WORLD BANK

OPERATIONS EVALUATION DEPARTMENT

THE FIRST 30 YEARS

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.

WORLD BANK OPERATIONS EVALUATION DEPARTMENT

World Bank Operations Evaluation Department

The First 30 Years

Edited by
Patrick G. Grasso
Sulaiman S. Wasty
Rachel V. Weaving

2003
The World Bank
Washington, D.C.

http://www.worldbank.org/oed

1818 H Street, NW
Washington, DC 20433
Telephone 202-473-1000
Internet www.worldbank.org
E-mail feedback@worldbank.org

0-8213-5549-X
e-ISBN 0-8213-5550-3

Printed on recyled paper.

Contents

Part III. Relations with Clients and Partners

Part IV. Challenges of Development Evaluation

Annexes

Foreword

James D. Wolfensohn
President, The World Bank

The development agenda for the coming years will have a significantly greater focus on development effectiveness. Development effectiveness is a matter of deep concern to the World Bank's Board and management, and fortunately the activities of the Operations Evaluation Department have put us in a very good position to act on this concern, by enhancing our understanding of what is needed to achieve sustainable development and poverty reduction and of what approaches have and have not worked.

As this volume attests, OED has much to be proud of. The department has come a long way in the 30 years since its foundation, from recording history to influencing current decisionmaking. And by identifying and bridging gaps, it has made great progress towards providing a full assessment of the Bank's development effectiveness, even though, as Robert Picciotto has often pointed out, we still lack an adequate database on some of the things that we need to measure.

With its substantial track record, I know that OED is well placed to make a central contribution to the new challenges that now face the Bank, and the development community more broadly. We are moving into an era in which everything we do must be assessed against the Millennium Development Goals. The question that will be asked now is not just "Did your project or program do satisfactorily?" but "How did it affect the achievement of the Millennium Goals?"

This question means that we in the Bank Group need to evaluate our efforts not just against our own targets and expectations, but against a larger set of expectations that have been agreed by the international community. Thus OED will need to extend the purview of its assessments: to measure achievements against the MDGs; to evaluate how the Bank's work relates to the totality of work that is being done in the development field and judge the integrating role the Bank Group is playing in the development community; and to advise on how what we're doing fits into, assists, and helps scale up the efforts towards achieving the Millennium Goals. The results from these evaluations will be vital in helping us determine how to achieve a more coordinated effort between the Bank, other multilateral and bilateral institutions, regional banks, civil society, and the private sector, all working within the frameworks that are set by governments to meet goals that are now agreed internationally.

I want to thank OED for its achievements and for the tremendous progress it has made. And I wish OED good fortune as we take on this new challenge.

Message from Former President Robert S. McNamara

On the Occasion of the OED's 30th Anniversary

Robert S. McNamara
President, World Bank, 1968-81

All large organizations, and all administrators, managers, and professionals, if they're honest with themselves, will admit that they have successes and failures. I've been an administrator for something on the order of 60 years, and God knows I've had failures, though also I hope some successes. We should learn from these experiences of success and failure. It was for that reason that I set up the Operations Evaluation function in the Bank something on the order of 30 years ago. In particular, it's very, very difficult for any of us who take pride in our activities to admit failure, and it's very difficult, therefore, for us to examine the causes of it. That can best be done by an independent organization, and that was the structure that we provided for Operations Evaluation in the World Bank.

We set up OED with professional people, well qualified to examine the activities of their peers and their colleagues in the World Bank, but with no line of authority from me or anyone under me to direct them in what to say. The evaluation reports were to be sent straight from OED, directly to the World Bank's Board of Directors, without my review, without my criticism or comment or change. The Board of Directors was encouraged to examine each report, debate it, and then transmit it to their governments. In effect, OED's reports were distributed to thousands of people across the world.

Certainly the Bank has had its failures. But the accumulated record from Operations Evaluation shows that on balance, the Bank has been successful, and that it has learned and is continuing to learn from its failures. Today OED has an extraordinarily valuable databank analyzing the causes of development successes and failures. Taking account of these lessons will, I believe, increase the Bank's rate of success for the future. That should be the continuing objective of Operations Evaluation. I congratulate you all.

Introduction

The story of independent evaluation in the World Bank is more than 30 years in the making. But it is a story that has never been told, at least not fully. That is why during the course of 2002-03 the Bank's Operations Evaluation Department (OED) held a series of major events, including seminars and workshops, leading up to the thirtieth anniversary of the founding of OED by Robert McNamara on July 1, 1973. This volume captures the history of OED as related by many of those who participated in making that history. It is divided into four sections.

The first part is drawn from a series of presentations made by former directors-general, Operations Evaluation, and senior OED managers. It relates, in roughly chronological presentations, the development of OED from an experimental program to a major department with a record of accomplishment and influence within the Bank and beyond.

The second part presents perspectives on OED from two of its most important internal constituencies—the Board of Executive Directors and senior Bank management. Part three considers OED's relationships with its external constituents, including its work on developing evaluation capacity among client countries and its evaluation partnerships with other development organizations. Finally, part four presents a forward-looking appraisal of the challenges of development evaluation from three of its leading thinkers.

Acknowledgments

These proceedings have been prepared on the occasion of the thirtieth anniversary of the establishment of OED. In addition to the authors and contributors, the preparation of this volume benefited from extensive contributions by the World Bank Group Archives. James M. Huttlinger (Archives) provided invaluable information and archived documents on the establishment and functioning of OED operations and activities.

Part of the proceedings are drawn from two major commemorative events that took place in 2002. The first event, OED—The First 30 Years, held on September 23, 2002, brought together over 300 participants, including OED alumni, the Bank's evaluation and quality assurance family, current and former CODE members, and international financial institution evaluation partners. The second event, a Panel Discussion on Development Evaluation, was held on November 6, 2002. Carole Smith offered exceptional guidance and advice in preparing the agenda and organizing the September event. Additional helpful support for logistics and displays was provided by John D. Snyder. From the Bank, Tracy L. Farrall provided assistance with conference facilities for both events. Special thanks also go to Michele Iannacci, who served as the photographer at the September event.

Extensive support and contributions were made by many OED staff members. Patricia Laverley contributed to all phases of this project. Julius Gwyer provided support in managing and coordinating the production of the documentary, and the poster that is part of this volume. Pierre-Joseph Kingbo organized exhibitions of OED publications at the events. Akiko Sawamoto worked with the Archives to create Annex C. The design, editing, and production of the event programs and OED's history working papers were directed by the Outreach and Dissemination Unit. The unit consisted of Elizabeth Campbell-Pagé (task manager), Caroline K. McEuen (editor), Juicy Zareen Qureishi-Huq (administrative support and assistance), and Esther F. Petrilli (graphics and layout). Betty F. A. Bain and Yvonne Playfair-Scott assisted in acquiring OED's institutional documents and locating OED alumni. Others who provided valuable assistance and input include Princess E. Moore-Lewis and Lydia Ndebele for administrative support and assistance, and Alex H. Mckenzie, Yang Soo Yoon, and Maria J. Mar for computer and technical support. Caroline McEuen managed the editing, design, and production of this book.

We thank Daphna Kalman for her work in designing the poster, and Deborah W. Campos for providing the photographs of former OED executives.

A video documentary chronicling the last 30 years of OED's work was produced in coordination with the Media Division of the World Bank. OED would like to acknowledge Francis James Dobbs for his advice and guidance, Dora C. Tzochevska for her work in shooting and editing the documentary, and Michael Kessler for his input and assistance.

PART I

Perspectives on the Evolution of OED

From its inception, OED has been an ongoing experiment. Building on some precedents during the late 1960s, Bank President Robert McNamara established an evaluation function in 1971 that became the basis for OED. The hope was to build an evaluation system that would act as a simple scorecard measuring Bank accomplishments, provide deeper feedback to improve organizational learning, and develop a more profound assessment of the Bank's contribution to development as a way of improving performance on the ground.

The papers in this section, based on a series of presentations delivered during 2002, provide a first-hand account of this experiment in action by former directors-general, OED directors, and OED managers. They make clear that the original intent of the evaluation effort has been maintained, even though there have been a number of changes in OED's organization, products, and evaluation processes.

Two noteworthy trends emerge from a consideration of OED's evolution. The first is the growing independence of the evaluation function. Starting out as a program within the budgeting department in 1971, by 1973 OED was a full-fledged Bank department, and by 1975 had achieved considerable independence through the appointment of the first director-general, Operations Evaluation, reporting directly to the Board of Executive Directors. Over the years since, OED's independence has grown as many of the administrative ties to the Bank's management have been severed, and, more important, as OED has demonstrated its behavioral independence through hard-hitting, critical reports.

The second trend is the growing complexity of the evaluation function. As the Bank's work has become more diverse, the tasks of evaluation have had to follow suit. At the same time, as the Bank has opened up more to public scrutiny, OED has had to find new ways to communicate its evaluation findings not only to the Board and Bank staff, but also to a much broader public, including Bank clients and partners, non-governmental organizations, civil society representatives, and not least the beneficiaries of Bank projects and programs. All this has required OED to develop new or adapt old evaluation methods, create new product lines, and bring on board evaluators with a wider variety of skills.

The authors of these papers tell this story from the inside, as they experienced it. Additional observations culled from interviews with OED staff are included in Annex A.

1

First Experiments in Operations Evaluation: Roots, Hopes, and Gaps

Christopher Willoughby
Successively Unit Chief, Division Chief, and
Department Director for Operations Evaluation, 1970–76

Ex post evaluation of the Bank's activities was not an entirely new concept in September 1970 when Robert McNamara instructed the Programming and Budgeting Department (P&B) to evaluate "the contribution of the Bank's operations to the development of member countries." Without dignifying their exchanges and memos with the formal title of evaluations, the Bank's operating departments had learned vastly from their own extensive experience of project implementation. A very stimulating study of the larger potential effects that projects may have (and may need to have for their own effective operation) on their countries' managerial practices, bureaucratic behaviors, and decision-making capacities had been published by Albert Hirschman in 1967, on the basis of thorough examination of eleven of the earlier projects that the Bank had helped finance.[1] The Bank itself had published in 1969 a thorough economic reappraisal, by its Economics Department, of a major road project completed in Iran in 1964.[2] But Mr. McNamara's decision elevated evaluation from a topic for occasional formal research and frequent ad hoc analysis into a technique for regular application, to identify lessons that would improve future Bank policies and procedures.

So began a period of intensive experimentation that was to benefit from suggestions and reactions from many sides. It took nearly five years to create an evaluation system with a structure that seemed commensurate with the complexity of the Bank as it then was. I try to trace here the travails of our gradual evolution over those five years, to give due recognition to the main formative influences, and to identify, like any retrospective evaluator, a few ways in which we might have done better.

So began a period of intensive experimentation.

P&B, established soon after Mr. McNamara's arrival at the Bank in 1968, was a main instrument of the more active management style he introduced. Led by John Adler, P&B had developed the elements of program budgeting for the Bank. So-called country program papers (the forerunners of today's country assistance strategies) had been introduced in 1969. Monthly reviews of the progress of operations were an important activity. P&B was intimately involved in the introduction of staff timesheets, to secure a more accurate picture of what things really cost. Dominant concerns were the effective delivery of the enlarged and more strategically focused lending to which Mr. McNamara had committed the Bank and, more generally, an increase in the quantity and quality of output relative to staff input. From the earliest stages in developing a system for more strategic management of the organization, it was recognized that once the foundations had been laid, feedback loops would need to be introduced, so that plans could be revised on the basis of actual outcomes. By early 1970, P&B papers were referring explicitly to "systematic evaluation and reassessment of our past operations—projects financed, lending conditions imposed, economic reports produced, supervision undertaken, etc." as tools that could help improve policies and staff time allocations, thus increasing the Bank's efficiency.

Management sought reassurance as to how far these expectations were being achieved in reality.

In addition to such lesson learning, Bank management also needed information to help convince the financial markets regarding the quality of the loans the Bank made. Expected economic rates of return, as calculated in appraisal reports, were meaningful to investment bankers and bondholders, and these numbers were compiled and cited, but management sought some reassurance as to how far these expectations were being achieved in reality.

The task assigned to P&B was thus to create a new internal management tool, a confidential and very objective information flow about how the Bank was really doing in the field. Recognizing the fears that could arise—for instance, that the aim was not only to identify "mistakes" but to pinpoint personal responsibility for them—Mr. McNamara gave great attention in the early days to nurturing the new function, discussing possible approaches, and encouraging support for the work from his associates. But he was always at pains to make clear that P&B should reach its own conclusions about whatever it examined, on the basis of objective analysis not influenced by his own views or those of other Bank staff members.

First Initiatives

By early 1971, it had been agreed that the pilot phase of evaluation should consist of two fairly ambitious but very different studies—a deep one of everything the Bank had done in one country, and a wide one covering loans to a broad spread of borrowers in one sector. Colombia was the country, and electric power the sector. Key Bank staff responsible for these areas accepted P&B's selection with grace and, despite the inevitable periodic controversies, patiently maintained cooperation and support through long succeeding months of P&B importuning them for advice, documentation, and comments. A small staff of two Bank economists, three graduating young professionals, and several research assistants was gradually brought together to lead the work.

Colombia was uniquely suitable because of the extent of the Bank's involvement in its development efforts. A first loan had been made in 1949, immediately followed by the fielding of a general survey mission whose report the Bank published.[3] A substantial flow of sectoral technical assistance had been maintained, and a wide variety of projects financed. Since the mid-1960s the Bank had been negotiating an annual memorandum of understanding with the government on macroeconomic policies as a framework for its own lending and that of other members of the consultative group it had set up for the country. Colombia had thus been the testing ground for numerous new Bank initiatives. The government graciously consented to being the guinea pig for the new experiment, and government agencies provided much technical support, as did Colombian research institutes and individual consultants.

A crucial issue for P&B in undertaking a country study covering a long period of time was to define the "development" to which the Bank was supposedly contributing. Since the purpose of the exercise was to identify lessons useful for the future, P&B decided to adopt what was considered a modern definition of development, consistent with the rising emphasis that Mr. McNamara was giving to poverty reduction as the central development challenge and *raison d'être* of the Bank. Thus development was characterized as "movement of the whole social system in such a way as to provide increasing opportunities to a growing proportion of the population of the country to realize more fully their mental and physical capabilities."

Measured against this yardstick, the Colombia evaluation concluded that the Bank had greatly contributed to opening new possibilities for the country by its substantial lending for trunk infrastructure (mainly roads and electricity, with almost all projects in those fields showing satisfactory rates of return) and by its advice on macroeconomic policy.[4] Management in these areas had substantially improved, and this had benefited, besides those already better off, at least a significant proportion of the large numbers of people who had moved from the countryside to the major cities.

But the Bank's work had thus far made little contribution to resolving Colombia's major problem of severe rural poverty, which directly affected more than a quarter of the national population and had been highlighted in the general survey mission report. For that, the institution-building and policy advice techniques needed to be applied in proj-

Closer attention was needed to national policies affecting use of the loan funds.

ects and programs that would directly address the country's weaknesses in mobilizing its own resources and its people. Closer attention was needed to national policies affecting use of the loan funds channeled into private manufacturing and agricultural investments.

To orient its operations, the Bank needed to develop and maintain a longer-run strategic framework showing how the resources available could be combined, to enable the problem of rural poverty to be overcome within a generation. More aggressive pursuit of land reform seemed to be a much-needed component and, if that was politically impossible, clear alternative policy options had to be defined to raise employment and productivity.

The electric power study showed that it was possible to cover a large volume of lending—some $1.5 billion, or 40 percent of the Bank's cumulative lending to developing countries for electric power—and still reach useful conclusions and suggestions for future improvements.[5] While obviously less deep than the Colombia study in consideration of possible alternatives to what had actually been done, it could provide reasonable reassurance on the main questions, especially the efficacy of the institution-building efforts and the economic validity of the plant choices made. It was also able to explore some other questions about Bank interventions, such as the cost effectiveness of withholding loan commitments pending borrower/government reform actions.

A Second Round of Experimentation

By early 1972, it was clear that such major evaluation studies took a long time. Quite independent of the initiative Mr. McNamara had taken, the U.S. Congress and its audit arm, the U.S. General Accounting Office (GAO), began to press for evaluations from the Bank, or to undertake them on its own. Mr. McNamara decided to release the two studies to the Bank's Board, where they received universal acclaim, except from the Executive Director representing Colombia. He felt, in many ways quite understandably, that it was unfair, and politically oversimplified, to assess past actions against the Bank's modern definition of development objectives, and perhaps he also deplored the fact that his government had not been given an opportunity to review the draft of the overall summary report before its was finalized.

Some important conceptual distinctions were becoming clearer in the minds of the staff, in light of their experience and discussions with Bank colleagues and GAO staff. The basic criteria appropriate for assessing Great Society programs in the United States could be used with interpretations directly related to the Bank: effectiveness (extent to which objectives fulfilled), efficiency (cost/benefit analysis), and economy (reasonableness of unit costs). But these three dimensions needed to be looked at for three different items: the physical project, institution-building aspects, and use of the Bank's administrative budget (the last-named covering most of what is included, in OED's modern terminology, under the rubric of a project's "relevance" as well as "Bank performance"). This could be done either in a relatively simple way, covering the extent to which objectives stated in appraisal reports had been fulfilled, or in a variety of deeper ways, reassessing the objectives against possible alternatives that might have done more for development. The term "performance audit" was introduced for the first type of study, and "evaluation" was increasingly reserved for studies of the latter type.

An important goal of the second round of experimentation was therefore to try out the concept of individual project performance audits, applying it to a variety of projects whose construction had been completed in earlier years. The bulk of the staff power available was devoted to three sectoral evaluations, covering fewer cases in greater depth than had been done in the electric power study. Work was also to start on an aspect that Mr. McNamara had always been greatly concerned about: review of operating department actions in response to the recommendations of the Colombia and electric power reports.

An important goal was to try out the concept of individual project performance audits.

