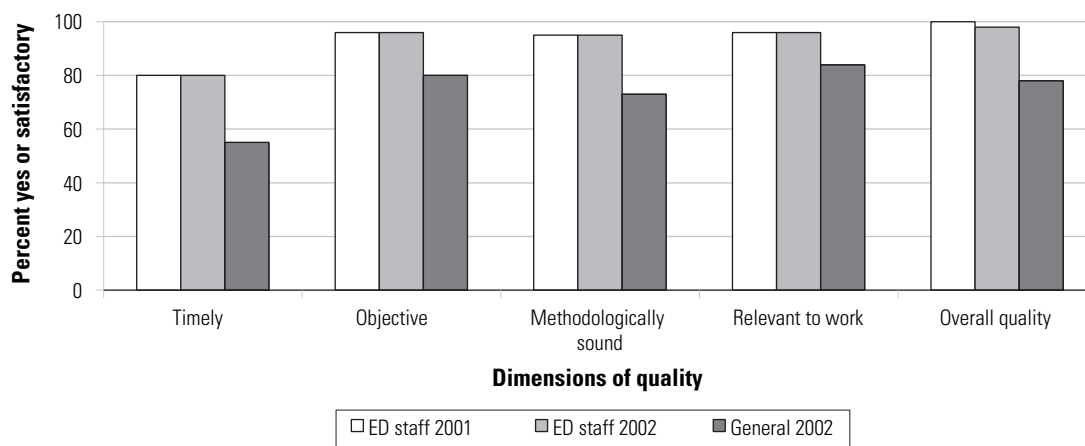
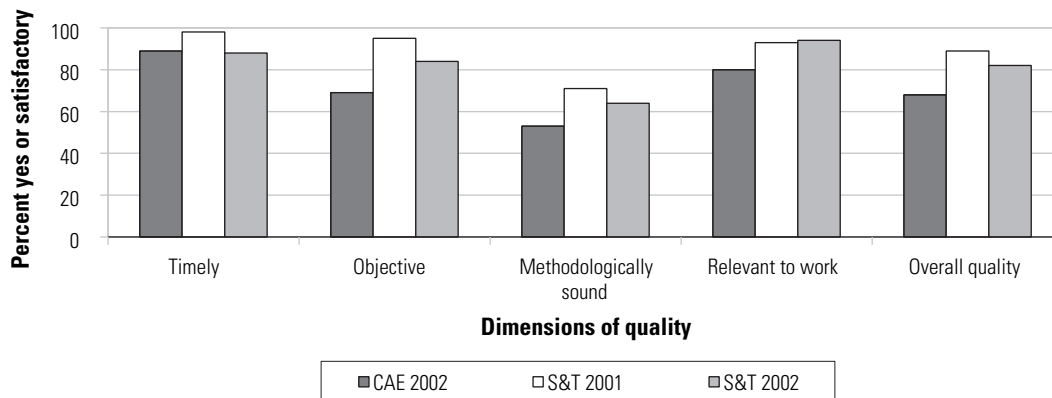
Figure A5.4a Quality of OED Evaluations

Figure A5.4b Quality of OED Evaluations

liness, but they found room for improvement regarding objectivity and perceived clarity of methodology (figure A5.4b). About one-fifth of the 45 respondents to the questions on CAE quality thought the CAE was of little or no relevance, one-third rated overall quality as fair or poor, and only a little over half of the respondents thought they were methodologically sound. Yet 93 percent of respondents said that if they were to start work on a new country, they would find a CAE to be useful. On the S&T surveys, over 80 percent of the 75 respondents thought the studies were timely, and relevant, while 64 percent thought the studies were methodologically sound (figure A5.4b).

To find out whether views of a CAE improved over time, OED compared the views on “old” CAEs undertaken before FY01 and “new” ones undertaken in FY01 or later. Comparing results from the surveys on CAEs carried out prior to FY01 to the more recently completed ones reveals differences that are statistically significant. CAEs completed before FY01 received more favorable feedback (table A5.1). It seems unlikely that the more favorable results on older CAEs are due to better quality or less controversial CAEs; if anything, OED has recently strengthened its CAE methodology. A more likely explanation is that as time goes on, staff either accept more readily the CAE findings, or with staff and manager turnover,

the new staff and managers look at the evaluation from a less personal perspective.

Influence of OED Evaluations

This year, the surveys asked whether the OED evaluation(s) read by the respondent or its recommendation(s) influenced the respondent’s view of development priorities and strategy and whether they influenced the Bank’s strategy, lending services, or nonlending services. Respondents were asked to rank influence on a four-point scale: strongly, somewhat, little, or not at all.

Among ED staff who responded to the survey, over 85 percent reported that their own views were influenced either somewhat or strongly by OED evaluations, in the concerned sector, country, or in the Bank as a whole (table A5.2). Two-thirds of ED staff respondents also said that OED evaluations influence the way Bank programs and projects are implemented. Some 75 percent of respondents to both the general survey and the S&T survey told us that OED evaluations influenced either somewhat or strongly both their own views of strategy and development policy at the sector level. Over two-thirds of CAE survey respondents reported that the CAE influenced both their own thinking on country strategy as well as that of the Bank. The percentages of S&T and CAE survey respondents who reported that the

Table A5.1**Quality and Influence of CAE
(percent)**

	Old CAE	New CAE
No. of respondents	20	25
Timely? (Yes)	95	84
Objective? (Yes) ^a	85	56
Methodologically sound? (Yes) ^a	80	32
Relevant to work? (very/somewhat) ^a	95	68
Very relevant	45	24
Somewhat relevant	50	44
Of little relevance	5	28
Not relevant at all	—	4
Overall quality (good or better) ^a	85	56

a. Statistically significant at the 5% level.

evaluation influenced Bank lending and non-lending operations was lower, at 50 percent to 64 percent.

Findings on Annual Flagship Reports

This year's surveys sought feedback on OED's apex reports: the Annual Review of Development Effectiveness (ARDE) from the ED staff and Bank staff (general survey) and on the Annual Re-

port of Operations Evaluation (AROE) from ED staff alone.⁵ Among ED staff respondents, the readership or awareness of the AROE (68 percent) is lower than for the ARDE (84 percent), although almost all ED staff who read these documents rated them as somewhat or very relevant to their work. Among general survey respondents, readership of the ARDE is low, at 29 percent. But of those who have read it, 89 percent

Table A5.2**OED Influence (percent somewhat or strongly)**

	ED staff		General	CAE	S&T	
	2001	2002	2002	2002	2001	2002
Number of respondents	56	49	358	45	125	75
Influence your view of strategy and development policy						
In the sector?	90	90	76 ^a	—	71	75
In the country?	84	85	—	66	45	—
In the Bank as a whole?	79	88	—	—	56	—
Way Bank projects/programs are being implemented?	—	67	—	—	—	—
Influence the Bank's						
Sector/country strategy?	—	—	—	71	—	83
Lending services?	—	—	—	53	—	50
Nonlending services?	—	—	64	—	—	64

a. The question in the general survey asked about influence on the respondent's own view of strategy/development policy in the sector/country.

Table A5.3 Findings on ARDE and the AROE

	ARDE		General		AROE	
	ED staff				ED staff	
	No.	%	No.	%	No.	%
Have read it	40	80	105	29	29	58
Have not read it but aware of its findings	2	4			5	10
Have neither read nor are aware of findings	8	16			16	32
Relevance (% very or somewhat):	42	100	95	89	34	97
Very relevant		57				50
Somewhat relevant		43				47
Of little relevance		—				3
Of no relevance		—				—

thought it somewhat or very relevant to their work (table A5.3).

OED Services

The general survey also aimed to measure the awareness and usefulness of OED's various evaluation services: project ratings, the OED Help Desk, summary information such as Précis, lessons learned, evaluation summaries, and other OED Web-based documents. Fifty-eight percent of the 551 respondents had used at least one of the services, and of these, 94 percent found them to be relevant to their work. Of the respondents who reported not using OED services in the past six months, 44 percent said they were not aware that such services existed and another 41 percent said they had no need of them.

Results by Demographic Profile

OED thought there might be differences in perceptions of OED work by staff level, particularly for CAEs, and possibly by headquarters versus field-based staff. Although demographic information was asked on all the surveys except the ED staff survey, the numbers of respondents by category are too small to draw valid conclusions. The results by demographic characteristic are available on request.

Recommendations from Respondents

All surveys asked respondents how they would suggest OED improve the quality of its products and were provided a list of recommendations the

respondents could check. Table A5.4 presents their responses. The most frequently checked recommendations among all respondents were for OED to broaden both internal and external consultations and to carry out wider dissemination. Views were more mixed on whether OED should emphasize ratings more. Over 40 percent of respondents among ED staff and 28 percent of S&T survey respondents said that more emphasis should be put on ratings, while among respondents to the general and CAE surveys, fewer than 20 percent of respondents thought that more emphasis on ratings was desirable.

Implications for OED

These survey results point to two areas of improvement for OED—dissemination and methodology.

- To improve dissemination, OED should work to:
 - Engage the Networks in identifying how OED products can most effectively reach interested staff.
 - Develop a plan for dissemination of major products in the approach paper, which should include a statement about the target audience and what the dissemination strategy will be.
 - Enhance the range of dissemination activities on which live presentations and workshops are conducted in addition to paper and Web-based document distribution.
 - Increase publicity for new evaluation reports.
 - Enhance awareness of OED and products and services at orientation for new employees.