Firmer Attachment to the Bank's Board of Directors

While the second round of experiments was under way, intensive discussions continued with the U.S. government regarding the constitutional position of the operations evaluation service within the Bank. It was eventually decided that, to reduce the risk of conflict of interest for Bank management—or even the appearance of potential conflict—operations evaluation should be removed from P&B and set up as a separate unit under the charge of a vice president who would have no operational responsibilities, but would also supervise the Internal Audit Department. And the operations evaluation work program should be approved by the Executive Directors' Joint Audit Committee, which had been created in 1970 to supervise Bank and International Finance Corporation (IFC) internal and external audit arrangements.

The Operations Evaluation Department (OED) was born on July 1, 1973, taking over the staff who had previously been working on the subject in P&B, and came under the supervision of Mr. Mohamed Shoaib, a Vice President of the Bank who had long been Finance Minister of Pakistan and had a deep interest in management auditing.

Another important development in the early months of 1973 was the issuance of an instruction by the Bank's vice president, Projects, that the operating departments should prepare project completion reports (PCRs)—today, more accurately called implementation completion reports—within one year of the completion of loan disbursements, on all projects for which disbursements were completed after July 1 1972. (A few such reports had already been prepared on earlier projects, at the initiative of individual staff and departments.) Assisted by this decision, as well as by its own initial experience in preparing audit reports on individual projects, the new department committed itself to issuing project performance audit reports (PPARs) on all loans and credits within a year or two of their closure. This was in line with Mr. McNamara's strong feeling that such a standard was essential, both to fulfill the Bank's accountability responsibilities and to provide a useful gauge of performance, recognizing that profits were a much less meaningful indicator in the Bank than in most stock corporations.

The new department committed itself to issuing project performance audit reports on all loans and credits.

An important innovation in operations evaluation's third round of experiments, which were the first under Joint Audit Committee surveillance, was an entirely new kind of study focusing on the Bank's own procedures and policies, with the history studied not limited to completed projects approved many years before. Issues such as delays in loan effectiveness and delays in project preparation, for example, could clearly be studied from a sample of more recent cases, with the hope of reaching useful suggestions for their reduction. Christened "management policy reviews" and very much inspired by Mr. Shoaib, these studies could deal with aspects of the Bank's effectiveness that might be of particular concern to the developing member countries. They were, perhaps, the forerunner of today's process reviews.

By the end of 1974, OED had issued no more than about three dozen reports (most of them PPARs). But a major concern was to find ways of making these reports more usefully accessible, so that a Bank staff member could quickly identify what past lending experience indicated about any issue that arose. The need for a solution was heightened by the efforts Bank management was already making to ensure that any relevant past experience would be reflected in proposals regarding new loans or policies. Given the technology of the day, the best we were able to come up with was yet another report, a "Concordance" that listed all issues that had been touched on and cited the specific references. The first such Concordance was issued in 1974.

At Length, an Emerging System

Continuing discussion with the GAO and Board members about the constitutional position of OED led to agreement late in 1974 that, upon the forthcoming retirement of Mr.

Shoaib, OED's independence should be further institutionalized by placing it under the sole charge of a director-general, Operations Evaluation, with rank of vice president, selected by the executive directors from a short list provided by the president, removable only by the executive directors, and ineligible for reappointment to the Bank Group staff except in unusual circumstances. The director-general would normally hold office in five-year terms, like the president of the Bank (but not of course coterminous). OED would report, through the director-general, to the executive directors, but retain an administrative link (reflected by a dotted line on the organization chart) to the president.

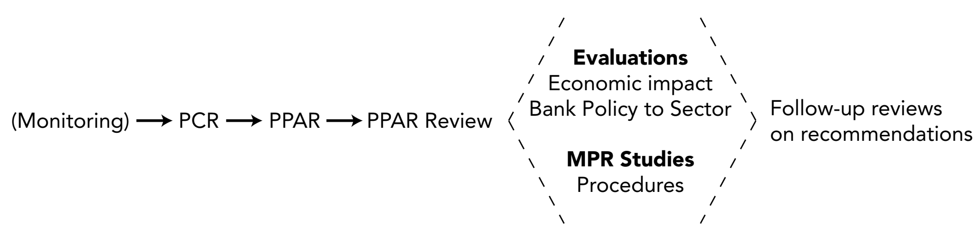
Thus, from all the experimentation, by 1975 there were beginning to emerge the main elements of a system of evaluation for the Bank that could be handed over as a starting point for the director-general. Figure 1.1 presents these elements as an extension to the Bank's then-standard project cycle.[6]

PCRs were coming forward from the operating departments by 1975 in increasing numbers and their quality was improving. Thus more of them could be attached to the PPARs, which in turn could ideally confine themselves to an attestation of apparent quality, on the basis of limited file research, and some elaboration of points of particular interest, especially connected with Bank performance.

The PPAR Review, which was expected to be an annual report following its first issuance in September 1975,[7] would summarize results of all the PPARs produced since the last review, provide a relatively simple, comprehensive scorecard of the performance of completed Bank-assisted projects, and highlight important lessons emerging from comparative review of all the PPARs.

Trends or problems identified in the PPAR Review would be an important source of ideas, themes, and relevant cases for the more evaluative work that would remain OED's

Figure 1.1. Main Elements of a Possible Evaluation System, as Envisaged in 1975



principal occupation. Sectoral and subsectoral reviews, based on deep study of a few cases within the context of the general Bank policy regarding the sector/subsector, seemed to be the most productive of useful lessons with a sufficiently broad foundation to be actionable Bankwide. Country studies were also expected to be useful, but after some time, if and as the Bank deepened its country programming and increased the macroeconomic aspects of its advice and loan conditionality. Management policy reviews seemed, from the initial work, to be very promising ways of dealing with Bankwide procedural and process issues.[8]

The final element perceived to be essential for a sound system was follow-up reviews. These assessed the progress that operating departments had made on the actions they had agreed to undertake in response to the recommendations of major studies. The purpose of these reviews was to help ensure that evaluation findings were not forgotten and that the evaluation effort would ultimately earn its keep in terms of operational improvements assisted. The first two such reviews were distributed to the Board in spring 1975, reporting on actions taken on the main recommendations in the Colombia and electric power reports. They were the first steps toward OED's modern Management Action Record.

The final element perceived to be essential for a sound system was follow-up reviews.

Hopes

The evaluation system was expected to bear fruit in essentially three forms:

1. A somewhat oversimplified but comprehensive and objective running report card on what Bank/IDA lending was actually accomplishing that would reassure the Bank's funding sources and help identify areas where improvements were needed.
2. A variety of deeper feedback analyses that would accelerate and improve learning from experience and outside criticism, and spread the learning effectively, thereby enabling the Bank to be less risk-averse in selecting projects, as required to improve its contribution to reducing poverty.
3. Particularly from the latter type of studies, a more profound assessment of the Bank's contribution to development that would increasingly supplement and strengthen the report card and facilitate further improvements on the ground.

Our first hope was, therefore, that we, along with the early writers of project completion reports, had established a tradition of thoughtful and self-critical enquiry into past operations, not hampered by the fears prevalent at the outset of our work but, on the contrary, conducive to lesson learning. This was essential to the integrity and utility of the system, for there were limits to the cross-checking that OED would be able to do with the rapidly increasing number of loans completing disbursement.

One important, though obviously not foolproof, incentive to objectivity was OED's solicitation of comments on draft reports from all in the Bank who might be particularly interested in the project and from the borrower. Drafts were widely circulated for comment within the Bank, not only for this reason but also because this was one of the better ways to reduce the constant danger that evaluation products would not get far enough up in people's in-boxes actually to be read. Though the Colombia report had only been reviewed with the borrower to a limited extent, because of its novel nature, drafts of almost all subsequent reports were discussed with, or sent to, the borrowers and governments concerned. It became standard practice to include their written comments as an annex to the final report.

The operating departments also increasingly asked borrowers to prepare their own PCRs. Sometimes this produced excellent documents, but most countries had too few appropriate staff to do more than provide essential data, discuss the evolution of the project, and comment on the Bank/OED drafts.

An important early emphasis of Mervyn Weiner, after he became the first director-general, Operations Evaluation, in October 1975, was to systematize these processes, with a much heavier accent than before on the self-evaluative aspect of the PCRs. The intention was to enhance lesson learning within the operating departments, and also to create a more efficient division of labor between the departments and OED—which was expected to be able to ratify most of the PCRs produced without intensive work, although it retained the right to go into detail on any project that seemed to need more thorough assessment. Attention also began to be given to assisting the development of self-evaluation capacities within the governments of interested member countries. A first formal statement of the Bank's Standards and Procedures for Operations Evaluation, covering in particular the content of PCRs and PPARs and processes for their preparation and review, was submitted to the executive directors and approved in April 1976, and then widely disseminated.[9]

Performance audits and evaluation reports that had been produced by the end of 1975 indicated quite impressive overall economic results. They had covered a combined total of 170 projects, or about 20 percent of the projects completed with Bank/IDA financial assistance up to that time. More than 90 percent of this large sample appeared likely to yield economic returns above the 10 percent minimum sought by the Bank, or an approximate qualitative equivalent in sectors where returns could not be soundly quantified. On the more than $10 billion investment involved in all 170 projects, overall estimates of the average economic return suggested a figure of around 15 percent.

But a cautionary note was also in order. A few of the projects, and in particular six on which audit reports had been issued in 1975, showed satisfactory returns only because deficiencies in the forecasts for costs, crop yields, and so on had been offset by unexpected increases in the value of project output, resulting essentially from OPEC's 1973 oil price increases. That event appeared to have redeemed what might otherwise have proved poor investment decisions by the Bank in these six cases.

This was one of the reasons why it seemed so essential to us to focus, even in project performance audits, as much on the performance of the Bank as on the performance of the project and its implementation by the borrower. Other reasons were the perennial doubts that arise as to how far the funds being provided by the Bank were releasing resources for quite other expenditures than the project ostensibly financed; the widespread evidence that the aspect on which outside help was most needed, especially in the case of poverty-reducing development, was the weakness of institutions and broad policies; and, as mentioned at the outset, prior preoccupation with the efficiency of use of Bank staff resources.

We were always trying to answer the question, "What difference did the Bank make?"

We were always trying to answer the question, "What difference did the Bank make?" And we were looking forward to the days when improved time-recording in the Bank would generate more accurate information about Bank staff expenditures on different aspects of a project and enable more precise analysis of the relationship between the differences the Bank made and the staff expenditures involved. That would be an important complement to project rate of return analysis.

Gaps

Many of our reports identified a strong need for a carefully selected set of indicators capturing project/program performance and changes in the economic environment, as a key means to improve management and facilitate Bank supervision. We urged identification and targeting of such outcome indicators at appraisal, and establishment of monitoring systems that would show what was really happening. We tried to help by proposing illustrative sets of indicators for different types of projects. In 1974 the idea got some support in a Bankwide staff instruction, and a unit was formed in operations to promote such efforts, especially for rural development projects, where the job was most difficult. However, when our follow-up reviews on both the Colombia and electric power studies indicated disappointing progress specifically, and almost solely, on the development and use of project monitoring systems, we made what was probably a serious mistake by treating this as a relatively minor, bureaucratic matter (as compared with most of the other, more substantive recommendations) instead of trying to reinforce our support.

Since we were stimulated by Mr. McNamara to devote much thought to the problems of poverty and since effective poverty reduction projects have proved so difficult to design,[10] we should probably have given more attention to specific audit and evaluation arrangements for projects and parts of projects that had this purpose. Instead, especially for projects in countries' backward areas, we strongly promoted "maximizing development impact" by complementary investments and initiatives. Many cases we had studied had shown the opportunity and need for such initiatives. However, in some cases where complementary initiatives had been undertaken, we had found their environmental and social side effects to be negative. We probably bear a bit of the responsibility for the practice, in the later 1970s, of sometimes incorporating multiple complementary investments in project design without sufficient attention to beneficiary views and local institutional capacities, and some of the resulting projects proved impossible to implement, or were at the least unsustainable.

Had we instead promoted more analysis of what experience showed about the relative importance of different complementary investments in different situations, ways of organizing them,[11] and the costs involved—both to the country, and in terms of Bank staff time—we would have come closer to helping solve a still outstanding problem. Only in very recent years has much attention been given, for instance, to assessing the impact of infrastructure projects specifically on poor people as opposed to poor areas.

Conclusions

We had expected that OED's work would enable the Bank to be more aggressive in testing new approaches by systematically analyzing actual results and facilitating quick adjustment or abandonment. Perhaps we were unrealistic. Skeptics would point to the usually long time required for disbursing loans, clarifying results, and generating and disseminating studies—by the end of this time large volumes of similar loans might have been committed. Our experience indeed suggested that the most productive use of OED was in assessing the results of work already done, rather than trying to second-guess the decisions of those responsible for current work (except, to a degree, in follow-up reviews). But this did not mean that our attention had to be limited solely to activities fully "completed"; a study could focus, for instance, on one or two carefully selected forerunner projects plus the early stages of more recent projects. Where applied, the improved monitoring systems mentioned above facilitate the effective coverage of the latter.

Certainly in some periods since 1975, and perhaps in most, the World Bank has moved into and out of different types of activity faster than it did before 1970. And the broad evidence is that OED studies played a part, along with research studies, policy papers, and the continuous absorption of experience by Bank management and staff, in bringing about these changes.[13] Sometimes, however, the changes may have been slower than they could or should have been,[14] or faster than what in retrospect seemed wise. OED's contribution would probably be most important in connection with the abandonment of (or major adjustments in) past approaches. It might be useful for OED to consider evaluating the process of policy change in the Bank by reviewing a sample of cases of this type, in part to identify whether and how it could improve its contribution to such an important process.

Notes

The author is very grateful to former colleagues for comments on an earlier draft of this paper, particularly John Blaxall and Robert Picciotto.

1. Albert O. Hirschman, *Development Projects Observed* (Washington, DC: Brookings Institution, 1967). The high utility, for evaluation work, of concepts introduced in this book and other Hirschman works is very interestingly discussed, and illustrated from the Bank's later experience, in Robert Picciotto, "Visibility and Disappointment: The New Role of Development Evaluation," in Lloyd Rodwin and Donald A. Schon, eds., *Rethinking the Development Experience, Essays Provoked by the Work of Albert O. Hirschman* (Washington, DC: Brookings Institution, 1994).

2. Herman G. van der Tak & Jan de Weille, *Reappraisal of a Road Project in Iran,* Staff Occasional Paper No. 7 (Washington, DC: World Bank, 1969).

3. World Bank, *The Basis of a Development Program for Colombia* (Washington, DC: World Bank, 1950). Report of a mission headed by Laughlin Currie and sponsored by the International Bank for Reconstruction and Development in collaboration with the Government of Colombia.

4. World Bank, *Bank Operations in Colombia: An Evaluation,* Report No. Z–18 (Washington, DC: World Bank, 1972).

5. World Bank, *Operations Evaluation Report: Electric Power,* Report No. Z–17 (Washington, DC: World Bank, 1972).

6. Warren C. Baum, "The Project Cycle," *Finance and Development* Vol. 7, No. 2 (June).

7. World Bank, *First Annual Review of Project Performance Audit Results,* Report No. 871 (Washington, DC: World Bank, 1975). This report covered 50 projects. The second in

the series, Report No. 1227, was issued June 30 1976 and covered 57 projects. Both reports fell short, however, of the intended standard of covering all projects within two years of completion of disbursements for them.

8. The first MPR study, *Delays in Loan and Credit Effectiveness,* was distributed to the executive directors in July 1975, and the second one, *Delays in Project Preparation,* in February 1976.

9. World Bank, *Operations Evaluation: World Bank Standards and Procedures* (Washington, DC: World Bank, 1976).

10. See, for example, OED, *Poverty Reduction in the 1990s: An Evaluation of Strategy and Performance* (Washington, DC: World Bank, 2000).

11. The late Arturo Israel, who was our invaluable principal collaborator through all the early years, moved on to other work but produced in the 1980s a very thoughtful, practical book on many of these issues: Arturo Israel, *Institutional Development: Incentives to Performance* (Baltimore: Johns Hopkins University Press, 1987). Another example of pin-pointed issue-focused monitoring and evaluation that could help accelerate refinement of the Bank's advice in difficult areas would be follow-up on cases where his suggestions were applied to projects addressed mainly to poor people.

12. Jocelyn A. Songco, *Do Rural Infrastructure Investments Benefit the Poor?* Working Paper No. 2796 (Washington, DC: World Bank, 2002); and the reports produced under the Asian Development Bank's major research project, RETA 5947: *Assessing the Impact of Transport and Energy Infrastructure on Poverty Reduction.*

13. World Bank, *The Role and Effectiveness of Development Assistance: Lessons from World Bank Experience* (Washington, DC: World Bank, 2002). A research paper from the Development Economics Vice Presidency of the World Bank.

14. William Easterly, *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics* (Cambridge, Mass.: MIT Press).

2

Institutionalizing the Evaluation Function at the World Bank, 1975–84

Mervyn L. Weiner
First Director-General, Operations Evaluation, 1975–84

Today is the Bank's birth date. On June 25, 1946, the Bank opened its doors for business. This happens to be my Bank anniversary date, too. It was on June 25, 1951, that I began what turned out to be a wonderful 33 years here. That perspective induces me to reflect briefly on the changing context in which OED has always functioned.

The Changing Context for Evaluation

The inception of formal evaluation in the early 1970s and the establishment of the Office of DGO in 1975 mark roughly the midpoint in the Bank's history to date. These have been years of virtually non-stop change. A few markers come to mind.

Soon after I joined the Bank I had the opportunity to attend a Board meeting that, among other things, considered whether the Bank should begin lending for municipal water projects. They decided not to, perhaps because an important executive director insisted that financing the provision of drinking water for people would be financing consumption rather than productive investment, and the Bank was an investment bank. Financing water for cattle, of course, would be productive and therefore OK!

Similarly, when the Bank later began lending for education, as I recall, this was limited to vocational education. Clearly, these past 56 years have been years of continuing evolution in the Bank's understanding of its development role.

Critics have complained about the incessant addition of special initiatives, which are said to have diluted the focus and coherence of Bank operations. True as that may be, another way of seeing these changes is as the Bank's evolving response to its accumulating experience and the broadening expectations of its members. Against the evolution of the past 56 years, the Comprehensive Development Framework could be seen to mark the closing an evolutionary circle. Past discrete special initiatives, along with new ones like governance and transparency, are now perceived as integral parts of one interconnected whole.

Of course, what the Bank does in country X in year Y will remain particular as to time and place. But I have the impression that the bean-counting approach to the portfolio profile is no longer operative. Virtually anything is now eligible, so long as it is appropriate as to time and place, well justified, and well done.

Origins of Evaluation in the Bank

At a Board meeting early in the Bank's agriculture and rural development initiative, an executive director wondered how the Board would know whether these new projects turned out after implementation to be worthwhile. That could have been the moment when Mr. McNamara first decided to establish an independent evaluation unit in the Bank. I doubt it, because my guess is that such a function was part of his idea of good management. He was, after all, the one who caused the Pentagon Papers to be written. But whatever the facts, it was not long after Mr. McNamara became president that Chris Willoughby was asked to establish the first operations evaluation unit in the Bank.

The bean-counting approach to the portfolio profile is no longer operative.

To ensure its independence, the new unit first reported to the Bank's budget director and, later, to the vice president in charge of administration. In the four to five years before the Office of the DGO was created, the unit produced path-breaking work that defined the basic evaluation products OED has produced ever since—performance audit reports and impact reports on individual projects, sector and country reviews of groups of projects, and process reviews—all with a small, young staff.

Forging the Relationship Between Evaluation and Operations

I first encountered OED as a projects director and found the variety and quality of their work impressive. But what I recall most were the reactions of colleagues who had been asked to comment on draft reports concerning operations in which they had been directly involved. They were deeply bothered by the way differences in views were handled.

That there were differences is not surprising. Multiple observers inevitably have differing perspectives, especially when their views are shaped by varying experience. OED's staff and consultants at the time had little experience with Bank operations or direct knowledge of their history and context. So staff who had been involved in these operations often challenged an OED observation. But they found that while some comments were accepted, those that were not accepted were simply disregarded in the final reports to the Board.

This absence of a countervailing operational voice in Board reporting was not appreciated! From where I sat, the resulting friction undercut the feedback benefits of OED's good work. That was my major concern when the president asked me to take on the new job of DGO.

Chris continued as OED director for a year, pursuing the excellent work he had begun. OED's link to the Board by this time was well established. But how might the OED-operational relationship be improved?

I soon concluded that it could only be improved by doing two things:

- Making operational staff an integral part of the evaluation process rather than having them just contribute to OED reports
- Giving them a protected voice in the Board alongside OED's.

The Bank had somehow to become a self-evaluating organization, with OED playing the key—but not the only—role in the evaluation process.

In fact, I always perceived OED as part of a self-evaluating Bank, and not as the sole actor on the evaluation scene.

ANNUAL REPORTING

To anchor the institutionalizing process, two annual reports were introduced. One was a review of the findings of all performance audit reports issued during the year. The other was a report on the status of evaluation in the Bank. The main focus of this second report was on OED's work program, but it also reported on the status of self-evaluation in the International Finance Corporation, the Economic Development Institute, and the Research Complex. OED did not evaluate what these other units did, but by reporting on their activities, it sought to encourage evaluation there, too.

Evolution of Completion Reporting

The early efforts concentrated on completion reporting. This was, after all, to be the foundation of the Bank's evaluation system, the basic evidence on which country, sector, impact, and process studies would be based.

The obligation of operational units to produce completion reports on all projects had recently been established. But confronted with scheduling and budget pressures to meet lending programs, it is not surprising that their performance was spotty. Some excellent completion reports were produced, and some weren't produced at all. Some turned out to be inadequate for submission to the Board, being little more than unfiltered compilations of supervision documents that just reported the implementation history of the project, and were completely insensitive to the cost and audience for these reports. Some were as much as 50–60 pages long; others reflected serious deficiencies in the underlying supervision reporting.

Without timely and comprehensive completion reporting by operational staff, the longer-term vision of completion reporting providing foundation evidence for larger OED studies was at risk.

Without timely and comprehensive completion reporting by operational staff, the longer-term vision of completion reporting providing foundation evidence for larger OED studies was at risk.

OED could not long continue to bring deficient completion reports to Board standard and also evaluate a growing Bank portfolio by itself. Operational views of completed operations had to be assured if the hoped-for evaluation benefits for the Bank were to be realized.

What happened next was a meeting with projects directors over lunch. I told them that OED could no longer remain the unit of last resort for delivering acceptable completion reports to the Board. Line responsibility for project identification, preparation, appraisal, and supervision would henceforth also have to include the self-evaluation of implemented projects. OED would audit completion reports, which would provide basic accountability evidence and also evidence for OED's sector, country, impact, and process reviews. But the foundation of the evaluation system would have to be timely and acceptable completion reporting by operational staff. Remaining free to criticize OED views without first having to formulate their own was no longer acceptable.

They were also told that whenever differences in views arose, their dissent would henceforth be included in the final OED report. I would ask them to confirm that their views had been reported to their satisfaction before sending the report to the Board.

All this was subsequently codified in the Standards and Procedures document that was reviewed and approved by the Board.

It's gratifying to see how these arrangements have since evolved from strength to strength. Accountability for portfolio performance and evaluating operating experience overall became firmly institutionalized and have continued to serve staff and shareholders well.

I don't know what supervision reporting is like today, but I have the impression that deficiencies in supervision still affect completion reporting and OED studies. Ideally, supervision reporting should track monitoring and evaluation indicators defined in the appraisal report and analyze major deviations as they occur, making it possible for the final supervision report to become the operation's completion report. But that seems not to have happened yet.

I've been too long out of the Bank to suggest why this should still be a problem after 30 years. It may reflect the ethos of forward-looking, problem-solving staff who prefer leaving to others the task of reporting and reflecting on experience. It may also reflect inadequate budgeting for supervision and completion reporting.

When I reported serious lags in completion reporting to the Joint Audit Committee, the concern became public and, for a period, supervision was declared to have first claim on budgets. But lending programs still had to be delivered, generating real tensions for staff. It took a while before budgets began to make adequate provision for these conflicting claims. Whatever the explanation, I was struck when reading recent OED reports to note that concerns about monitoring and evaluation persist.

Strong Presidential Support

Mr. McNamara, in whose tenure evaluation was introduced, was its unqualified supporter. He made that very clear in everything he said and did. We would meet for lunch every few months, along with the director of the Central Project Staff, to review how things were going and alert him to things he should know or do something about. Had the president been less interested and more distant, I dare say the history we are reviewing today might have been very different. All credit to him.

Consistent Board Interest

The Board too was consistently interested in the conduct and evolution of OED's work. One could see this especially in the discussions that took place over a period of years about the portfolio coverage of the evaluation system. I referred earlier to the evaluation system embracing the whole portfolio, under the oversight of the Joint Audit Committee. When OED's first work program was presented to the Committee, one important director took exception to the planned full coverage of the portfolio. As best I can paraphrase, he said: we already know the Bank is a pretty good organization and that most of the things the Bank does are sound and will probably turn out well. So that's not what we want to hear. What we want to learn about are the problems, the failures, and how these shortfalls should be addressed. His concept of the DGO was that of an inspector general.

My response was direct. I said that if the only OED reports that the Committee ever received were about problems, that would color their image of the Bank's entire portfolio, for they would have no way of knowing whether failed projects represented one percent or 91 percent of completed operations. The discussion was extended, to the point that I was eventually forced to tell the Committee that they would have to find another DGO if they insisted on what was being asked. I would, of course, report fully about portfolio problems, but context was essential. The Committee finally agreed that full project performance audit coverage would be the rule.

Some years later, the 100 percent coverage rule was challenged again, this time by a new president for whom sample coverage made more sense. It was also challenged by the budgetary and staffing reality of OED's growing portfolio and studies program. The feasibility of projecting 100 percent coverage indefinitely into the future had become a valid question. The solution then proposed to the Committee was to reduce the performance audit coverage to 50 percent, bearing in mind that projects not covered in this way would be reviewed later by OED in special studies of groups of projects in a sector or country. That proposal created a problem for some Board members, who reported that 100 percent coverage was an important element of political support for the Bank in their governments. To reduce this to 50 percent could thus be politically awkward. But that, too, was sorted out, and 50 percent coverage was agreed.

Reporting Relationships

The president's place in the DGO's reporting channels was an early issue. Direct DGO reporting to the Board was the rule from the outset. But what was not immediately clear was the DGO's reporting link to the president. Reporting through the president would clearly be inconsistent with the independence that the Board required. But reporting only to the Board would bypass the president, who was chairman of the Board, and create potentially awkward situations were the Board to question the president about reports he had not seen before. After considerable debate, it was agreed that the DGO would report simultaneously to the Board and the president. That proved to be a satisfactory solution and, as far as I know, continues to be the arrangement today.

Staffing

I've referred to operational staffing constraints when full coverage completion reporting became the rule. OED's growing workload also created staffing problems for itself, of a different kind. What kind of staff should be sought? Where? How long should OED staff remain in place before reassignment?

Since a major intended benefit of evaluation work was feedback to operational staff, recruiting experienced operational staff for limited periods and then having them move back into operations seemed to be the most appropriate staffing policy for OED. But this proved very difficult to implement, because of the way Bank managers and staff perceived career incentives. Managers were understandably reluctant to release their best people, and only too happy to transfer their poorest performers! To dissuade staff who expressed interest in OED work for a period, some managers even threatened re-entry difficulties after their evaluation stint. There was nobody to manage staff reassignments in the best interests of the Bank.

But not all managers were obstructionist. The Bank's first review of Training and Visit Agricultural Extension criticized aspects of that experience. The author of this excellent report was clobbered by his former agricultural colleagues for the views expressed, and—for the first and only time in my experience—the central staff refused to comment on a draft report, asking instead to have it killed. I said I would withdraw the report if they offered a good reason for doing so. Otherwise, the report would be sent to the Board and, if they persisted in refusing comment, their refusal would be noted in the transmittal note to the Board. To avoid the potential embarrassment implied, the senior vice president was informed of this unusual situation and saw to it that it was soon resolved. (I mention this episode because it was so unusual. Overall, the system functioned as intended.)

To dissuade staff who expressed interest in OED work for a period, some managers even threatened re-entry difficulties.

The best part of this story is that the author of the study, whose reassignment prospects were thought to be in great jeopardy, was welcomed at his preferred reassignment option soon thereafter. Rather than fearing his candor, his new manager sought him out for his experience and professional integrity. I would like to think that episode had a positive demonstration effect on subsequent OED reassignment, but I'm not sure.

Efforts to Extend Evaluation Beyond Bank Operations

I recall vividly my first meeting with senior IFC managers on evaluation. Like that first meeting with projects directors, I was welcomed and offered all assistance except the undertaking to produce evaluations of their own. But that has changed. IFC now has its own firmly established evaluation arrangements.

My only relevant recollection about Research is the overlap in a number of Development Research Center (DRC) and OED studies. It seemed to me that overlapping studies should be planned in more complementary fashion than had been the case. I therefore raised the matter with the then-director of DRC. The reply I received after a few weeks was brief: "DRC plus OED equals zero. End of story."

Well, it was apparently not the end of story. Today, 30 years later, I see a former director of Research is now the director of OED, and former DRC staff/consultants now work for OED. I don't know whether OED and research studies still overlap, but clearly some good things can happen if one only waits long enough!

Let me now turn to other significant features of the DGO's early years.

Evaluation Criteria

The Bank's lending objectives have been in constant evolution over the past 56 years, a fact that evaluation needs constantly to bear in mind.

The Bank's lending objectives have been in constant evolution over the past 56 years.

An unusual episode offers an implicit commentary on relevance as an evaluation criterion. As I remember the story, the explicit objective of this project, a rather old one, was to increase cattle production for export. The re-lending beneficiaries of the loan were therefore big cattle ranchers. The project was successful as designed, but the consultant's otherwise excellent report found the project devastatingly wanting for its lack of poverty alleviation. Now, it would have been entirely appropriate to argue that export promotion did not address poverty alleviation and should thus no longer be eligible for financing by the Bank, although that proposition might be argued either way. But to assert years later that this old project's export promotion objective was wrong in its own time would have been difficult to justify.

Audience for Evaluation

In my time the general public was not yet an audience for OED reports. I recall discussions in the Joint Audit Committee, which then oversaw OED work, at which proposals to publish particularly interesting reports were rejected because of the perceived sensitivities involved. All that has since changed.

Outreach

I should have noted that Chris Willoughby began early to seek borrowers' views about particular operations. The thinking was that if borrowers were invited to comment routinely on all OED drafts, they might begin to see value in preparing their own completion reports, or in having local officers join OED missions, or, best of all, in establishing their own evaluation arrangements whose reports we could then begin to use.

We received comments from many borrowers, which were included in OED's reports in the same way as operational staff comments. But the larger objective was not realized, at least at the time. Some borrowers were reluctant to offer views that would be disseminated to all Bank members. Others, less inhibited, used this opportunity to blame the Bank, not always fairly, for every problem they encountered.

I pursued further the idea of borrowers evaluating at least some of their Bank projects at two regional seminars, one in Nairobi and one in Kuala Lumpur. Participation was widespread and the exchange of experience was lively. But when I reported on these initiatives later to the Board, I had to say that while the two seminars were well attended and useful, they failed in their purpose of provoking the beginnings of a local evaluation response.

OED maintained a presence at the OECD's periodic evaluation group meetings at which members' evaluation experience was exchanged. Whenever the opportunity presented itself, the DGO and the director also shared OED experience with borrower officers and auditors general to invite their participation in OED's work. But these outreach efforts produced little follow-up or result. A major reason, in my view, was that encouraging borrower evaluation was never really internalized by the operational staff who managed the Bank's day-to-day relations with borrowers.

I hope both borrower governments and Bank staff now engage more fruitfully with OED than was the case in OED's early years.

To conclude, perhaps the most important thing to be said is that OED is not just still here, but thriving. Given the tensions inherent in evaluation in any organization, even one such as the World Bank, that is an achievement to be celebrated. Let's now open for questions.

QUESTIONS AND ANSWERS

Participant: I understand that the rule is that if you accept the position of DGO, you cannot accept another position in the Bank. Where did that originate and did that apply when you accepted the position?

Mr. Weiner: Yes, it applied when I accepted the position. It originated in the debate within the Board about how this office should be set up. There was great emphasis on independence, and different national experiences were reflected in that conversation. This rule was the outcome.

Participant: You mentioned the uneven quality of supervision reports. How did you perceive the quality of the Bank's record-keeping during your tenure as director-general?

Mr. Weiner: I retired before the 1987 reorganization and those that followed, but the stories I've heard about their impact on records are appalling: some records were broken up, others just disappeared. I don't recall earlier problems of that kind. Although there are always deficiencies in the record, my recollection was that you could go to the unit responsible for a particular project and expect to find binders with the relevant records there intact.

Participant: My name is Bob Banerji, and I'm a former staff member of OED who left in 1992. A few months ago an editorial in *The Wall Street Journal* criticized OED's independence. One of the issues it raised was the fact that the appointment of the DGO is renewable. Does this compromise the independence of OED, because to some extent, the DGO becomes beholden to the president of the institution?

Mr. Weiner: Your question has implications for the integrity of the Board, for the DGO is an appointee of the Board. The president submits a short list of up to three names, but it is the Board that then appoints. In the event, the president submitted a short list of one, and after five years they reappointed me, as they have since reappointed Bob Picciotto. I don't see how this arrangement has anything to do with independence because the Board, not the president, makes the appointment. Now, if you are seeking to stick it to the Bank, then this arrangement can provide a hook on which to challenge the notion of independence, because the Bank's president is also the chairman of the Board. Experience with Boards in other institutions may well explain outsiders' skepticism about the independence of the Bank's Board. However, my experience of these arrangements was that independence was never an issue in my time.

By way of elaboration, let me tell you of a revealing incident. When an important Part I executive director, who was on the Board in 1975 when these arrangements were being discussed, was about to return home several years later, he invited me to lunch, just the two of us. I was surprised because, while I knew him, we had never had that kind of relationship. We had a very pleasant lunch, at the end of which he told me what was on his mind. He just wanted me to know that when the DGO arrangements were first made and I was appointed, he had had grave reservations about their appropriateness. The idea of the president submitting a short list of just one, and a staff member at that, while being chairman of the Board, made him uncomfortable.

In his quiet way, he hadn't said anything at the time, but he remembered these deep misgivings. What he wanted to say now that he was returning home was that his initial concern about independence had turned out to be misplaced. After observing how the arrangements functioned and how executive directors were involved, he concluded that they were just fine.

Robert Picciotto: I just want to add that at the end of my first term, somebody had been nominated to replace me, and then the Legal Department pointed out that the language of the mandate is that the DGO can only be removed by the Board. Because the position is renewable, management cannot propose a list of one at the end of the first term; removal can only be initiated by the Board. So I'm serving my second five-year term. I was also appointed on a list of one. This apparent ambiguity probably laid the groundwork for the current selection procedure, which is essentially fully controlled by the Board. There is management consultation, but essentially it's by a committee of Board members chaired by the chairman of CODE.

Participant: Along the same lines, there is the notion that the DGO had to be above the fray, not only vis-a-vis management but to play a balancing act. He had to protect the management from bad evaluation. That was the reason why I thought you had the DGO and then you had the director. The director was staff, and he had the management of the department, but the DGO had to be above the fray. So from an independence viewpoint, he had to play a tricky balancing act. Is that correct?

Mr. Weiner: The short answer is I don't know. The balancing issue is there, as we've been discussing, but I don't think it was conceived that way. My guess is, like most of these things, the explanation is simpler. The department with a director existed before the idea of DGO was ever thought of, and the DGO was superimposed on that existing situation. That's my explanation.

Participant: You emphasized the learning process from evaluation. Are there some examples of changes in policies and programs that we can attribute to evaluation results? Second, you didn't make any reference to accountability. Was this also an issue of concern during your period?