Table A5.4 Recommendations

	ED survey	General survey	CAE survey	S&T survey
Number	51	373	51	75
Percent responding				
Broader external consultation	73	56	36	39
Wider dissemination	51	42	18	44
Broader consultation with Bank staff	41	67	49	59
More emphasis on ratings	41	19	11	28
Less emphasis on ratings	16	21	17	9
Introduce new type evaluation	12	14	—	
Track a particular sector/theme	4	9	—	
Other	16		40	15

To improve methods, OED should:

- Include an annex in every study, stating the methods, data sources used, and possible shortcomings and biases in the data.
- Improve the consultation process within the Bank. With field office staff, OED should make greater use of technology such as video conferencing.
- Improve consultation with stakeholders outside the Bank, with an explicit plan for focus groups, surveys, structured interviews, and other appropriate instruments.
- Continue to work on OED methods for evaluation, at the project and country level in particular.

ANNEX 6: CAE TRACER STUDY

Introduction

This annex provides an assessment of the impact of a sample of OED country evaluations on country strategies and Bank operations. The findings are based on a study conducted to trace the impact of specific Country Assistance Evaluation (CAE)¹ recommendations on country strategy and Bank operations. The study included a structured review of CAEs and CASs and interviews with CAE and Country Assistance Strategy (CAS) task managers, country team members, and sector specialists. The interviews provide lessons to enhance the CAE process and will supplement feedback on CAEs obtained from the staff survey conducted for the AROE.

Methodology

There were two components to the study. The first, a “tracer study,” identified CAE recommendations and tracked their application in country strategies and Bank programs. It is based on a review of CASs, active country portfolios, and structured interviews with key Bank staff. The second included countries that have had more than one CAE: this study reviewed the extent to which CAE recommendations have been incorporated into Bank assistance based on the assessment of “follow-up” CAEs.

The tracer study had two phases. First, CAEs were reviewed to distill key lessons and recommendations. This included, where available, interviews with CAE task managers to validate key points. Each of the CASs that followed was then rated for consistency with OED recommendations. The second phase analyzed the extent to which CAE recommendations have been reflected in actual assistance. This involved a review of country lending and nonlending programs and in-

terviews with a sample of country team members, based on a questionnaire tailored to each country.

Sample

Fifteen countries were chosen from the 52 country evaluations produced by the OED Country Evaluation and Regional Relations Unit. Ten countries were chosen from the sample of countries that had CAEs in FY98 and FY99 to ensure that sufficient time had elapsed for some implementation of the assistance program to have occurred. The CAEs were chosen to ensure that all Regions were included with at least one country, and no more than two. These ten countries were: Albania, Bangladesh, Bolivia, Indonesia, Malawi, Mozambique, Philippines, Sri Lanka, Ukraine, and Yemen. The other five were selected on the basis of having follow-up country assistance analysis by OED on that country several years later. These five were: Argentina, Ghana, Kenya, Morocco, and Zambia. In all cases, the CAEs analyzed were from 1998 or 1999; for the five countries that had follow-up country assistance analysis or a second CAE, the focus was on the earlier CAE; the more recent analysis was used as a reference only.

A total of 25 Bank and OED staff were interviewed for both components of the study. This includes CAS and CAE task managers, country coordinators, country officers, country economists, and sector and thematic specialists.

CASs Are Highly Consistent with CAEs

Overall, CASs are highly consistent with recommendations in OED CAEs. Of the 140 recommendations identified and tracked for the ten CAEs, only about one-fifth were not reflected in CASs.² The level of consistency between OED

recommendations and Bank strategy was high across most countries. In *Yemen*, for example, the FY99 CAE was said to have prompted efforts to reduce the autonomy of Project Implementation Units (PIUs) and to reduce the scope and complexity of projects. In *Indonesia*, the shift in the focus of assistance to direct poverty alleviation was said to have been prompted by the FY99 CAE. The consistency between OED recommendations and country strategies is rated moderate in only one case, *Sri Lanka*, mainly because no full CAS has been produced since the FY99 CAE due to the political situation (a progress report was produced, however, in FY99). A full CAS was drafted but was initially on hold pending a resolution to political uncertainty in the country and is expected to be finalized once the new government has prepared a Poverty Reduction Strategy Paper. According to the country team, CAE recommendations will become more relevant with implementation of a full CAS in a less constrained operating environment. Thus, in most countries in the tracer study, the CASs were strongly consistent with CAE recommendations.