Mr. Weiner: My own view is that accountability came first, hence the emphasis on 100 percent coverage completion reporting and annual reviews. Learning was a product of all this, but the foundation was accountability. The mechanisms for accountability generated the information for learning. You can emphasize the accountability or learning aspects of evaluation, but in my view they're indivisible, two sides of the same coin.

As for examples of learning through evaluation, institutional change comes about in many ways, in response to external pressure as well as learning from experience. That's especially true today, when the Bank has a much higher and more contentious profile than in the past. Unlike the supportive halcyon days that I remember, 10,000 people are reportedly demonstrating against the Bank in Oslo this very day. Those influences, as much if not more than internal influences, now affect internal policies and procedures.

But I would add something else. If operational staff are competent, conscientious, and working properly, the probability of learning for the first time from evaluation should be quite low. Good quality staff will always be learning from their experiences. What they may not know of is relevant experience in other units. So evaluation can be very important in disseminating this experience and thereby enhance learning throughout the Bank.

I had hoped that the then-central operational staff would also disseminate lessons learned, for demand-driven learning is bound to be more effective than supply-driven lessons learned. Unfortunately, only the central agricultural unit took up this role. Behavioral change does not come easily, even in the Bank!

Participant: Bill Stevenson. Nice to see you again. Mr. Weiner interviewed me 30 years ago for the Bank's Young Professionals program. And then he interviewed me ten years ago when I was up for Director of the EBRD's Project Evaluation Department. Your comments reminded me of a saying attributed to Harry Truman, that "the only thing new in this world is the history that you don't know." You've taken us through issues you were facing back in the 1970s that are still current today. The one thing that's close to the hearts, I think, of all the managers, is the issue of staff rotation. You mentioned the problem of people being damned for their professionalism when it comes time to move back to operations, and you told a happy outcome story. But as a practical matter, isn't people's candor constrained by the knowledge that candid treatment of results might put them in a bad light? What can be done about this?

Mr. Weiner: There is only one solution that I can see, but it may not be feasible anymore. It is to have a centralized personnel system that, after taking into account staff preferences, ensures that at the end of five years staff are reassigned in the larger interest of the Bank. That is very different from having to knock on doors and then learning that candor is held against you, or at least fearing that people hold your candor against you. But if I recall correctly, the decentralized personnel management was already in place in the 1970s, making things more difficult. The downside of the market approach toward career development is that it affects not only people's ability to move out, but also people's willingness to move in.

Mr. Picciotto: I think you've put your finger on it. But I think we are, Greg and I and Bill, trying to find a solution to this, namely to see if we could finally institute an evaluation network to which evaluators could belong, in addition to having a professional network. So that there could be a default mechanism where the DGO would, in fact, make a decision. We are hopeful to institute this, but it has not yet been set up. I hope it will be set up before I leave.

Participant: A much easier question. Sometimes I find that my colleagues on the operational and policy side frankly deny actual facts. Did you come across that kind of thing, and what might explain this rather unusual behavior?

Mr. Weiner: I don't think it's unusual. When people feel threatened, they behave accordingly. I once encountered that in a very unusual way. In the 1970s, the UNDP produced an evaluation of nonlending assistance to national planning offices, for which the Bank was executing agent. So it was a U.N. evaluation of something being done by the Bank. Since it was relevant to the Bank, I did the unheard of thing of sending the U.N. report to the Board with a brief covering note explaining that evaluation does not have to be done only by OED. This was a very critical report and elicited the same kind of reactions you encountered. The Board asked predictable questions, the staff gave predictable answers, and there it was. I don't see that the situation you describe is unusual.

Participant: So we still are working in the same department then?

Mr. Weiner: Yes, some things never change.

Participant: Beer brewing companies have tasters who make sure that Carlsberg this year tastes the same as Carlsberg ten years ago in Zaire and Denmark and so on. What's your impression about the stability of OED standards over time from what you've seen?

Mr. Weiner: Let's talk about investment projects to begin with. In my day we would review the experience, re-estimate the economic return wherever feasible, evaluate the institution-building effects, and put it all together in a one-dimensional judgment about whether the project was worthwhile or not. That's the way we rated them then: worthwhile or not. That judgment may have been fuzzy, but it was fuzzy because its constituent elements were fuzzy. Institution-building and economic re-estimate assessments can change with time. In any event they are estimates of future developments, so we chose not to be too precise.

That has since changed significantly. Performance judgments are now crisper—satisfactory or not satisfactory. Underlying classifications now permit comparisons that wouldn't have been possible earlier. So on balance, I think the changes constitute a marked improvement.

But I'm not sure what they say about the work of the Bank over time, which I think is the thrust of your question. I recall seeing an annual review some years ago in which the author stitched together portfolio performance indicators beginning with the first annual review. The resulting series revealed a dramatic break in the ratings. Years of performance ratings in the 80/90 percent range were followed by a sudden decline to a 60 percent average, compelling the author to address your question. What I recall about his conclusion was that the break was more evidence of the increased rigor of OED's evaluation analysis than of a marked deterioration in the performance of the Bank's portfolio.

Well, that was no doubt part of the explanation. But I wondered whether a much larger part of the explanation was that the indicators reflected radical changes in the composition of the portfolio and in the way more recent operations were reviewed and processed.

I've been out of the Bank too long to say much more on this. The Bank is doing very different things today, which may intentionally imply increased risk and call for different performance standards. But I have no basis for saying that more recent portfolios are more or less worthwhile. One should bear in mind that this long-term change in the focus of Bank operations also reflects long-term change in the countries in which the Bank is working.

Mr. Picciotto: Mervyn, let me, on behalf of all of us, really thank you very much for taking the time and putting together those very, very useful recollections.

3

Building up the Capacity and Influence of Evaluation in the Bank and Outside, 1986–92

Yves Rovani
Director-General, Operations Evaluation, 1986–92

By the time I became DGO in 1986, the need for evaluation in the Bank had vastly increased. The operations then being completed reflected the new policies the Bank had introduced in the 1970s and 1980s, addressing poverty alleviation in general, but also energy. Also being completed were the first adjustment loans, a new instrument introduced in the early 1980s, with a shorter life than traditional projects. Moreover, evaluation was increasingly expected by the international community; this was the time when development effectiveness was being increasingly questioned, not in the streets as it is today, but by scholars who needed factual and honest answers.

This was the time when development effectiveness was being increasingly questioned.

But while the need for evaluation was growing, I think the evaluation capacity of the Bank was being undermined to some extent. Certainly it had been receiving much less support than under Robert McNamara, who had introduced the function and played an active role in making sure it was used. When Mervyn Weiner retired, management and the Board appointed Shiv Kapur, then the director, for an excessively short term, and recruited externally a director to be his successor.

This did not work out, and I was chosen as DGO on short notice. I happened to have been a supporter of evaluation (and self-evaluation) from its beginnings. I brought a lifetime of operational experience gained "in the trenches" as a project staff member and as a manager responsible for sector policies, strategies, and operations.

I saw my mandate at the time as twofold: to breathe new life into the Bank's evaluation system, and to encourage the further development of evaluation capacity outside the Bank. I'll discuss these two in turn.

The Bank's Evaluation System

On strengthening the Bank's own evaluation system, I'll talk successively about the people, the output, and dissemination and feedback.

People, Culture, Management

Starting with the people, I was told as I arrived on the scene that at a staff meeting of OED, Tom Clausen, the Bank's president, had asked "What's your impact"? At the time, some in OED saw this as a rather impertinent question. They took the view that "Our job is accountability, and accountability justifies our existence. Not the marketing of our products, which in effect is in competition with our accountability function." This made me believe we had a cultural problem to solve.

We also had a management problem, because the department as such existed only on paper. There were no central administrative functions to speak of. There was no budgeting. There was no cost information. Common services were absent. A divisional structure had been adopted about two years earlier, but had been effectively implemented only as far as one division was concerned, and that was the division of Graham Donaldson, which was responsible for agriculture and the social sectors. His division was to me the mother of what I believed OED should develop into. The rest of the department included some very good people—for example, the late Carl Jayarajah, who was an outstanding fellow. It also included a bunch of prima donnas who absolutely did not want any form of structure over them.

We solved both the cultural and the management problem.

As regards the culture, my first move was to set up a staff retreat in Annapolis, putting on the table the issues of development effectiveness and policy toward the environment, and opening discussions about our own objectives, products, and issues that were confronting support staff and operational staff in their day-to-day work. There was a lot of grumbling about this, but out of these discussions I think life started to spring. We put a number of task forces to work: one was to review the project completion reports issue; another to look into support staff issues; and another to look into office technology.

Inside of two or three years, we had the most modern office technology in the Bank, and we were well on our way to having an evaluation data system using modern technology, and all the credit was, in effect, the staff's.

The culture I was trying to instill was a participatory style of management. No style of management, in my view, is more applicable to a unit of self-starters like OED. (Also—and I'll make this point later when talking about dissemination—you need high-caliber people to interact with operational staff if you want to have an impact on the quality of operations.)

On the management side, my first job was to appoint a director. I was told that I did not need one, and that the DGO-OED director relationship had been fraught with problems. But I insisted on retaining the position and I looked for the best person I could find to meet the needs of the time. I found Ram Chopra, then an assistant director in the Economics Staff—who was not on the list I had been given. Three years later, when Ram left to take a job as country director in the Middle East, I chose Eberhardt Köpp to replace him. Eberhardt was a senior director with outstanding credentials in the Bank's operational policies and country operations management. The DGO-director relationship worked extremely well throughout—clearly the source of past problems had been the appointment of the director by the Bank's management, rather than by the DGO!

The culture I was trying to instill was a participatory style of management.

In the first two years of my mandate, we experienced a major staff rotation of close to 50 percent—due in particular to the effects of the Bank's reorganization of 1987. We did attract top talent, and were able over time to maintain high recruitment standards for staff as well as for managers. We also were able to adjust our staff profile to the changing needs of our programs, which our longer-range plans helped us to anticipate. As rotations continued, several staff left on promotion to other parts of the Bank, helping to dispel the widely held view that OED staff were less than welcome in the rest of the Bank.

One problem I did not have was a continuity problem. Coming into my new job as DGO at short notice, and without a director, I found my life was greatly eased by the DGO's assistant, Ms. Rosa Maria Duncan, now the late Mrs. Rosa Maria Wiemann. Throughout my tenure, the smooth running of the evaluation function owed much to Rosa's institutional knowledge and to her professionalism. I admired and was very fond of her.

Output

Turning now to evaluation output, by 1986 we had a significant resource problem in the face of mounting demands for evaluation.

All of our friends and not-so friends were saying, drop the audit ratio, do fewer performance audits, and you will generate the resources needed to support the program. We did drop the audit ratio a little bit, setting a lower target of auditing no fewer than 40 percent of completed operations. But there was a limit to how far we could go if we were to live up to our mandate of ensuring the Bank's accountability—especially because the system of project completion reporting was in decay.

Project completion reports were declining in quality and submitted with increasing delays. A large backlog had accumulated. The reason for this was that someone had the bad idea around 1980 of shifting the responsibility for preparing project completion reports away from Bank operational staff, the authors of the operations, to the borrowers. A covenant had been introduced in loan agreements requiring borrowers to do the project completion reporting.

Project completion reports were declining in quality and submitted with increasing delays.

Of course, this idea was well meant. The borrowers were supposed to do a good job of writing the reports, and the staff were supposed to do a good job of reviewing them, and the reports would be sent to OED, and everybody would be happy. But the facts were otherwise. Anybody who has been involved in project supervision, as I have for a good part of my life, and who has seen the kind of reporting that comes out of borrowers on projects, could tell you that PCRs would be handled in many cases by a clerk in a backroom, to satisfy the letter of what the Bank was asking. So we suddenly found ourselves with operational staff trying to do their best with these reports, which sometimes wasn't much, and with the backlog accumulating.

The decay meant that the whole system as conceived by McNamara—which was for comprehensive evaluation, with self-evaluation by operational staff as an essential component—was about to crumble.

I spent the next six and a half years pestering everybody about reducing the backlog and about improving the quality of PCRs. We had any number of task forces, reminders to the Board, reminders to the president, and broken promises. After six and a half years, however, the system was back on track.

Our limited resources meant we had little left for evaluation studies. As you all do, I think that special studies are the best means by which evaluators can distill experience and disseminate it. I also think, like yourselves, that impact studies are absolutely vital if we're to find out the reality of what projects contribute, and if we are to understand the determinants of sustainability. Where I thought I could make a contribution here was to open the net for the selection of studies, presenting them in a two- or three-year program that would be large enough, despite the modest resources, to show some kind of balance among topics. By opening the net, we were, in effect, trying to mobilize ideas with which to design a program that would be relevant.

Of course, relevance is half of the solution to the dissemination problem. I knew situations where we were producing a study, but I was hearing from operational staff, "Well, that's very nice, but we don't do this type of operations any more." Or "This is all very fine, but it's old stuff, you know. It's old stuff." One reason they were right is that our studies took forever to produce. In the production of the study itself, you had all of the fishing for evidence built in. Instead, I was in favor of investing more time in preparation, in designing the approach to the study, so that the approach paper sent to the Board would be the end of a process rather than the beginning, and so that the study could be done faster. That was maybe a dream, but it was a way to improve the quality of the preparation and save time.

The studies were controversial if they had to be.

The studies were controversial if they had to be. If you deal with important and controversial issues, and you do a good job of it, you are present; you are not a mouse under the table, even though the inclination of management was understandably to push us down. For example, for a study on the environmental impact of projects in Brazil, I had to take several trips to Brazil to persuade the government to accept the study. I used the argument that "I can't afford not to do the study, because of the bad image it would give us." For the same reason, I told them, "You cannot afford not to allow us to do it." So the Brazilian government agreed and, because we were helping develop evaluation capacity in Brazil, they agreed to support the exercise with staff inputs. And the deal was clear from the beginning: "at the end, we have our study, which we are fully responsible for, and you will have your study, if you want to, but we won't have a mixture of what you like and what we like. Otherwise, we are bound to have a diluted, a very diluted output." And that's the way it went. And I think the end result was a very good study.

We did other important studies: for example, reviews of adjustment lending; the four country studies started by my predecessors, on Pakistan, Tanzania, and later on Senegal and Tanzania; the rural development study, one of the major studies that shook the Bank; and the industrialization study, which looked at the early phases of industrialization in three Asian countries. This last also gave the Bank a shock. It hit at the raw nerve of the Bank's ideology of the time, and so the management response to it was very negative.

As to the administration of the studies, I thought that the experience gained in managing these exercises should stay with us, so we should have our own staff be the study managers, as was the case with Gladstone Bonnick on Tanzania.

In addition to the studies program, we were doing as many cluster audits as we possibly could; not only of series of projects, but also comparative audits of several projects. Part of the reason was to economize, but it was largely because comparison is so useful for drawing lessons of general application. Almost 60 percent of our performance audit output was in the form of clusters. That was just about the absolute limit. Many of these cluster audits were very interesting and enriching experiences.

The single largest piece of output of OED (and the most costly) was the *Annual Review of Evaluation Results.* This also became the most controversial one because—my luck being what it was—we began to show a decline in the performance of the Bank. And that decline continued until my last day, when the proportion of satisfactory operations had fallen to 63 percent.

When I arrived on the scene, the annual review was a headache because there were so many loose ends about the data that were not resolved until the very last minute of the exercise. Instead of merely massaging statistical information on projects completed in a given year, we transformed the exercise into a tool to present the important output of the department in the past year. Each year, we covered substantive topics. We were able then to treat the statistics differently, analyzing patterns by operations' year of approval rather than by their year of completion, and thus to derive more interesting messages. We also incorporated the results of impact evaluations and introduced a focus on the sustainability of benefits.

All this made the annual review a much more interesting exercise. I was perusing past annual reviews recently, fearing the worst, but found that they were actually pretty good pieces of work. The other thing we did was appoint the author each year, way ahead of time, to let that person manage the data retrieval and organization and also to do the conceptual work. We put some of our best people in charge of this exercise—which accounts, I think, for the good quality of the thing.

Dissemination and Feedback

In the early days of evaluation the Bank was centralized, and dissemination and feedback processes could be managed effectively as part of its central systems. There were central systems for quality control, and central departments that had the responsibility to take the messages from OED and in effect react to them and disseminate them. This was a very effective arrangement in a centralized operation such as we had then, and fine when you have a president, who is center stage by definition, with a personal interest in seeing that evaluation results are effectively used.

But shortly after evaluation was introduced, the central quality control system of the Bank began to crumble. Operational responsibilities were decentralized, first to the Regions, starting in 1972, and then to groups of countries in 1987, and the decentralized units became increasingly intolerant of any form of central oversight or advice.

This is where the participatory style of management became important: we needed the active participation and initiatives of all the staff of OED in the dissemination process. We needed them to talk at seminars, to interact with operational staff, to lecture at the Economic Development Institute (EDI, the precursor of the World Bank Institute), to develop courses for EDI, and to exchange views with the research people. To paraphrase Mao Tse Tung, we needed to be in operations, in EDI, and in research, as a fish is in water. Everybody had to be involved in dissemination. And we needed to back up their efforts by providing ready access to reliable evaluation data and easy retrieval of evaluation findings and lessons.

Of course, you need able people and responsible people to know how to do this selectively. The requests for help from operational staff were always there, and it was easy to find oneself doing other people's work. So we had to be selective, and this selectivity had to be administered not just by the director or division chiefs but by the staff themselves. They had to make the most effective use of the little time on their hands.

OED Précis, where in one to four pages, we would highlight the findings of a study or an audit and draw lessons from experience.

The other key thing we did on dissemination was to improve the presentation of our products. I pride myself in having reduced the annual review in 1986 from three volumes to one, and of a more modest size, to boot. No one missed the other two volumes! And we instituted a system where the annual review and the more important studies were professionally edited. At the beginning, in the confusion and the chaos of the Bank pre-1987 reorganization, all kinds of good people, really very good editors, had been helping us for free. As the dust settled, we had to find the resources to continue the effort. Ram Chopra and I were finally able to persuade Rachel Weaving to join my office. She managed the recruitment and supervision of editors as well as editing a few things herself, and she took charge of developing a new concept, the *OED Précis,* where in one to four pages, we would highlight the findings of a study or an audit and draw lessons from experience. These were an immediate success.

Feedback, or the reinvestment into new operations of experience gained in past ones, is an integral part of operational quality and controls and is the responsibility of operational staff and managers. The DGO reports annually to the Board and management on the effectiveness of the feedback process. The DGO and evaluation staff play a limited though important role in facilitating the process.

Helping countries to develop their own capacity for evaluation is just as important a task for the DGO as promoting self-evaluation in the Bank.

We promoted a few ideas. I tried, for example, getting the Regions to indicate what they had done to reinvest the lessons from earlier studies of value to their particular Regions, and to have their signed statements incorporated in our annual report. Some of these statements were quite good. Others were not. Where the wooden tongue of the bureaucrat was speaking loudly, it was very difficult to get uniform quality out of this effort, but we tried.

The other thing that we tried to do was to support best practice. The Regional vice president for East Asia once told me, "I have an idea. What if I asked, I instructed, staff to have a section in the appraisal report that states how experience relevant to this project has been used and applied, how it has been reinvested in the design of this project?" I said, "Great." He said, "Would you help me?" I said, "Yes, we will help to the extent we can," though I was not very welcomed in OED when I announced we have this extra piece of work, which is to read appraisal reports. The Latin American and Caribbean Region also took up this system.

So, in dissemination and feedback, obviously, in the decentralized system of the Bank, there are myriad ways to proceed.

Evaluation Capacity Development

As my predecessors, Mervyn Weiner and Shiv Kapur, also recognized, helping countries to develop their own capacity for evaluation is just as important a task for the DGO as promoting self-evaluation in the Bank. Both tasks are key to the effective process of learning and enhancing operational quality.

Why evaluation capacity development should be a priority was obvious to me. In an ideal world, borrowers would evaluate their development policies and investments, and donors would evaluate their contributions, as we were evaluating ours. Of course, trading experiences among participants would add value and impact to the work of each of them, and as more participants joined in, the value to each of them would go up—just as in a telecommunications system, the addition of one subscriber makes the system more useful for all the others. Participants should not be limited to the borrowers and ODA participants; there is considerable benefit in expanding the evaluation network to private sector actors, or agencies such as highway authorities in developed countries, that have valuable experience to research and share.

Coming back to the real world, I learned quickly that I had to focus selectively on countries that showed real commitment and willingness to allocate resources to evaluation, and that would build and animate the decentralized systems necessary to do effective evaluation and to sustain their efforts.

I saw my role as to initiate the dialogue, wherever possible both with countries and other development institutions, and to follow up wherever prospects—that is, the commitment of the other party—made it worthwhile. To ensure follow-up, I assigned the function initially to an advisor in my office, Robert van der Lugt, and when Robert left, I appointed as its manager Pablo Guerrero, whom I had met when he was heading the internal evaluation unit of the Inter-American Development Bank. Together, we proceeded to involve the staff of OED whenever their work was relevant and the effort needed was affordable.

I would like to bring out two stories in this connection. I had been told that in Brazil, a central evaluation unit had been created a few years earlier but had ceased to exist. I was asked to meet with the Brazilian delegation to the Bank's annual meeting and was informed that evaluation had been made mandatory in Brazil's last constitutional reform and that our help would be welcome. I suggested that the Brazilians appoint a senior official and send him to spend time with us to develop the project with whatever help we could provide. So they did, and we—Pablo Guerrero in particular, and others in OED—developed a fruitful cooperation over several years. As part of his central duties, Dr. Lapa undertook the difficult task of ensuring meaningful comments on our evaluation work in Brazil. When we undertook the complex evaluation of the environmental impact of Bank projects in Brazil, the government agreed to contribute to the effort.

An example from Colombia also illustrates the role that a good evaluation study, and collaboration with evaluation staff, can play in helping to launch evaluation in a country. At my first Bank annual meeting I met a minister from Colombia, who asked that we do an evaluation of the Bank's involvement in the power sector in Colombia. I agreed without hesitation, because I was already aware there was a problem. Rene Ribi, civil engineer and a former power division chief, and Trevor Byer, who had led the Bank's Colombia energy assessment mission and was a star energy planner, undertook this work, helped by a financial analyst, and did, I think, a superb job. Then the Colombian Minister of Finance, Mr. Holmes, showed up in my office, on the occasion of a later annual meeting, and I said, "Mr. Minister, you are aware, I suppose, that one of your predecessors asked for an evaluation of the power sector in Colombia?" He said, "Yes, certainly, that's why I'm here." He leaned forward and said, "The study is very critical of the Bank." Then he smiled and said, "It is also very critical of Colombia." And so it was. He proceeded to tell me he was arranging a seminar in Colombia where all policymakers and interested parties would be invited to discuss the needed reform of the power sector, and he asked for the participation of the authors of the OED study to help lead the discussions.

Then I said to the minister, "You have just seen the power of evaluation to turn things around. Would you consider creating an evaluation capacity of your own?" He said, "Not a bad idea, but how do I go about this?" I said, "You choose a high official whom you trust and you ask that person to think through the idea and propose a program to you. We would be delighted to welcome this person, and share our knowledge and the network of our friends to help in this endeavor."

And so he appointed Eduardo Wiesner, the retiring executive director for Colombia, who came to spend time with us, meeting with others in the evaluation community—including Eleanor Chelimsky, who was heading the evaluation arm of the U.S. General Accounting Office—and developing in the process a passion for evaluation. The minister approved his proposals and the project took off rapidly. A year or so later, Eleanor Chelimsky, Pablo Guerrero, and I were invited to a very interesting seminar on evaluation in Colombia where we were presented with four evaluation studies done by consultants and where, in effect, every official in the country, including the attorney general, wanted a piece of the action. (In Colombia, as in Brazil, evaluation had been made mandatory in the last constitutional reform.)

Conclusions

In retrospect, referring back to Mr. Clausen's question, we did achieve impact in the Bank. I will cite only one example in this regard: the Wapenhans Review of the Bank's portfolio performance. This review would not have been launched were it not for OED's reporting of deteriorating portfolio performance in a series of annual reviews and studies. And it would not have been as credible and effective except for the material and analyses provided in a number of OED studies—on project supervision inter alia—or an important review of economic analysis that I had asked management to undertake, and which Joanne Salop carried out so effectively with the help of our own Pablo Guerrero and a few others.

As my term ended, I felt I was leaving a stronger base for the future, and a lot more wanting to be done by my successors.

At this point, I think I would like some questions.

QUESTIONS AND ANSWERS

Participant: I am Patrick Grasso, and I am the knowledge manager here. Mr. Rovani, one your points that caught my attention was about using advanced technologies, and particularly, I gather, for capturing our reports electronically. I think it's worth mentioning that the effort you made there is still paying off quite a number of years later, and is very useful. In fact, I don't think we could do our knowledge management job without it. It's been a very effective tool for us to have this kind of technology. What gave you the thought to begin to move in that direction?

Mr. Rovani: Well, two things. First, I was hit in the face when I came to the OED by something that was called the Concordance. The Concordance was a huge book, manually, clerically, assembled by someone looking for key words. It was awful and I stopped that work. Also I was struck by the fact that every year, when the annual review had to be prepared, the compilation of data had to start from scratch. Then, at this famous retreat I mentioned, there was interest in introducing personal computers.

So everybody got equipped, and we made a deal with the information technology department of the Bank to have one of their staff in residence to upgrade our data management. The support we got from them, particularly from Peter Gutterman, is what helped us develop a state of the art data management system and also to conceive the idea of the searchable evaluation text base and make it a reality. I was not an actor in this. I just was interested; I opened the door, and supported this thing. I went with Pablo to see what the Canadians had been doing, and they had a very good system. I think that had cost them several million dollars, and we couldn't spend $100,000. So we waited a little, and when off-the-shelf technology had progressed a little further and costs had come down, OED and Peter were able to develop a "user friendly" text retrieval system tailored to our needs.

Participant: A little-known fact is that the rural development report was only the second Bank report published on the desktop. The precursor was the Uruguay Round report.

Mr. Rovani: I did not know this. It is another piece of evidence of the talent and innovative spirit in OED. Let me use this opportunity to make another point. Nowhere in the Bank is there under one roof as much diverse expertise, experience, and concern for quality as you have in this outfit. I think you are blessed with the fact that you cross the hallway and you meet somebody with a different background, different ideas. This is one of the great things about OED, and one of the things that made OED attractive to prospective staff members.

Participant: You stressed the importance of impact evaluation for getting a better understanding of the determinants of sustainability. But in practical terms, the meaning of impact evaluation is rather elusive. One multilateral development bank defines it as four years after project completion. Here in OED, it's sort of a mix. And as you move the time horizon out, the evaluation becomes much more challenging, more like a research project, with difficulties of attribution, and extra expense. I'd be interested in your thoughts on this issue.

Mr. Rovani: I could not give you the expert answer you would like. A number of staff, I am sure, could answer this better than I can. But for me, two things matter: to be able to report meaningfully on the economic worth of projects well into the operational life of the project, and to do this cost-effectively—because ultimately this is what makes evaluation credible. Of course, the task of evaluating the impact of projects varies very much with the type of project. Irrigation projects are probably the most difficult ones because of their long life and the many facets of their impact. In my time as DGO, we could not afford, without substantial borrower involvement, the costs of the long-range research involved. I thought we could do better by building into projects, from the beginning, the basic data needed for eventually evaluating the operational work. This said, we did produce meaningful evaluations of the impact of projects, with some surprising and worthwhile results. I can recall an impact evaluation of an irrigation project in Morocco, from which material was developed for EDI courses, and comparative evaluations of area cotton projects in several countries in West Africa.

Participant: How do you see evaluation in the immediate operating context of today? The Bank is facing new kinds of criticism, new pressures, and new shaping forces for its future. Evaluation is being questioned and certainly being looked at more critically. So the question of accountability is key. Do you see this as a shaping force for the future of OED and its role in the Bank?

Mr. Rovani: Quite clearly, the Bank's evaluation function cannot escape the wholesale criticism being laid at the Bank's doorstep. The function is indeed part of the Bank; I know of no other course. Accountability will always be a key feature of the evaluation function. The issue is what you evaluate and how well you do it. I would say, stay the course, do the best job you can, and strive to continue to be independent.

I will touch on your second question in a different way. When I was briefing the executive directors in 1986, the United States executive director told me, "Mr. Rovani, evaluation is very important. We wish you to get to the top." And I said, "Mr. Keating, I promise I will climb. I don't know how far the top is." And my answer to you today is that we did a lot of climbing but, on my watch, evaluation did not get to the top.

Participant: I just want to make a little postscript, Yves, if you'll excuse me, to say that for those who have known him as long as I have, Yves hasn't changed. The very first time I arrived in OED, in 1987, I think it was, he was holding a meeting, a staff meeting, and I got there late, and I inched open the door, and Yves was already in full flow, and I got in, and there was a little chair against the wall. I thought I'm going to make this, you see, and I was just sitting down in my little chair, having not made a sound, when he interrupted what he was saying, looked straight at me, and said "Nice of you to come."

Participant: The latest review by the Bank's Quality Assurance Group shows that among the Bank projects entering the portfolio last year, 93 percent were satisfactory. If you put out a report showing that 93 percent of your projects are satisfactory, do you have a credibility problem? A lot of the measurement techniques that we use are looking at Bank *processes* and how effectively they are performed. And they miss in some key respects the ultimate *impact* the Bank is having on the beneficiary. Some elements of measurement are beginning to capture impact, and I think they should be part of the further discussion that really needs to take place.

Participant: I wanted to make one comment that sums up my personal experience in OED. It is about what evaluation does to the evaluator. It's a tremendous discipline for training your mind for doing other work, where you have to look at a situation and discern the essence of what is determining that situation, in order to prescribe and make suggestions for policy and other things. I retired from OED seven years ago now, and since then I have been busy, but I must say that I always can tap into the reservoir that I accumulated by my experience in OED, because it's a tremendous training for the evaluator. Quite apart from that, you benefit from the quality people you work with in OED and the interchange that happens there.

Mr. Picciotto: Yves said two things today that particularly resonated with me. One is that his Bank career was always at the center (Projects, Central Projects Staff, Energy, Evaluation). V. S. Naipaul has a book called *Finding the Center,* and I think that is a job for evaluation, as Yves highlighted. Yves also talked about getting to the top. I think today we are still trying to get to the top. I don't think we'll ever make it, but I think that OED also still has both continuity and change, as Yves mentioned. So it has made us very proud to have you here today, Yves.

4

From Working Methods to Business Plan, 1984–94

Graham Donaldson
Chief, OED Division 1, 1984–94

For me, OED was a very important assignment in a 40-odd year career inside and outside the Bank. I think there was no place where the things that I was able to bring to the workplace were better used than in OED.

During the period I was there we started out with very simple working plans and then had to develop some working methods. And we ended up, of course, doing business plans, which is quite a sophisticated tool for controlling a very intensive program in a complex outfit with a whole lot of outputs. And so there was really a great change.

When I came to OED I had three advantages. The first was that I'd run a division for 10 years, and it was quite a big division; it went from 15 to 30 people in that period, and I found that a very useful experience. Second, for 15 years before that, I'd been doing empirical studies of various kinds in various places around the world, all based on field work and field observations, so I had useful field experience. Third, I had been a major user of OED reports as an input into policy work. In the Bank's Department of Agriculture & Rural Development, where I had come from, whenever we wrote a policy paper, we first looked at the project record, which was best documented in the reports of OED. I also used to review PARs and studies in response to requests from the director of agriculture. He used to walk into my office and say, "I was just at a lunch where Merv Weiner said . . ." and he would repeat something that had struck him as unusual, and then say, "perhaps you'd better check it out, see what's in that report, and tell me whether there's anything we need to do." And so I read a lot of OED reports simply because of Merv Weiner's dissemination efforts.

OED in 1984

When I arrived in 1984, OED was in a very difficult situation. Certainly the department was up and running. It had some very good staff, as well as some pretty weak ones, and they had produced some outstanding reports. But the standard of the reports was not consistent, and some were pretty poor. The management style was authoritarian. There were no divisional budgets, and everything had to be signed and cleared through the director's office. I couldn't send a mission anywhere without prior clearance from the director. Knowing that a Bank division chief had some freedom to initiate actions, I made a plea that we institute a few of the basic things that went with forming a division, and Shiv Kapur was really very constructive in his responses.

The OED staff in 1984 were very alienated. There was no team spirit. Many had been transferred to OED against their will. They were offended by the fact that they'd been moved to the M Building, which was way down near where IFC is today. (Just a month or so before I arrived, OED had moved back to the sixth floor of the H building.)

The middle managers' (division chiefs') positions had recently been created, and most had been filled by elevating the senior evaluation officers. Few of these managers had any prior managerial experience; none had much experience in empirical analysis or field work; and they not only lacked a division chief's prerogatives, they also lacked the division chief's grade. All of these things created a pall of aggravation and alienation throughout the whole staff. Also fundamental to this problem was that the staff were working as individual units and there was no mechanism by which they could exchange information or learn from one another. Very rarely did they go on more than single-person missions.

The divisional work program I inherited was the sum of 12 individual programs, each formulated independently. The selection criteria for performance audits were very weak. There was no requirement that we examine series of projects, and the earlier projects in a series were completely overlooked. Yet, particularly in agriculture, projects build upon one another; for example, sometimes remedial actions are added into the next project to make up for deficiencies that were detected in the previous one, or the previous several. So just to look at one without looking at the others doesn't make sense. Further, there was no connection between audit work and impact evaluations. Essentially we had "targets of convenience" rather than "targets of opportunity," which might have been a better way to go.

New Operating Procedures

So we set about introducing some operating procedures, and I must say that everybody cooperated on this. When I say "we" did it, the division did it, because we did it collectively. No one could have done it nearly as rapidly or as well if it hadn't been the division's shared concern. It was teamwork.

The first thing I did, and that Shiv Kapur allowed me to do, was to take the division on a retreat. That had never happened in OED before, and there we let all those who were naysayers and complainers say their bit, and then let the others gradually beat them down. It was a great exercise, and we came away from it with a whole lot of people strengthened in their positions and their leadership roles within the group. In particular, I remember Christian Polti and Paul Duane as being two very important players who decided that what was being talked about made a lot of sense to them, and they just gave leadership to the rest of the group.

As the business schools say, first you have make the change, and then you have to embed it in the culture. The way we did things, the changes essentially became part of our culture, and people were proud of that. We realized just how strong a culture we'd created when, after the 1987 reorganization, the infrastructure group (who were serious hardhats) was merged into our division. They became converts; they even admitted that they enjoyed coming to division lunches. The other way that I could tell how far we'd moved was that some people who had earlier left OED came back to us, including Ted Rice and Ralph Romain.

Let me review some of the changes we made in operating procedures.

When I arrived, missions had no terms of reference. There was no process for identifying issues before going to the field. There was no real preparation requirement before going out to do an audit. Sometimes people read the project completion report on the plane out and then did the file search later. So my response was, first of all, to introduce terms of reference. The requirements were that the terms of reference must specify issues only after the file search, the PCR review, and interviews with the Regional offices had been completed. And the terms of reference had to say who in the Region had been talked to (so I could check). We pulled people into line very quickly.

We standardized the reporting to make our results easier to use.

Because there had been no cross-fertilization on working methods, standards of quality and content varied widely. One of the things that I had learned back in the Agriculture Department was that you couldn't find things quickly in the OED reports. The reports would cover the same topics, but not necessarily in the same order or the same format. So we standardized the reporting to make our results easier to use. We also provided technical assistance for recalculating internal rates of return. That meant that we were able to get a fairly standard quality product, which was at least an improvement over what was there before.

To encourage cross-fertilization, we introduced debriefing sessions. That was basically the division lunch. But it meant that people had to come back and talk about what they had done in the field, and in this way others got to know what each was doing. And we introduced peer reviews so that each draft had to be reviewed by two other members of the division. Usually one of the reviewers had to be an economist and the other a technical person; one or other of them had to be a senior staff member. The peer reviewers were otherwise selected by the respective audit authors.