The tracer study tried to determine why some of the recommendations were not reflected in CASs, to see if there were any patterns that could provide lessons for future CAE recommendations. A number of reasons were cited by Bank staff, including disagreement about overall relevance, or specific recommendations considered either too vague and nonactionable or too specific. CAE recommendations were not addressed in the following areas:

- *Monitoring*: CAE recommendations for increased monitoring were not adopted by CASs in a number of instances. This included establishing indicators to serve as early financial crisis warning systems (*Albania, Indonesia*); poverty monitoring; environmental monitoring; and setting quality indicators for social services. The recommendations were considered by country teams to be either unnecessary or too vague to be useful.
- *Pace of reform*: In some cases, CAE recommendations were considered to be unrealistic or premature in terms of the proposed pace of reform. This was particularly the case with recommendations on privatization in *Bangladesh*

and *Ukraine* and devolution in *Sri Lanka* (where devolution had to be addressed in the context of a larger political solution to the conflict).

- *Specific instrument or specific ESW*: CAE recommendations for specific pieces of ESW or specific instruments were not consistently adopted by the CAS. For example, in *Sri Lanka* a range of CAE-recommended nonlending was not addressed.

In some cases, recommendations not mentioned in CASs were later implemented. In two instances, specific OED recommendations on selectivity were not taken up in the CAS produced close to the CAE, but were subsequently adopted (*Mozambique, Philippines*). In *Mozambique*, the CAS went to the Board one month before the CAE, and although the draft CAE was available to Bank management, its recommendations may have been too late to be fully adopted. In the 1997 CAS for *Mozambique*, the principle of selectivity was accepted, for example, but it was argued that it would take some time to implement with respect to Bank financing because both government and donors looked to the Bank to be involved in many sectors. Following a donor strategy retreat in FY99 and by the time of the next CAS in 2000, selectivity was further advanced. The *Philippines* CAS in FY99, although building on the FY98 CAE's recommendations and incorporating most of them, diverged from a specific OED recommendation to discontinue lending in the social sectors, given the riskiness of the portfolio and the availability of donor funds. The country team noted that the recommendation, considered contentious at the time, was later adopted as no lending in either health or education was approved.

Implementation Has Been Less Consistent with CAE Recommendations than with CASs

Implementation varied either somewhat or considerably from the proposed strategy or time-frame in most of the countries. Implementation of 45 percent of the recommendations tracked was weak or absent.³ The most commonly cited reasons were low country capacity and commitment, and shortfalls on the Bank's part. In some cases, unanticipated changes in the operating environ-

ment and, therefore, in country priorities reduced the relevance of CAE recommendations or proposed CAS programs.

Country Factors:

- *Overestimation of country capacity or commitment (Bangladesh, Philippines)*
- *Unforeseen deterioration in the operating environment (governance, Bangladesh; conflict, Sri Lanka)*

Bank factors:

- *Shifting priorities impact nonlending programs.* In *Albania*, the FY99 CAE recommendations for some nonlending activities taken up in the FY99 CAS were not implemented given changes in the priority and allocation of country resources, as well as activities of other donors.
- *Pressures to lend result in deviations from strategy.* In *Kenya* the Bank moved from a no-lending scenario—recommended by OED in its FY97 CAE and proposed in the FY97 CAS—to a program of emergency and adjustment lending. This higher lending scenario was implemented without some of the key governance-related triggers having been met. In *Morocco*, although the FY97 CAS incorporated many recommendations made by OED in its FY97 CAE, and some of them were followed (for example, a Resident Mission was established and ESW has been better disseminated), there were important deviations between the FY97 CAS and subsequent lending. In particular, OED recommended greater selectivity in adjustment lending and a focus on structural aspects of fiscal problems; the FY97 CAS base case called for one relatively small adjustment loan, and focus on fiscal issues was prominent in the CAS. During the period covered by the FY97 CAS, however, more than half of the lending commitments were for adjustment lending, the triggers to reach the high-case lending scenario that included higher adjustment lending were not met, and fiscal problems were not adequately addressed and continue today to be a serious concern.
- *Internal Bank incentives inhibit selectivity.* Internal Bank incentives were cited as a constraint to greater selectivity in lending and

nonlending activities. Examples include staff/sector units rewarded for projects and not for nonlending, and mandatory ESW requirements taking up an increasing share of resources, regardless of priorities in the country concerned (*Malawi, Mozambique*).

A number of country officers/coordinators would have liked more specific guidance on *how* the Bank could/should implement CAE recommendations as opposed to strategic goals. For example, the country coordinator for *Malawi* found that the CAE presented an excellent agenda of the priorities for government action, but the challenge the team faced was to find ways the Bank could effectively support this agenda in the prevailing country conditions.