The other challenge was to forge a link between audits and other work. To achieve this, each evaluator in the division had to commit to producing an evaluation study, and feed all of his or her audits into that study. The problem here was that, by listing all the studies in the pipeline, we created an expectation for the delivery of reports that were a long time in preparation. Once OED became internally transparent (when Yves Rovani came in), the Joint Audit Committee knew our work plans at a very much earlier stage, and they got really frustrated with the fact that we would take three years, and maybe five, to deliver a study report. (What we had done early on was just to not list studies until we had essentially all the audits done, and then we could say, here's a list of studies that we'll deliver in 12 months.)

To solve the backlog problem we went to "delinking."

Another feature in 1984 was the "backlog" of projects for which no completion reports had been prepared. I believe something like 400 out of 1,600 projects had never been audited for lack of completion reports. Looking at the lists of projects without PCRs, we knew that high proportions of them were in agriculture, and especially in Africa. The backlog seriously undermined the credibility of OED's accountability reporting. Given that we had no data for one in four of all the Bank's completed projects, what then was the value of what we said in the annual review? When Yves Rovani became director-general in 1986 he acknowledged my concern. This was the first time I managed to get one of his splendid responses. I had found out that he didn't mind bad news, and so I'd come in and tell him about something that was not going right. He'd say, "Oh, my God, 'Graahm' what are we going to do?" and then he'd hear my ideas on what we should do.

To solve the backlog problem we went to "delinking." That is, we removed the requirement that OED could only audit a project if a PCR had first been prepared. Then we moved to clear up the backlog. The result was predictable: the evaluation statistics on the performance of agricultural projects plummeted. The change was widely misinterpreted as a decline in the quality of projects, but it was merely the result of introducing data on projects that had been overlooked in the past.

We introduced the concept of "group audits," which were done wherever there was a project series. These were a great success, I think because they covered such a long time span that they became like impact evaluations in some ways. Then we got even more daring, and introduced what Yves called "cluster audits," where we would take similar projects in several neighboring countries and audit them collectively.

One of the consequences of group audits was that the output and performance of staff went way up. Previously staff were scheduled to do something like four audits a year, and, for a good many of them, that was the full workload. I remember that Paul Duane and Christian Polti, and probably Kathryn McPhail, too, did as many as eight to ten audits in a year. So we cleaned up a lot of the backlog projects in a very short time.

Then we introduced essentially a new working method for putting the whole audit program together. The procedure became that the division chief determined the projects to be audited, albeit often on the suggestions of division members. That caused a lot of aggravation at first, but in fact these rules on new working methods became the Operating Procedures that were introduced throughout the department not long after Yves Rovani came into office as DGO.

In addition, we standardized impact evaluations to deal with clusters of projects. It was quite easy to evaluate the whole of, say, the irrigation program in Indonesia because it consisted of irrigation projects one through eight. Also, because we clustered these evaluations, we could afford to hire local consultants to do field surveys.

I also used to put out ephemeral memos on specific issues, which said, look, this is happening, therefore, henceforth we're going to do this. Staff called them snowflakes and once gave me the suggestion that I get them printed on especially soft paper. They proved to be a good way of getting information out, and staff followed these guidelines faithfully. Everybody could see what the ground rules were, and often the support staff were the instrument by which they were implemented. (I see Jacintha Wijesinghe down there smiling.)

The new procedures had the effect of bringing performance up to essentially best practice for the division as a whole.

I think that together the new procedures had the effect of bringing performance up to essentially best practice for the division as a whole.

The relatively large size of the agriculture portfolio in the period we're talking about made it easier for us to do clusters and group audits and so forth.[1] But the new initiatives worked all right in education and for infrastructure, too.

Drivers of Change

So what were the drivers of what was essentially an organic process of change? I would stress the following:

- The Rovani impetus for rigor and for getting the product right
- Changes in staffing
- The changing role of the Joint Audit Committee
- A growing service orientation.

The rule became "No speculation." This proved very popular with the Regions.

The arrival of Yves Rovani as DGO in 1986 changed things dramatically. Mr. Kapur's office had been on the seventh floor and I don't think we had ever seen him on the sixth. So it was quite a shock when, on his first day as DGO, Yves walked into my office and sat down in the chair and asked, "Well, what's going on?" And he used to come down regularly, more or less once a week, and say, "You know, what are we going to do about this? Have you noticed . . .?" Et cetera, et cetera. So he created a much more relaxed environment. And, of course, he very quickly moved his office to the same floor as the rest of us.

RIGOR

Yves contributed a great deal to the rigor of OED's work. First, he clarified OED's perspective. He has mentioned the fact that people in OED saw their job only as accountability. He went to the first departmental retreat and said, "I want you all to understand that OED is in the business of transparency." That marked a tremendous change in the way that our work was looked at, and it ruled out a lot of the axe-grinding element of OED reports. The rule became "No speculation." An audit could say anything it liked based on fact, but no speculative statements were acceptable. This proved very popular with the Regions.

STAFFING

As to staffing, when I came to OED I began bringing in some new recruits: Dell Myren, Julian Blackwood, and Kathryn McPhail. During my first two and a half years, 10 people left the division either by retirement or otherwise, and the change in our staff mix greatly improved morale and participation and the caliber of staff.

When he arrived, Yves did for the rest of OED what we had begun in OEDD1. Something that was very well managed by Yves and Ram Chopra was to use the Bank's 1987 reorganization as a golden opportunity for OED. With Ram Chopra as the talent scout, Yves filled OED with top-level performers and dropped lots of "strong individualists." OED became a different place overnight. At that time they switched from three to two divisions. Alec Nowicki was in charge of one-half of the department and I had the other half, and the two of us could work together very well, which was really wonderful. And with the other managers being very supportive, it was a very productive period in a relaxed atmosphere.

MORE DEMANDING OVERSIGHT; BETTER SUPPORT

Something else that Yves didn't mention is that the Joint Audit Committee changed a great deal under his tutelage. He managed the Committee superbly: he got a great deal more support from them than ever before, on budget, new initiatives, and staffing. The way the Committee was managed encouraged some very dynamic and committed executive directors to volunteer to be members.

GROWING SERVICE ORIENTATION

About this time, too, we began a process of outreach to the Regions, driven by the belief that we must be more of a service operation, while still maintaining enough detachment to make independent observations. Under Yves we had a lot more contact with Operations and with the Regions, working at all levels. Yves worked with VPs. Ram went to directors' meetings, as did Alec Nowicki on occasions. I went to division chiefs' retreats and Regional staff retreats and kept spreading the word.

OED staff went to other divisions' retreats around the Bank. This gave us a chance to explain OED's purpose and procedures to the rest of the Bank. During these sessions we urged them to come talk to us if they had a problem with OED draft reports they were reviewing. And on occasions we'd jointly review the evidence relating to the offending statement or statements. If we in OED rejected their concern, we'd call them up and tell them why, but quite often the burden of proof rested on OED and we had to accept that there was an element of speculation and make an appropriate change.

These changes helped to restore a working relationship with Regional staff.

We also tried to get the Regions to improve the quality of project completion reports. One of the arguments we made to the Regions was that their best defense against critical audits (about which they were complaining) was to have a good PCR, which clearly showed outcomes and problems and was reliable reporting. In some cases we helped out by finding appropriate consultants who knew projects and who were prepared to do PCRs for the Regions.

Another of Yves Rovani's contributions was that he cleaned up the annual review. Since we didn't have very good data anyway (given the aforementioned backlog), how could we justify issuing two volumes of key performance indicators? Presumably presenting two volumes of tables had been an effort to essentially bury people in detail. After some consideration we just dropped the second volume, without announcing the change. We removed a lot of the vagueness and obfuscation from the annual review. And as we started to give our messages and conclusions a lot more simply, more people started to hear us.

Those are some of my recollections. I'm not suggesting that they're all consistent, but I hope I didn't rewrite history too much.

QUESTIONS AND ANSWERS

Participant: I'd like to amplify what you were saying on impact evaluations, where I did a lot of work. We used these tools to help advance the Bank's agenda as well as OED's agenda. In those days the Bank was increasingly coming under criticism on a number of issues such as environment or gender or health impacts. These issues were not much covered in the typical audit; indeed impacts in these areas had tended not be documented in projects. So we used the impact evaluations to address issues that were high on the concerned public agenda, but on which we didn't have information.

I think another important feature of that period in OED is the development of the link to organizational learning. More and more we were looking to see what were the lessons to be drawn, what was the replicability, of the operations we were evaluating.

Mr. Ingram: A lot of your comments have been about evaluation methodology, about increasing rigor. Was evaluation methodology an issue that people talked about? Did clients demand to know what was going on inside the black box? How did you deal with that?

Mr. Donaldson: Yes, there was a good deal of discussion about it, including in the JAC and with other agencies, and during Yves' period, as part of our reaching out process, we got officially involved in the OECD evaluation group. I think we became a lot more transparent in our operations, just as the Bank was becoming a lot more transparent in its operations.

Participant: Essentially your division was a pioneer for the rest of the department, as Yves acknowledged. I got the impression that the strong culture in your group was not always replicated in the rest of OED.

Mr. Donaldson: OEDD1 had a bigger group from a bigger sector, and that made it easier to build camaraderie. I did have to intervene in division lunches to make sure that it wasn't just agriculture guys all talking with one another and the others sitting on the sidelines.

But one of the things that began to happen was that within the division we started working across sectors. For example, Hans Thias (education) used to do some work in agriculture, Robert Burns (transport economist) started doing some irrigation projects, Christian Polti on one occasion audited a bunch of road projects, and then we had Ron Ridker who could work in all three sectors at once. Perhaps you can put together a better team by building it up piecemeal. The flood effect, when lots of high-caliber staff came in after the 1987 Bank reorganization, didn't necessarily help in building a team. And in the other division there was not a lot of continuity in management, until Mark Baird came on board. I was in OED continuously for 10 years. But in the other major division there was Alec Kirk, then Otto Maiss, then Alec Nowicki (who died in office), and then Brian Shields (who died in office). I believe that really affected the other division.

Participant: You mentioned the service orientation. I'd like to ask you about your support of the Bank's monitoring and evaluation systems and evaluation capacity development.

Mr. Donaldson: We spent a lot of time talking about and working on the Evaluation Capacity Development Program. In any client country that had created an evaluation capacity, when we were going to do auditing, we would go to the evaluation unit first. This was really helpful in that, for the first time, OED had a destination, whereas previously we were sort of supernumerary; we didn't really belong with the Treasury and we didn't really belong with the sector ministries. And we often used to take one of their evaluation staff with us to the field. Sometimes we commissioned them to do the survey for impact studies. And I think that was very useful. In addition, a lot of people came in and worked in OED for a while and went on our missions and things like that. Those kinds of things helped.

Participant: In your experience, was any report suppressed because it upset an audience. Were you forced to kill any study for the wrong reasons?

Mr. Donaldson: No, I don't think we were ever forced to kill a study. I think we'd avoid them strategically rather than waiting to get blocked. But there was always opposition to any study we wanted to do. Evaluation is quintessentially a political process. Every time we had to ask, "Okay, is this worth it?" But I don't think we ever backed off entirely.

There are people who just want to turn OED's findings aside. I would hope that there are better, cooler heads who appreciate them better. That said, one of the things that we had problems with was trying to discern what the Regions would like to see done. They would suggest, say, evaluating all the rural development projects in country X, but largely because that division chief was tarred with those projects as part of his history, and he wanted to see them evaluated to provide a cover. Because once OED comes out with a fairly—even a hard-hitting—report, it allows the Bank to say, yes, okay, we see it, and to turn the page and put it behind them. By doing this, to some extent we've let them off the hook.

Participant: Against the rules that you've imposed now, could you speculate a little bit about the feedback loops into project design on the project side of the Bank?

Mr. Donaldson: I don't have any outstanding examples of feedback loops. There was quite a lot of search for information going on by Operations staff, not always overt and not always by senior management, or with knowledge by senior management of what was going on in the Regions. People like Dennis Purcell would have lots of Operations staff come and talk to him. Christian Polti used to have more visitors from the Regions in a week than was really tolerable.

Notes

1. During the 10 years FY 75–84, agriculture was half of the Bank's total lending program; the Bank supported agricultural sector projects to the value of $50 billion in that period.

5

Promoting Professional and Personal Trust in OED

H. Eberhard Köpp
Director, Operations Evaluation Department, 1990–95

I'm glad to come back to OED for the first time since I left in 1995. I was director here for almost five years, and I took the initiative to join OED after quite a satisfying and varied career in what used to be called the Operations complex. I served as country director for Pakistan and Turkey, with about 100 projects under supervision, accounting for almost 10 percent of Bank/IDA lending at the time. I also spent five years with Ernie Stern, reviewing for the Loan Committee about a thousand loan packages. I liked operational work, but at the end I wanted to see what had been happening as a result of this work: what was the outcome of all our good intentions, our goodwill, and well-meant efforts? What could be a better place than OED to find out? So I was very grateful when Yves Rovani hired me at that time.

Up to then, I must confess, I could not say that in the Operations complex I had been very strongly influenced by OED's output, either directly, or even indirectly through the so-called lessons learned that the central departments in Operations Policy staff supplied to the Loan Committee. These were the so-called "good old times" of central quality control in the Bank, which was unable to prevent a very obvious decline in portfolio performance. As a country director from 1987 onwards, I had become more aware of OED's findings and criticisms. But at that time, the typical reaction of operational staff to these findings was still to dispute the factual details and to question the usefulness of the findings, and to attack OED's methodology—an alleged black box of unknown evaluation criteria. Most of the learning from past failures and successes still took place mostly through experienced, motivated individuals in Regional project or sector divisions.

When I joined OED, I had the feeling that I should focus on promoting professional and personal trust in OED, both from the Operations complex and within the department. The dictionary definition of trust is, inter alia, assured reliance on the character, the ability, the strength, or truth of someone or something in which confidence is placed.

Yves Rovani has provided an overview of 1986–92, the last two years of which I shared with him in OED. He has described the difficulties of staffing, of general recognition, of technology, of dissemination, and so on. So I will just highlight a few things that were strengthened or abandoned or changed during the time I spent at OED. These changes were prompted or initiated by the varied experience and ideas of the staff and managers of OED, by new directions and approaches in the Bank at large, and by constructive criticism from outside, as well as by budget constraints.

The goal was to enhance the strength of OED and to deepen the relevance and truthfulness of its output while strictly preserving the department's independence. We all wanted to increase the reliability of OED's work, both as an important source for learning from experience and as an objective assessment of the successes and failures of the Bank's operational work. To accomplish this, we saw the need to adjust the character of OED—meaning the staffing profile and the working relations—to changing needs and constraints, by improving methodology and evaluation criteria, by interacting more with others in the Bank and outside, and by providing more timely evaluations. It was not always easy to reconcile these often-conflicting demands.

Yves Rovani has mentioned in his paper the unevenness within the department in terms of management structure and style, skill mix and behavior. Many of these problems were already being solved when I joined, but some, of course, continued. Just as I joined, one of the three division chiefs, Alexander Nowicki, suddenly passed away, leaving a big gap in the department, and OED was also moved out of the H building to distant offices on H Street—not at all a popular move for the staff.

We tried to build up the divisional structure of the department, while making the walls between divisions more permeable—for example, through the use of cross-divisional teams. From my experience over 26 years in the Bank, I always found that any structure could work if you chose the right people and they chose to make it work. Mark Baird, who came from Operations, filled the division chief vacancy left by Alex Nowicki. The fact that Mark was relatively soon promoted to director elsewhere in the Bank and later to vice president underlined that the image of OED had been changing. The department was clearly seen now as a place where you could use your experience and critical sense and develop new skills to help build your career in the Bank.

We tried to practice transparency and participation in management and general working relations. One example relates to work planning. Up to 1990, when I joined, OED always prepared a work program, which was endorsed by the Board or relevant committee and served as the basis for budget discussions with management. But there was no tight or formal agreement on the work plan in budget terms. That certainly couldn't continue: budgets were tightening and the Bank was introducing more rigid rules that made all departments accountable for delivering an agreed work program. So we started a process, from the bottom up in the divisions, of work planning and costing out and projecting staff needs, then comparing division with division with division, and then bargaining within OED and with the Programming and Budgeting Department, and eventually seeking the endorsement of Bank management and the approval of the Board. This was tedious but it certainly led to more budget discipline, and to more predictability and timeliness in our output.

We tried to practice transparency and participation in management and general working relations.

I think the issue of budget approval was the weakest aspect of OED's independence in my time. Of course, it is the Board that takes the final decision on OED's budget, but in my time, the Board rarely deviated from management-approved proposals. And thus the work program and its content and direction were subject to strong influence by Bank management through the back door of budgeting. (I should add that, at least during my time, even Programming and Budgeting recognized the cost of OED's need for especially qualified people and experienced staff, and gave us a little more money or budget for higher-level staff.)

Perhaps even more important for the quality, transparency, and credibility of OED's work were changes and extensions or clarifications of OED's evaluation criteria and improvements in its tools. We undertook to review and clarify the criteria that were listed in our project information forms and to try to get the ownership of operational staff of these criteria. Some of you may remember that Bob Picciotto initially felt, "This review and clarification of the criteria, we can do this in a week or two. We just do it, period." But we in the department just said, "No, we really have to do it from the bottom up," and we set up an interdivisional team. The work took quite some time but it led to a much better, clearer set of criteria, and the acceptance of the criteria gradually increased substantially.

I will quote a little anecdote here. In 1995, after leaving OED, I worked in the Secretariat of the Multilateral Development Bank Task Force. We had a meeting with about 20 nongovernmental organization (NGO) representatives to get their views, and at that meeting Bruce Rich, a well-known critic of the Bank from the Environmental Defense Fund, started almost preaching about the need for the Bank to use clear criteria in assessing, appraising, and evaluating its operations. He read out a list of the criteria that he would recommend, and believe it or not, the list was the OED project information form. I said, "Thank you very much that you like these criteria, but that's what we *are* using in the Bank!"

To build ownership for the criteria within the Bank, we had a long process of consultation with Operations. This prompted them to change their guidelines for the project completion reports and rename them, since it became clear that these were really implementation completion reports. Thus the ICR was invented. After a long back and forth, they almost gave up and proposed just to fill out the OED project information form, but we in OED said "No, you should provide the facts, the judgments, the experience, which then gives the basis for an evaluation rating," and that's what they eventually did.