In several cases (*Albania, Mozambique*), PRSPs or SWAPs were cited as playing an important role in enabling the Bank to overcome constraints and making it possible to implement CAE-CAS proposals relating to consultation, participation, donor coordination, and country ownership. The PRSP/MDG process is also noted for highlighting the importance of setting indicators that country teams hope will assist both CAS M&E and evaluation in the country (*Albania, Bangladesh*).

Country Team Views on the CAEs

The tracer study included a structured interview with country teams to obtain feedback on the CAE process as a whole. Bank staff were asked to comment on both the country evaluation process and product, including the level and nature of interaction between OED staff and the Region and the structure of CAE recommendations.

Perceived OED impact is mixed. Overall, staff reported that CAEs provide useful input for the CAS process and for incoming country directors and country officers/coordinators. Dialogue with the country team in preparing the CAE was said to have been useful in stimulating new approaches within the team. In other cases, however, CAEs were said to have reinforced conclusions the Bank had already arrived at (*Bangladesh, Philippines, Ukraine*).

Timeliness. Almost one-third of respondents said that CAEs needed to be scheduled to allow sufficient time for lessons to be incorporated into

Bank strategy. One country economist noted that the high level of interaction with OED during CAS preparation had proven “somewhat taxing.” Only in the case of *Mozambique* was the CAE thought to have come too late to influence the CAS.

CAE recommendations should be actionable but not restrictive. According to respondents, the type of CAE should vary depending on the type of country. For countries with an established history of engagement with the Bank, strategic directions are considered clearer and detailed sector-specific recommendations more useful. In-depth sector analysis was also said to be particularly relevant for big countries with large lending programs. Many CAE recommendations were found to be too generic to be useful, or lacking internal consistency. For example, recommending greater selectivity without identifying activities to be cut back, or advocating the mainstreaming of environment or gender in the CAS without evidence from the CAE.

Overly detailed sector recommendations were seen as inhibiting the flexibility needed by the Region to respond to changes in the implementing environment. One example of overly comprehensive recommendations is the *Mozambique* CAE, which includes sector-specific recommendations in each chapter as well as overall recommendations for country strategy. Identifying individual potentially monitorable elements from each recommendation leads to hundreds of in-

dicators, without any clear guidance from the CAE itself on prioritization. In addition, many of its recommendations endorsed existing approaches.

Implications for the CAE Process

The tracer study highlights the need for greater consistency in both the structure and purpose of CAE recommendations. Based on the review conducted, suggestions to improve the structure of CAE recommendations include:

- i. Include more specific guidance on *how* recommendations are expected to be implemented.
- ii. When the CAE is linked to a new CAS, ensure it is completed before the CAS is due.
- iii. Limit or exercise greater selectivity in the number of recommendations.
- iv. Separate recommendations from endorsement of existing approaches that are already internalized by the country team.
- v. Specify indicators where relevant.

Changing country conditions and institutional priorities are liable to make specific recommendations less relevant over time. It is the spirit of recommendations or the strategic direction of the CAE that is most important. This points to the need for self-monitoring and self-evaluation and the value of follow-up CAEs.

ANNEX 7: CHAIRMAN'S STATEMENT, COMMITTEE ON DEVELOPMENT EFFECTIVENESS (CODE)

On July 17, 2002, the Committee on Development Effectiveness (CODE) met to discuss the findings and recommendations of the *2002 Annual Report on Operations Evaluation (AROE)*, along with the draft *Management Response to the AROE*, prior to their both being discussed at the Board on August 1, 2002.

Background. CODE and the Board review the OED AROE on an annual basis, as part of their oversight function to assess, the adequacy of and progress in the Bank's control, monitoring, and evaluation processes. The FY02 AROE provides a ten-year perspective on developments in risk management and monitoring and evaluation. It confirms that solid advances have been made in M&E and development risk management. But substantial gaps also still remain. Key recommendations to enhance the Bank's evaluation and control framework are that: (a) the Bank design performance measures to track achievements in relation to the MDGs; (b) strengthen the evaluation and control framework for sector strategies, partnerships, and grants, including developing an operational policy for evaluation; (c) improve self-evaluation and risk assessment at the country level; and (d) improve and widely disseminate OED's evaluation methodologies.

Conclusions and Next Steps. The Committee commended OED for an impressive report and welcomed its focus on risk, monitoring and evaluation (M&E), and results-based management (RBM). *The Committee endorsed in general the findings and recommendations of the AROE* and felt that it raised a number of important and difficult issues, but cautioned that the Bank must be clear about the limits of outcomes measure-

ment. The committee stressed that the focus on measuring results will inevitably be complex and exposes the Bank to risks. Members stressed the need *for evaluation to support better the Bank's move toward results-based management*, especially in an environment where accountability of international institutions has greater visibility. *Management agrees with the overall thrust of OED's recommendations.* The Board will discuss the report on August 1, 2002. The revised Management Response will be shared with the committee prior to its adoption into the FY03 MAR.