Another important step towards getting uniform criteria for the whole operational cycle was a joint study by OED with Operations staff on the so-called discrepancy between the rating of an operation by the final supervision mission and the often more negative rating by (usually the same) operational staff in the ICR, which was typically written about a year later. Usually OED had no problem with the ICR rating. I think the findings of this study were a revelation to many operational staff, and were a milestone in destroying the myth of OED's "black box."

There had always been a demand for relevant and timely and useful output, to give concrete guidance to Operations staff on what to look for and what to avoid.

Many of the things I have mentioned strengthened OED, but the interaction, formal and informal, of OED with other parts of the Bank was particularly important. For example, as Yves Rovani has mentioned, the Wapenhans Report was largely prompted by OED's findings and based on data from OED about deteriorating operational performance. Both Yves and I served on Steering and Review Committees throughout the preparation and drafting of the Wapenhans Report. I remember vividly the first meeting we had in the group I belonged to, which tried to delineate the work program. Some members proposed that we would take projects as approved by the Board, and look just at what happened to the projects during implementation. But based on my experience in Operations, and in particular OED, I asked whether this approach wouldn't mean that we assumed the project had been perfectly identified, perfectly prepared, perfectly appraised, and perfectly negotiated, and that we would therefore miss half the story. That if we put garbage in, we would get garbage out. Though this question was rejected at that first meeting, the eventual outcome was the concern for "quality at entry," which was highlighted in the Wapenhans Report and subsequently.

As to OED's output, there had always been a demand for relevant and timely and useful output, to give concrete guidance to Operations staff on what to look for and what to avoid. Bob Picciotto, especially, intended that all of OED's work, especially studies, should gradually extend and shift the department's focus. In the past, the main focus had been on whether or not Operations had met their objective, which may have been set quite some time ago, and under policies that had since been changed. That focus remained important in OED's work for accountability reasons. But in addition, we were looking for lessons that were relevant to current operational work. So we increasingly also compared operational outcomes with the intentions of current operations and policies. Graham Donaldson's division did a study on the relatively few success stories among past rural development projects. They asked, what made these projects a success? Also, as you know, the general focus of evaluation moved away more and more from the project level up to the country and global level.

An important strength of OED is its independence. Most important, none of OED's output went to the Board in draft for clearance; nothing had to be cleared by the Bank management. OED, and formally the DGO, took full and final responsibility for OED's output; the Board or its relevant committee and Bank management added comments, but they had to leave the original final version intact for all to see. This had not been the case in the other multilateral development banks.

An important strength of OED is its independence.

When I worked for the Multilateral Task Force on Development Evaluation, after leaving OED, I was pleased to see that the DGOs and their equivalents and directors of the MDBs were already starting to follow our recommendation to exchange experience, and try to evolve a common methodology and set of yardsticks. Again, the Bank really was ahead of the others in most respects.

Let me end by saying that I learned a lot here from Yves and Bob, but most of all from being in an environment of professional evaluators. I must note, too, that as the director of OED I was helped very much by Betty Bain. This job in OED was probably my most demanding job as a Bank manager, but I have fond memories of the experience and my associations with OED, and I look back with some pride and some sense of achievement.

H. Eberhard Köpp: The question of evaluation standards has been a perennial one for OED. I would suggest you can never achieve complete consistency of standards because evaluation is done by people. As good criteria as you may have, in the end it's people who apply them. That said, I think I would trust at least the comparison between today's evaluation results and the results of 10 years ago.

Gregory Ingram: Let me make a quick comment, going back to a point that Eberhard made earlier. OED is more independent nowadays, in the sense that we don't negotiate our work program with management, although we do tell management what we're going to do. We no longer even discuss our budget with management. Both the budget and the work program go straight to the Board Committee on Development Effectiveness. Independence has also been strengthened to some extent on the personnel side. But a lot of the other things that Eberhard mentioned are bones that we are still gnawing on in OED.

Participant: Does OED have access to trust funds and other sources of funding outside the regular budget?

Mr. Ingram: Yes. We have a self-imposed limit, which is that we don't accept more than 15 percent of our budget from those sources. The funds come principally from four partners: Norway, the Netherlands, Switzerland, and the Department for International Development in the UK. We use those funds largely for outreach and dissemination and for consultations.

The Logic of Renewal: Evaluation at the World Bank, 1992–2002

Robert Picciotto
Director-General, Operations Evaluation, 1992–2002

I was asked to talk about the renewal of OED in 1992–2002, but found it hard to gather together the various strands of this period in OED's history. Time shrinks in periods of rapid change, and it is as if my two terms as DGO (a third of OED's history) have gone by in a flash. V. S. Naipaul has written that "it takes thought (a sifting of impulses, ideas, and references that become more multifarious as one grows older) to understand what one has lived through or where one has been." I may need more distance and more thought to "find the center" of what has happened in OED over the last 10 years.

But looking back, it is clear that OED has never broken the golden thread of its history, originally spun 30 years ago. Preserving the best of the past has always been a characteristic of DGO transitions. And, therefore, the story of the recent OED renewal may have more to do with continuity than with change. Only when I recollected in tranquility what Chris Willoughby, Mervyn Weiner, and Yves Rovani had said about their own tenures as DGO did the logic of OED's renewal emerge.

What I discovered is that the conceptual foundations of the renewal were present at OED's creation. Robert S. McNamara viewed evaluation as an absolute necessity for an institution that was growing rapidly and facing risks in an operating environment characterized by diversity, possibility, and difficulty. Feedback and follow-up were central to the conception of OED from the start. Feedback is about learning and follow-up is about accountability. They are two sides of the same coin.

Modern psychology teaches that to adapt to the demands of a changing and sometimes hostile environment, all organisms must be able to profit from past experience. Therefore the textbook definition of learning is in three parts. First, to find out about the operating environment. (When a child tries to walk, for example, he has to see where he's going.) Second, to recognize signals about its evolution. And third, to adjust behavior in order to survive and adapt. Such learning is critical to the survival of all living organisms and it also applies to organizations.

But organizational sustainability requires more than individual learning. To overcome free riding and other dilemmas of collective action requires leadership, participation, and—above all—incentives to align individual behavior with corporate and social goals. Hence the critical role of accountability for organizational learning, and the correlation between demand for evaluation, corporate transparency, and management receptivity to change. Stagnant and insular organizations do not make use of evaluation. Open, adaptive organizations do.

Stagnant and insular organizations do not make use of evaluation. Open, adaptive organizations do.

Eric Hoffer, the longshoreman-philosopher, has described change as an ordeal whose only cure is action. But for an organization battered by external shocks, the remedy of action can be worse than the disease if corporate strategies and core competencies are not aligned with emerging challenges. Hence, in a turbulent enabling environment, continuous learning is needed to achieve development effectiveness.

It follows that periodic renewals that address all three aspects of learning are needed both for the World Bank and for its evaluation function. Thus, the history of OED's renewal is the history of its own capacity to learn, first to adapt to the external and internal operating environment; second, to identify new evaluation priorities; and third, to put in place the programs, skills, and processes needed to implement them.

Let me go over these three aspects in turn.

As to OED's external environment, the last 10 years have been full of surprises. Policy dogmas have been shattered. Predictions have been disproved. The geopolitical order has been reconfigured, with the end of the Cold War and again with the rise of terrorist threats. The end of the Cold War did not deliver the expected peace dividend. Instead, regional conflicts multiplied and military outlays expanded along with the arms trade. In the former Soviet Union, rather than generating prosperity, the transition from the plan to the market led to a rapid rise in inequality and poverty.

In 1997, despite their robust macroeconomic fundamentals, the tiger economies of East Asia, commonly perceived as economic miracles, succumbed to financial contagion. More recently, the information technology revolution bred an investment bubble instead of ushering in the brand new economy, immune to business cycles, that its advocates had expected.

In Latin America, widely hailed market-oriented reforms did not trigger the growth needed to stem the rise of poverty or to avert the meltdown of Argentina, a star performer of the international financial institutions. In the Middle East, a peace initiative crowned by Nobel prizes floundered, while politics and economics stagnated. In Africa, violent conflict and civil strife dominated the headlines, while adjustment programs failed to yield the expected supply response, and few policymakers foresaw the devastating impact of the HIV-AIDS pandemic.

Neither the ideology of market triumphalists nor the religion of state planners has delivered the expected results. According to Martin Amis, ideology is a belief system with an inadequate basis in reality. Stubborn reality is far more intricate, elusive, and influential than pre-established doctrine. Since evaluation concentrates on reality, it is an effective antidote to ideology (though cognitive dissonance and denial can neutralize its effectiveness). Conversely, as development dogmas lose credibility, evaluation gets a boost.

What about the internal operating environment?

Throughout its history the World Bank has been caught in the crossfire of competing ideologies. On the right, market enthusiasts characterize the Bank as a bureaucracy committed to excessive state interference in the economy. On the left, anti-capitalist protesters picture the Bank as a sinister tool of private interests. In fact, the Bank cannot dictate to its clients, and its Articles of Agreement have helped it to transcend ideological conflict, and, through principled leadership and cautious deliberation, to focus on the shared objectives of its members.

Occasionally, in a well-meaning effort to be responsive, the Bank has become too sensitive to development fashion. But Bank policy shifts have broadly reflected the consensus of the development community. The early 1980s saw a neo-classical resurgence, culminating in a major focus on economic adjustment. The late 1980s saw a major reorientation in sectoral policies, to focus on social and environmental concerns. By 1992, further pressures for change had built. There was great turmoil about the Bank's alleged lack of compliance with its own environmental and social policies. From all quarters, reform was advocated and the Bank was urged to become more open, accountable, and responsive.

From all quarters, reform was advocated and the Bank was urged to become more open, accountable, and responsive.

Hence the ensuing realignment of corporate strategy, quality management practices, and evaluation processes. In other words, the foundations for the recent changes were laid during the Preston presidency, when the "50 Years Is Enough" campaign was already brewing. The Portfolio Management Task Force (Wapenhans) report was sent to the Bank's Board in September 1992 and led to a major drive for better quality management.

At that time, as VP for Planning and Budgeting, I interacted with the Wapenhans Task Force and seconded some of my operational planning staff to the task force. But the pivotal actors who helped design the quality management drive included Joanne Salop, who contributed to the diagnostic of the report, and Jim Adams, Director of Operational Policy, who put together the management action plan, "Next Steps," that the Board adopted in July 1993.

Organizational change accelerated in the mid-1990s. Historically the Bank had sought to adapt to its operating environment through frequent, albeit disruptive, reorganizations. But not since Robert McNamara was change embraced so eagerly as under Jim Wolfensohn. From the very start, he worked to make the Bank more oriented to learning and knowledge; more focused on quality; more responsive to its owners, borrowers, staff, and development partners in the private sector and the civil society, and more concerned with governance and institutions.

Inducing change in the Bank has been akin to sailing against the wind.

This then was the operating environment that induced new priorities for evaluation.

Organizational history is not a linear process, especially in a large and complex institution subjected to a wide range of external demands. Ideas and people drive change, but it takes time to nurture consensus, build coalitions, and induce the multiplicity of decisions needed to shift corporate agendas and business processes. Hence, inducing change in the Bank has been akin to sailing against the wind. One often has to practice triangulation and adopt a twisting path in order to reach port.

The Bank's Board played a very important role in the change process. In the early 1990s, given the declining trend that OED had documented in the development effectiveness of Bank-financed projects, executive directors sought to strengthen portfolio management and evaluation. In February 1993, four executive directors proposed to graft on to OED a quality management and inspection unit. That proposal was opposed by management and did not secure majority support from the Board. Eventually it was superseded by an even more exacting scheme driven by international NGOs, for an Inspection Panel to investigate the Bank's alleged failures to comply with its own policies. Management and the Board debated the Inspection Panel proposal at length; it was eventually endorsed in the context of an IDA replenishment negotiation, and the independent Inspection Panel was created in September 1993.

Another landmark was the creation of the *Committee on Development Effectiveness* in December 1994, through the bifurcation of the Joint Audit Committee. Ruth Jacoby assumed the CODE chairmanship and was instrumental in the design of Board oversight processes connected to development effectiveness. She led CODE until September 1997 and was succeeded by Surendra Singh, who served until November 1998. By then the OED renewal was well under way.

The renewal received sterling support from Board members Jan Piercy (from November 1998 to October 2000) and Pieter Stek (from November 2000 to the present). Each had a distinctive impact on the evaluation function. Both exercised great skill in the delicate tasks of overseeing evaluation, building consensus among committee members, and regulating the delicate interface between independent evaluation, self-evaluation, and management decisionmaking.

The creation of the Bank's Quality Assurance Group (QAG) soon after Mr. Wolfensohn's arrival in 1995 was another critical development. It symbolized management's determination to strengthen self-evaluation and quality management in the Bank. It was triggered by an OED report on the quality at entry of Bank operations that led Mr. Wolfensohn to set up a small task force, on which I served along with Gautam Kaji and Armeane Choksi.

The task force recommended the establishment of QAG. QAG's evaluation criteria drew on OED's prior studies of quality at entry and portfolio management. But the adoption by QAG Director Prem Garg of innovative, cost-effective, and businesslike methods to implement just-in-time evaluations proved the key factor in promoting quality management.

Finally, the Strategic Compact served as an indirect but powerful agent of change for OED. It created new needs for Board oversight and monitoring and generated new operational emphases. A new evaluation strategy became necessary. Such a strategy was endorsed by CODE in March 1997, and provided the blueprint for the OED transformation that followed.

This brings me to the third part of learning: the implementation of the renewal itself.

Within this complex force field, the OED renewal articulated five strategic objectives:

- Move evaluation to a higher plane
- Shorten the feedback loop and fill evaluation gaps
- Build evaluation capacity
- Invest in knowledge and partnerships
- Manage for results.

Conceptually, the design of the renewal relied on the prior findings of an Evaluation Learning Group composed of management and senior staff, commissioned by Gautam Kaji, Managing Director, Operations, and the DGO in August 1996. The Group was cochaired by the OED Director, Francisco Aguirre Sacasa, and by the Director of Operational Policy, Myrna Alexander. The main drivers of the painful but necessary changes within OED were Elizabeth McAllister, who assumed the post of director in May 1997, and Greg Ingram, who took over from her in May 2000. Both deserve enormous credit for the results achieved.

Participatory management was critical in turning the strategic objectives into results.

Participatory management was critical in turning the strategic objectives into results. A mission statement was crafted in consultation with all OED staff: *Enhancing development effectiveness through excellence and independence in evaluation.* We spent a great deal of time on defining our core values—integrity, excellence, teamwork, innovation and risk-taking, transparency, and a balanced work environment. We redefined the accountabilities of OED managers, launched a training program, and redesigned internal processes.

Next, the department was restructured. The hierarchy was flattened and we eliminated the DGO front office. We enhanced connectivity with the Bank's Regions and networks through the appointment of coordinators. We made the operating structure more flexible and we emphasized the importance of teamwork in personnel evaluations. The three divisional cylinders were reshaped into four interacting groups, including two dedicated to corporate oversight and to knowledge management and evaluation partnerships.

Finally, we improved the diversity and the gender balance of OED. It is fortunate that we did so since, according to Chekhov, "men deprived of the company of women become stupid."

The first of the strategic objectives—move to a higher plane—was driven by the comprehensive development agenda adopted by the Bank. It paralleled the shift of the Bank's privileged unit of account from the project to the country assistance strategy. The move to the higher plane called for systematic evaluation of country assistance strategies and for feeding evaluation findings into the Bank's sector strategy papers. To implement the new evaluation emphases, and in recognition of the improved quality of implementation completion reports, CODE embraced OED's recommendation to reduce the performance assessment ratio from 40 percent to 25 percent of operations in 1997.

The second objective—shorten the feedback loop—complemented QAG's tracking activities and connected evaluation producers more systematically to the Bank's business processes. Country assistance evaluations were sequenced to feed into the CAS cycle. Sector evaluations were geared to the production of sector strategy papers. The *Annual Review of Development Effectiveness* was timed to influence the Bank's Strategic Forum, and the *Annual*

Report on Operations Evaluation was reconfigured to feed into the corporate COSO cycle. The introduction of a Management Action Record process responded to the same logic.

A corollary objective was the filling of evaluation gaps. Independent assessment of economic and sector work is now undertaken as part of country assistance evaluations and sector evaluations, and it has been taken up by management under the "Fixing ESW" initiative. ESW evaluation now falls squarely within the QAG mandate.

The third objective—build evaluation capacity—carried forward the work of an earlier evaluation capacity development task force led by operational staff (Anil Sood) and supported by OED (Pablo Guerrero). The task force completed its work in July 1994. The substantial progress achieved since then has been outlined in the first annual report on evaluation capacity development (ECD), produced this year. In addition to its catalytic role and operational support, OED has sponsored an international program for development evaluation training in collaboration with Carleton University of Ottawa. In partnership with UNDP, it has also helped to create an International Development Evaluation Association, which was launched in Beijing in September 2002.

The fourth objective—strengthen knowledge and partnerships—was far-reaching and multi-faceted, and aligned with the Bank's emphasis on both aspects. It was implemented by:

- Organizing biennial conferences on evaluation and development
- Creating a state-of-the-art Web site
- Taking the lead in the development effectiveness section of the Development Gateway
- Designing and implementing more open evaluation disclosure policies (August 1993 and August 2001)
- Creating a new look for OED products, diversifying OED products through a mix of user-friendly publications (Précis, fast-track products, Lessons & Practices, working papers, books)
- Providing leadership to the rest of the development evaluation community through proactive participation in the Evaluation Cooperation Working Group of the Multilateral Development Banks and the Working Party for Aid Evaluation of the Development Assistance Committee
- Connecting to the evaluation organizations of the Swiss Development Corporation, the UK Department for International Development, the Dutch and Norwegian Ministries of Foreign Affairs, and the UN Development Program through formal partnership agreements
- Partaking in parallel evaluations with the African, Inter-American, and Islamic Development Banks
- Taking the lead in a multi-stakeholder evaluation of the Comprehensive Development Framework.