Committee members discussed a number of issues, including risk assessment and mitigation; outcome measurement and the MDGs; links between assessment and results; M&E capacity and OED evaluation methods; and the management response to the report. The following specific points were raised:

Risk Management. The committee asked to hear more about how management is proceeding in this area, including the status of the Risk Assessment Committee. Management noted that risk management and better risk taking were now widely accepted as key to securing good development results and that the challenge was to integrate the many control mechanisms already in place across the Bank into a more coherent framework. Members would have *appreciated a broader focus on risk beyond development and operational risks.* The DGO noted that while this report focused on development risk, other aspects of risks were being addressed in other committees. The committee *agreed with the AROE's recommendation that an integrated risk management framework was essential* to address all aspects of risks facing the Bank. The committee stressed that *it*

would be essential to look more carefully at how to align incentives and processes to improve risk management, and to avoid layering heavy and costly controls onto Bank operations.

Outcome Measurement and the MDGs. The committee *recognized the limited collective ability of the Bank and its partners to measure outcomes and address attribution issues at the country level*. The committee stressed the need for the Bank *to be clear about the limits of outcomes measurement*, including with respect to the MDGs, to avoid unrealistic expectations and heightened reputational risk. It also stressed the importance of moving quickly to develop some intermediate indicators to allow the Bank to track progress. OED said that the next AROE will focus on the MDGs and that it would benefit from management's ongoing work on measuring for results.

Links between Assessments and Results. The committee said it would like to hear more about the *links between results and OED's evaluation approaches*. The DGO highlighted that evaluation had demonstrably improved the Bank's performance, especially at the project level, and that it was now crucial to fill the evaluation gaps at the country and the global levels. Some speakers said it would have been helpful if the paper had *highlighted more strongly where OED feels internal arrangements are resulting in better outcomes for borrowers, and how best to address concerns about gaps in the methodology*. OED pointed to prior evaluations dealing with various aspects of the question and concluded that areas in the Bank's programs with weaker risk assessment processes tended to have poorer outcomes and performance. Speakers noted that *accountability within matrix management, including with regard to the role of the Regions and networks in quality assurance, remained a challenge* and they suggested that it would be useful for CODE to revisit this issue.

The committee noted that programmatic lending presented particular challenges to evaluation and to tracking results and asked for further work by OED about the relative merit of investment versus programmatic lending. The committee

agreed with OED's proposal to strengthen the control and evaluability of grant-funded Bank activities. Some also supported *applying the same control and evaluation standards for traditional investment loans to grant funds*, as a way of increasing accountability and results. OED informed the committee that it is working with management on addressing the challenges presented in evaluating programmatic lending, given its longer time frame and objectives.

M&E Capacity and Evaluation Methods. The committee was *cautious about supporting OED's recommendation that monitoring and evaluation (M&E) be made a fiduciary requirement*. OED noted that there is a statutory requirement to undertake an evaluation on every lending operation at completion and that M&E is mandated as a requirement as well. Members requested more information on the implications of rigorous implementation of such a requirement for the Bank and for its clients. Members raised other issues related to M&E, including the importance of *the Bank as a knowledge Bank making demonstrable progress in its efforts to build countries' capacity for M&E*. It would be important to encourage country ownership and management of M&E systems, but at the same time to avoid building institutional enclaves. Members also stressed the *urgency of OED and management agreeing on an approach to country evaluations*. The committee encouraged efforts to incorporate the views of clients and the expanded use of tracer studies into OED's self-assessments. Management informed the committee that it, with OED's advice, was well advanced in the development of an umbrella operational policy for evaluation.

Board and Committee Review of Annual Reports. A member suggested that in the future it would be useful for papers coming for Committee and Board review, including the Annual Reports on Evaluation, to *state clearly in the paper the reasons for Board consideration; areas in which Board comment was being sought; the choices open to the Board; and the actions expected of the Board*. OED noted the committee's request for more specificity in its reports on what

was being sought from the Board and also noted member's requests for more detailed recommendations. OED *underlined that CODE and the Board were the main clients for the apex reports.*

Management Response. The committee welcomed the alignment of the management response with the AROE's recommendations but would have appreciated *more specificity with respect to the actions management planned to*

take to address the AROE's recommendations.

It was agreed that management will update its response to OED's recommendation regarding performance measurement in the context of the Board's discussion of the report and the Development Committee's review of the paper *Better Measuring, Monitoring, and Managing for Development Results* in September 2002.

Pieter Stek, Chairman

Chapter 2

1. Over ten years up to FY89–91, the proportion of unsatisfactory outcomes of lending operations had increased from 13 to 34 percent and the share of problem projects in the portfolio, from 10 to 17 percent.

2. OED's *Review of Aid Coordination and the Role of the World Bank*, Report No. 19840, October 1999, found that survey respondents (Bank staff, government officials, and donor representatives) thought that aid coordination was highly relevant for poverty reduction, but were neutral to negative in their assessment of the impact of aid coordination on poverty reduction (pp. 9–16).

3. AROE 2000–01, pp. 11, 21–22.

4. *Evaluation of the World Bank's Knowledge Services and The World Bank's Approach to Global Programs: An Independent Evaluation*.

5. Bank research found evidence that IBRD borrowers prefer other sources of financing (domestic, bilateral, Regional) to Bank funding for infrastructure such as dams, slum upgrading, transportation.

Chapter 3

1. For example, if benefits were delayed by a year, if costs went up or output prices declined by 10 percent, etc.

2. Management notes that appropriate and consistent identification, mitigation, and management of risks are just as important in investment projects as in adjustment operations, with the record showing no significant difference in the treatment of risks between the two lending instruments.

3. AROE 2000–01, p. 11.

4. An OED study of grant programs made five specific recommendations (see Précis 224, "Grant Programs: Improving Their Governance"). OED's ongoing evaluation on Global Programs reviews progress on each of them.

5. "The World Bank's Approach to Global Programs: An Independent Evaluation," March draft, p. 36.

6. This applies equally to partnerships. Some progress has been made recently by DGF guide-

lines on in-house secretariats, but DGF still needs to strengthen the arm's-length relationship with programs that are essentially carried out within the Bank, such as the Public-Private Infrastructure Advisory Facility.

7. To establish a baseline for the ongoing quality improvement effort, QAG, at the request of management, reviewed the quality of CPARs and CFAAs carried out between January 2000 and June 21 (prior to their being mainstreamed as ESW products) and found, as expected, that the quality was lower than for ESW, when measured against the same standards.

8. Relevant IMF instruments are the Code of Fiscal Transparency, Safeguard Assessments (relating to central banks), and Red Cover Reports. In December 2001 the Bank and several partners launched the Public Expenditure and Financial Accountability Program to reduce transaction costs borne by client countries from potentially overlapping diagnostics. The program will inventory diagnostic instruments and develop a comprehensive approach to their use.

Chapter 4

1. Management believes that this conclusion neglects the fact that Regional sector managers report to both Network and Regional sector management.

2. In an attempt to compare OED and QAG findings on ESW, we found only 13 reports that had been assessed by both OED and QAG out of a total 265 examined by QAG. OED looks at ESW primarily in the context of country, sector, and thematic evaluations, and often relies on QAG evaluations for assessment of quality. For the 13 products that OED had reviewed, there were few disagreements, and they focused mainly on likely impact (assessed by QAG) versus actual impact (assessed by OED).

Chapter 5

1. Inspection Panel Reports are simultaneously distributed to the Board and senior management.

2. Management notes that although management and staff have considered the OED method-

ology in some detail, there is no consensus that it is the tool needed for more consistent self-evaluation. As noted in figure A5.4b of Annex 5 of this report, the quality of Country Assistance Evaluation methodology is an issue. To address the challenge of improving self-evaluation in country assistance strategies, management will focus on this topic in its on-going work on better measuring, monitoring, and managing for results; a report is scheduled for discussion by Executive Directors in September 2002.

3. "Measuring Outputs and Outcomes in IDA Countries," Board paper prepared in context of IDA13 replenishment process.

Chapter 6

1. Even in a performance-based evaluation system, it is useful to have background documentation of the process by which a given strategy was arrived at, or a project design put in place, to learn from successes and mistakes. Such documentation has become scarcer since decentralization and email.

2. "Building Staff Capacity for Development: A New Learning Framework," May 2001, p. 2.

Chapter 7

1. The tracer study was a desk review of Bank documents, supplemented with interviews of Bank staff. Several of the countries mentioned here have had more recent CAEs; the period reviewed was between the earlier CAE (FY98 or 99) to December 2001. A more detailed presentation of the tracer study is in Annex 6.

2. Management does not concur with this interpretation of events. The lending in question was in response to opportunities to assist clients in areas of strategic priority.

Chapter 8

1. Management notes that the record shows no significant difference between adjustment and investment lending instruments in the extent to which risks are appropriately and consistently identified, mitigated, and managed.

2. OED has been carrying out sector-, thematic-, and country-level evaluations for over five

years, but for purposes of an operational scorecard that measures performance in a consistent and regular way, and to enhance organizational learning, self-evaluation at all levels must be strengthened.

Annex 1

1. The drop in the assessment ratio after FY96 reflects the agreement between OED and the Board to reduce the percent coverage to allow OED to focus on evaluations at a country, sector, and thematic level.

2. For FY96, this translates into a total of 63 ICRs with satisfactory or better performance indicators for the future out of a total of 180 ICRs reviewed, or a rate of 35 percent satisfactory; for FY01, the corresponding number is 53 percent. For the plan for monitoring and evaluation, a total of 53 ICRs were rated satisfactory or better out of 180 ICRs reviewed for FY96 exits, for a rate of 29 percent; the corresponding figure for FY01 exits is 34 percent. The projects themselves may have had such indicators or plans for monitoring and evaluation, but the ICR did not reflect this.

3. QAG 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 unsatisfactory for achieving their development objectives." The term "net disconnect" applies to only one of the four instances for "rating change."

Annex 3

1. Four of the 46 closed projects were not rated for outcome. They were: (i) Mexico's Federal Roads Modernization Loan, \$475 m, never signed, withdrawn 18 months after Board approval, withdrawn because of lack of counterpart funds and difficulties in implementing resettlement; (ii) Mexico's Aquaculture Loan, \$40 m, cancelled after three years with only \$1 m disbursement; Ministry of Finance did not agree with project; (iii) Indonesia's Maluku Regional Development Project, \$16.3 m, cancelled after about two years due to civil unrest; and (iv) Russia Second Highway Re-

habilitation and Maintenance Loan, \$400 man, allowed to lapse nine months after Board approval because conditions for effectiveness (PIU) were not met. All four of these were considered satisfactory by QAG at the time of entry.

Annex 5

1. In 1999, OED piloted very short surveys on S&T evaluations and CAEs, PPARs, ICR Reviews, and Outreach. The following year the surveys were substantially changed. This year, sampling was somewhat different than last for the CAE surveys and the wording in some questions was also changed to make the meaning clearer. In both 1999 and 2001, the number of responses to some of the surveys and some questions was low. Comparisons of this year's survey results with previous year's survey results are therefore valid only for certain questions in some of the surveys, and these are presented here.

2. The response rate in 1999 for CAEs was 39 percent (n=41) and for S&T evaluations was 46 percent (n=59).

3. In 1999, the survey targeted the various OED publications products: working series, FTBs, précis, evaluation reports, etc. No individual breakup of the evaluation reports was undertaken. Hence results from 1999 are not comparable.

4. Respondents also rated the evaluation on quality overall using a 4-point scale: excellent, good, fair, and poor. A scale of highly satisfactory, satisfactory, unsatisfactory, or highly unsatisfactory was used in 2000–01.

5. The AROE reviews the control and evaluation frameworks in the Bank and is not directly relevant to operational staff.

Annex 6

1. For the purposes of this note, the term CAE refers to all OED country evaluations, including early evaluations entitled "Country Assistance Review" and "Country Assistance Note."

2. Recommendations are not weighted for significance.

3. Recommendations are not weighted for significance.

REFERENCES

- COSO (Committee of Sponsoring Organizations of the Treadway Commission). 1994 "Internal Control—Integrated Framework." New York.
- OED (Operations Evaluation Department). Ongoing. "The World Bank and Global Public Policies and Programs: An OED Evaluation." <http://www.worldbank.org/oed/gppp/#1>
- . 2002a. "Grant Programs: Improving Their Governance," OED Précis No. 224, Summer 2002. Washington, D.C.
- . 2002b. *Annual Report on Evaluation Capacity Development*. Washington, D.C.
- . 2002c. *IDA's Partnership for Poverty Reduction: An Independent Evaluation of Fiscal Years 1994–2000*. OED Study Series. Washington, D.C.: World Bank.
- . 2001a. *2001 Annual Review of Development Effectiveness: Making Choices*. OED Study Series. Washington, D.C.: World Bank.
- . 2001b. *2000–2001 Annual Report on Operations Evaluation*. Washington, D.C.
- . 2000. *2000 Annual Review of Development Effectiveness: From Strategy to Results*. OED Study Series. Washington, D.C.: World Bank.
- . 1999. *Aid Coordination and the Role of the World Bank*. Report No. 19840. Washington, D.C.: World Bank.
- World Bank. 2001a. "Building Staff Capacity for Development: A New Learning Framework." Washington, D.C.
- . 2001b. *Assessment of the Strategic Compact*. Washington, D.C.
- . 2000. *World Development Report 2000/2001: Attacking Poverty*. New York: Oxford University Press for the World Bank.
- . 1990. *The World Development Report 1990: Poverty*. New York: Oxford University Press for the World Bank.
- WBI (World Bank Institute). 2001. "The Performance of WBI's Core Courses, FY99–01." WBI Evaluation Brief. Washington, D.C.