The fifth objective—manage for results—was largely a matter of practicing what OED preached, that is, by subjecting OED itself to results-based management processes, including tracer studies of OED products, client questionnaires to gauge reactions to evaluation products, and internal staff surveys to guide OED administration and internal management.

In parallel, a participatory process of programming and budgeting was put into place to connect evaluation priorities to OED staff views, management concerns, operational needs, and Board oversight requirements. We strengthened quality assurance through one-stop reviews for all approach papers and major reports. Entry and/or exit workshops were organized to ascertain the diverse perspectives of borrowers, voluntary organizations, and private sector partners. Finally, for most major studies, advisory panels were set up to inject expert guidance from eminent development personalities and specialists, and increased efforts were devoted to outreach and connectivity throughout the evaluation process and beyond.

What have been the results?

When asked by Henry Kissinger what he thought of the French Revolution, Chou En Lai famously replied that it was too early to tell. Similarly, it may be premature to attempt a definitive assessment of the OED renewal.

Still, it may be worth noting that the qualitative changes achieved by the renewal did not undermine the volume of output. Under my tenure, OED has produced each year an average of 279 evaluation summaries, 92 performance assessment reports, 7 sector evaluations (including sector and thematic evaluations and country sector reviews), 5 country evaluations, 8 impact reports, 2 process evaluations, 2 apex reports and 58 secondary products (Précis, Lessons & Practices, Fast Track Briefs, workshop proceedings, working papers, and the like). This is in addition to guidelines, speeches, and books, and comments to the Board/CODE on pertinent matters. This represents a substantial contribution to the Bank's development record, a distinctive corporate asset.

Of course, output does not mean outcome, let alone impact. But over the last five years, the relevance of OED's work has improved markedly. Specifically, OED has delivered country assistance evaluations for a majority of active borrowers. Sector evaluations have been issued to guide most sector strategies. Policy evaluations have shaped the recasting of all major safeguard policies. A comprehensive assessment of IDA 10–12 was delivered in the context of the IDA 13 replenishment. And the recent review of the Bank's global programs provided timely input for the Board deliberations on the budgeting of grants. Evaluations of the Heavily Indebted Poor Country Initiative (HIPC) and of the Comprehensive Development Framework are soon to be issued.

Finally, the positive trend in the performance of Bank-financed projects is largely due to the efforts of Bank managers and operational staff. But QAG, OPCS, other evaluation and control units, and OED have contributed handsomely too, since, to paraphrase Tom Peters, "what gets evaluated gets done better."

No less relevant have been OED's efforts to promote improved evaluation methods throughout the development community. By now, development evaluation has found its place in the sun as a legitimate discipline within the mainstream of the evaluation profession.

What lies ahead?

According to Kierkegaard, "life must be understood backwards, but it must be lived forwards." In today's volatile operating environment it is not practical to forecast what lies ahead for OED. But, as Peter Drucker often reminds us, "the future is already here." With Board guidance, a number of emerging challenges have begun to be tackled. This effort may need to be intensified, given the disappointing trends in development that bring to mind H. G. Wells' quip that "human history is more and more a race between education and catastrophe."

First and foremost, OED, working closely with DEC, should attend to a large amount of unfinished business in development policy evaluation. OED must continue to probe the issues underlying the current public debate about aid effectiveness. This is critical, since growing public skepticism about aid has contributed to a decrease in aid flows, despite escalating demands for resources to meet complex humanitarian emergencies and to implement an ever-expanding development agenda.

OED must continue to probe the issues underlying the current public debate about aid effectiveness.

Many countries that have implemented Bank/IMF advice are not doing as well as expected. What explains the sluggish growth of developing countries despite major improvements in their economic management? Were the reforms advocated too timid? Was the implementation of reforms less impressive than commonly believed? Was the sequencing of reforms ill planned? How can growth be accelerated without exacerbating environmental stress?

Evaluation and research have established that aid works best in combination with good governance and improved policies in developing countries, but this begs the question of what to do in the majority of countries where the right enabling environment is not in place. In such a situation there is no substitute for the project instrument. To help illuminate risk-reward decisionmaking and the centrality of side effects in project results, highlighted by Albert O. Hirschman, OED is poised to recapture its prior intellectual leadership in impact evaluation research.

Too little is known about the role that rich-country policies play in the fortunes of developing countries. Could it be that without larger debt reductions, larger private sector flows, more trade liberalization, more migration, more effective environmental policies in the North, and a freer flow of technology and knowledge toward the zones of turmoil and transition, more aid will not be enough to achieve broad-based poverty reduction? If DEC and OED do not join forces to deal with such issues, who will?

Second, there is no turning back from focusing on country assistance evaluations, which Ruben Lamdany did so much to create. Tailor-made country assistance strategies have become the main instrument of corporate management. In post-conflict countries, the Bank is called upon to fund rehabilitation programs and help restore institutions. In middle-income countries, the Bank complements the macroeconomic role of the IMF and plays a counter-cyclical funding role to help manage financial crises; to facilitate coherence in external support, the Bank is called upon to validate the quality of structural, fiduciary, and social policies. In low-income countries, the Bank is playing a central role in debt reduction and in assessment of poverty strategies, and its role at the country level is now much more important than simply lending. How well is the Bank doing in all these areas? The recent establishment of the IMF's Independent Evaluation Office offers an opportunity for close interagency collaboration here.

If DEC and OED do not join forces to deal with such issues, who will?

Third, sector strategy evaluations will need continuous upgrading. The recent emphasis on social sectors and the environment marks a welcome rebalancing of policy priorities, away from narrow parochialism of sector specialization and toward a more holistic view of development. The belated recognition that ownership and partnership matter has been a useful counterweight to donor-driven conditionality, and raises the corresponding evaluation challenge of evaluating partnerships.

A complementary challenge will be to promote a results-based focus to drive intersectoral interactions while pursuing the Millennium Development Goals. It will also be critical to focus on the winners and losers from current policies to ensure that the Bank's development activities retain a human face. Mainstreaming the safeguard policies at the country level is the great operational challenge of the next decade, and OED must continue to bring field evidence to bear on the outcome.

Of equal importance is the systematic tracking of the rural development and private sector strategies, since they will largely determine the prospects for poverty reduction. OED (working in concert with the Operations Evaluation Group in IFC, OEG, and the newly independent Operations Evaluation Unit in MIGA, OEU) must assess how the emergence of globalization, information technology, and knowledge should be taken into account in the implementation of private sector development strategies. With benefit of time, future evaluation history events will focus not simply on OED's history but on the joint history of OED, OEG, and OEU.

Fourth, new evaluation frontiers will have to be explored. Beyond its country assistance and sector strategy evaluation roles, the Bank has become a platform for collaborative programs focused on the provision of global public goods. A bewildering variety of global partnerships have sprouted to deal with cross-country development challenges. Here too, evaluation will have to be retooled and reoriented to fill evaluation gaps on the global plane. Equally, OED will need to become more active in promoting performance measurement as the platform for systemic evaluations cutting across all development assistance agencies.

Since the Monterrey Conference on Financing for Development (March 2002), an emphasis on results is being advocated as a remedy for the numerous problems faced by the aid industry. Monitoring has indeed been neglected, but improved monitoring can only go so far in improving aid performance. Just as serious is the gap in program evaluations. In particular, joint and country-based evaluation will have to put a sharper spotlight on such systemic problems as lack of coherence in aid practices, a proliferating number of aid givers, excessive transaction costs, and the misallocation of resources created by development fashions and tied aid.

This is an ambitious agenda, but OED has the skills, the momentum, and the motivation to deal with all of these questions. As Yves Rovani said on a similar occasion, OED has made considerable progress in climbing the mountain of evaluation excellence, but we have not yet reached the top. The air is brisk, the road ahead is clear, and I have no doubt that the leaders selected to manage the independent evaluation function for the next leg of its journey will bring OED to the top.

7

OED Renewal: Participation, Partnership, and Results

Elizabeth McAllister
Director, Operations Evaluation Department, 1997–2000

By 1997, the world around OED had changed dramatically. "Fifty Years is Enough" had been a shattering experience and message for the Bank at the 1996 Annual Meetings in Spain. In April 1997, President James Wolfensohn came forward with the $250 million Strategic Compact to address the criticisms of the Bank, and to deliver a fundamentally transformed institution that was quicker, less bureaucratic, and more responsive to client demands and global development opportunities. The Bank was to be a knowledge bank working in partnership with donors, clients, and organizations. By the 1997 Annual Meetings, Mr. Wolfensohn was reframing the development paradigm as a compact in itself, between the rich nations to give more and do more, and the poor nations to create an enabling environment for change—all with a view to creating a systemic, global impact to reduce poverty. His vision implied a major change in how the Bank would engage its partners and its clients. It highlighted a more comprehensive

Box 7.1. OED Renewal — A Three-Part Story

- Hearing from the Authorizing Environment
- Setting the Vision, Strategy
- Launching Organizational and Cultural Change

framework for understanding development that valued the role of the social, cultural, and economic sectors and the need for partnership among all the key actors in these sectors to sustain progress.

The Director-General of OED, Bob Picciotto, knew that for OED to be relevant, it would need to align itself with the Bank's new orientation. OED could no longer focus only on the project as the "privileged unit" of development intervention. It needed to connect to a broader development agenda and reflect new, more ambitious corporate priorities. Bob quickly crafted a new strategy for OED—top down—for the Board of Directors' approval. Staff criticized their lack of participation in setting OED's strategic directions. But in terms of organizational change theory, such an approach is not without its proponents: leadership sets out the vision for the new organization and bottom up, staff lead on how to make it happen. This is the approach that worked for OED. Many staff were part of the "authorizing environment" that needed convincing and Bob needed to position OED quickly if it was to be a player in the knowledge Bank. The new strategy proved to be not only timely, it was resilient and—I believe—on the mark.

The central goals of OED's renewal were the alignment of the strategy with a supportive authorizing environment and a high-performance organization.

Bob invited me to join OED from the Canadian International Development Agency in May 1997, to help implement the renewal program for OED. My appointment was a signal of the change he envisaged—partners and external viewpoints would be valued in the new Bank. The central goals of OED's renewal were the alignment of the strategy with a supportive authorizing environment and a high-performance organization. The strategy behind our strategy was a sustained, persistent commitment to these few and clear strategic objectives, to expanding our client reach, and to developing new patterns of organizational and individual behavior that would be sustained and perfected over three to five years.

This presentation aims to give you the story behind the story. For those of you who are denizens in the current OED home, I hope it helps you to value what you have. It may also help you to think about how you might make the changes that this decade will demand of OED.

A key feature of OED renewal was the use of stakeholder feedback. Our stakeholders believed that our products needed to create opportunities for learning, "platforms for debate." Managers wanted real-time advice. In OED's October 1997 retreat, we invited the Bank's President, selected executive directors, managing directors, and vice presidents to give us their feedback as key clients. Along with the Billson report and four groups led by OED staff, we heard that we were the best evaluation group in the world. Our evaluation criteria and our methods were clear and had created trust in our work. We had a committed staff, and we had gained a level of independence attained by no other evaluation group in the development community. But though our products were of high quality, the world had moved on, and we were missing the bigger picture. Our lessons had become repetitive. Our products arrived too late to make a difference, and we were "a fortress within the fortress." I remember Caio Koch-Weser, a managing director, saying OED was "the history department of the Bank"—a tough message for committed evaluators to hear. Executive directors called for better dissemination and clearer lessons and closer partnerships with clients.

Box 7.2. What We Heard by 1997...

- Best evaluation group in the world
- Committed, independent staff
- High-quality products

but...

- Missing the big picture
- Repetitive lessons
- Too late to make a difference
- "A fortress within a fortress"

Bob and I used the October 1997 retreat to deepen appreciation for the new OED strategy, starting with a compelling presentation on the new global order given by Keith Bezanson, who heads the International Development Institute at Sussex University. Bob led staff through his thinking behind the five strategic directions at the heart of the new OED strategy. Since that retreat, OED's management team has been consistent in using these five goals as a compass to guide OED's progress.

Of the five strategic directions, I believe the most important was *moving to a higher plane*. Led by Ruben Lamdany and Greg Ingram, OED's innovation and leadership on country evaluations and thematic and sector evaluations had the effect of raising the level of accountability from the task manager to higher managerial levels in the Bank, while changing the architecture of OED products. Along with Wendy Jarvie, Uma Lele, and Nagy Hanna on the ARDE and global evaluations, "moving to a higher plane" had the effect of hot-wiring OED into the policy agenda of the Bank. The country assistance evaluations were timed to feed into new country assistance strategies and sector evaluations into the sector strategy papers. This alignment meant that

Box 7.3. Conceptualizing the Strategy

- Move to a Higher Plane
- Shorten the Feedback Loop
- Build Evaluation Capacity
- Invest in Knowledge and Partnership
- Manage for Results

report deadlines were absolutely binding. We could no longer let a product slip into the next fiscal year to be perfected. By aligning ourselves with the Bank's policy agenda we created the necessary internal incentives to produce client-oriented and timely products.

This move precipitated healthy debates over whom OED's clients should be: the Board of Directors, senior management, or task managers. It has always been clear that OED's number-one client is the Board, but a consensus was reached that senior management, vice presidents, sector leaders, and country directors are also extremely important clients, because they ensure that our evaluation findings affect decisionmaking on the ground. Working to inform their work created the engagement and the "platforms" for debate and learning that we sought.

We *shortened the feedback loop* largely through inclusion and participation. This was an extremely important change. Starting within the Bank, we launched a series of annual consultations with the operational vice presidents to determine their learning and accountability preoccupations. We used these consultations to shape our work program. As good evaluators, OED staff thought hard about the threats to our independence we could be creating. We decided that as long as the Board held the authority over our budget and work program, based on our recommendations, our independence would not be jeopardized.

OED also increased participation to shorten feedback by involving partners, clients, staff, and civil society in entry and exit workshops that were meant to inform the design of our evaluations and to quickly inject our findings into the work of practitioners. In designing a study, the evaluator has scope to choose from among a number of key questions to guide the inquiry. Our choice of key questions based on the preoccupations of end-users of the evaluation, and of the Bank's assistance, increased the demand for our products. It was quite memorable when the vice president for East Asia asked us to find out what in the culture of program management in Indonesia was preventing the truth from reaching the top. That was an unusual and a brave question, which pushed the boundaries of the country evaluation to include internal organizational influences.

Build evaluation capacity development. OED approached capacity development by developing training and outreach programs and by pursuing joint evaluations with other partners and with country clients. In preparing for this presentation today, I called our friends in the OECD Development Assistance Committee network of bilateral evaluators to ask for their views on what OED accomplished with its renewal. One colleague said that through evaluation capacity development, OED has established development evaluation as an international profession; that it has internationalized evaluation. "Development evaluation is no longer an adjunct to the American Evaluation Society or the European Evaluation Society, but it is now a multicultural and multidisciplinary discipline in its own right."

By working in *partnership with* evaluation units in other development assistance agencies in client countries and with some large NGOs, OED has given legitimacy to these units. The joint work helped to reinforce the use of evidence-based evaluation and thus evidence-based decisionmaking as more of a norm in public administration.

Partnership and participation in evaluation expands the learning created in the course of an evaluation to other organizations. It also enhances the Bank's reputation by exposing the Bank's activities to the assessment of others. In many ways, by inviting civil society and the private sector to participate in our evaluations, OED increased the legitimacy of the Bank through increased transparency.

This was even true in unexpected events such as in the case of the Indonesia country assistance evaluation. We faced tremendous anger from our colleagues for a leak of an early consultation draft that happened in the Region itself. The anger abated when there was a call from *The New York Times.* One of the paper's senior journalists said to senior management that he was impressed that the Bank supported an institution within the institution to speak truth to power. He apparently noted that the Bank was modeling the good governance we were expecting from our clients. I would submit that the openness of our forestry and gender evaluations also enhanced the Bank's transparency and thus its credibility. This is a lesson that many in the corporate world are just now learning as they face the increased scrutiny of transparency networks.

The Bank was modeling the good governance we were expecting from our clients.

The Swiss Development Agency was an important partner helping OED's renewal. They provided moral, intellectual, and financial support to the "opening" of OED. They participated in our studies and funded our entry and exit workshops, publications, and numerous participatory events that were used to bring developing country perspectives and civil society research and opinion into our analysis. I will be forever grateful for their partnership, leadership, and friendship.

The fifth element of OED's new strategy was to *manage for results*: OED moved from managing outputs to outcomes. Staff working on our Unit Compact argued that an evaluation report was an output, not a result. An evaluation department achieves its purpose through the use of its reports. We began to measure our own effectiveness by indicators that reflected the use of our studies to shape decisions. Setting targets for product use and timeliness provided incentives to spend more resources on the dissemination and communication of our evaluation findings: publications, Web sites, the Help Desk, and exit workshops. A good example of the new OED was the

Fast Track Brief. These products are consistently short and crisp: within three or four bullet points they give key decisionmakers the main conclusions or messages of the report. I remember Dennis Whittle saying at our 1998 retreat that people don't read old-fashioned reports anymore. In the Information Age, reports need to be layered for easy absorption. It is tricky to be brief without being platitudinous, but OED has done it, and I encourage you to keep at it.

OED's Unit Compact and its report on the attainment of its targets in the *Annual Report on Operations Evaluation* (AROE) is a best-practice example in the Bank of managing by results. The *Annual Review of Development Effectiveness* (ARDE) and AROE encouraged the Bank to adopt a results-oriented approach for managing and measuring performance as good real-time management practice. In this area, we were less successful, and the best that can be said is that we created a readier environment then might have been when external pressure mounted for the Bank's move to results management.

Box 7.4. Launching a New Organization

- New Structure and Processes
- New Culture
  - ACS staff as players
  - Evaluators as facilitators, partners, while retaining profession ethics
  - Leaders as advisors
- New Values

Our results approach included attention to cost effectiveness. OED's renewal created new product lines and processes without a major increase in budget, and while maintaining the same staffing levels, despite the conversion of non-regular staff to regular appointments.

Structure

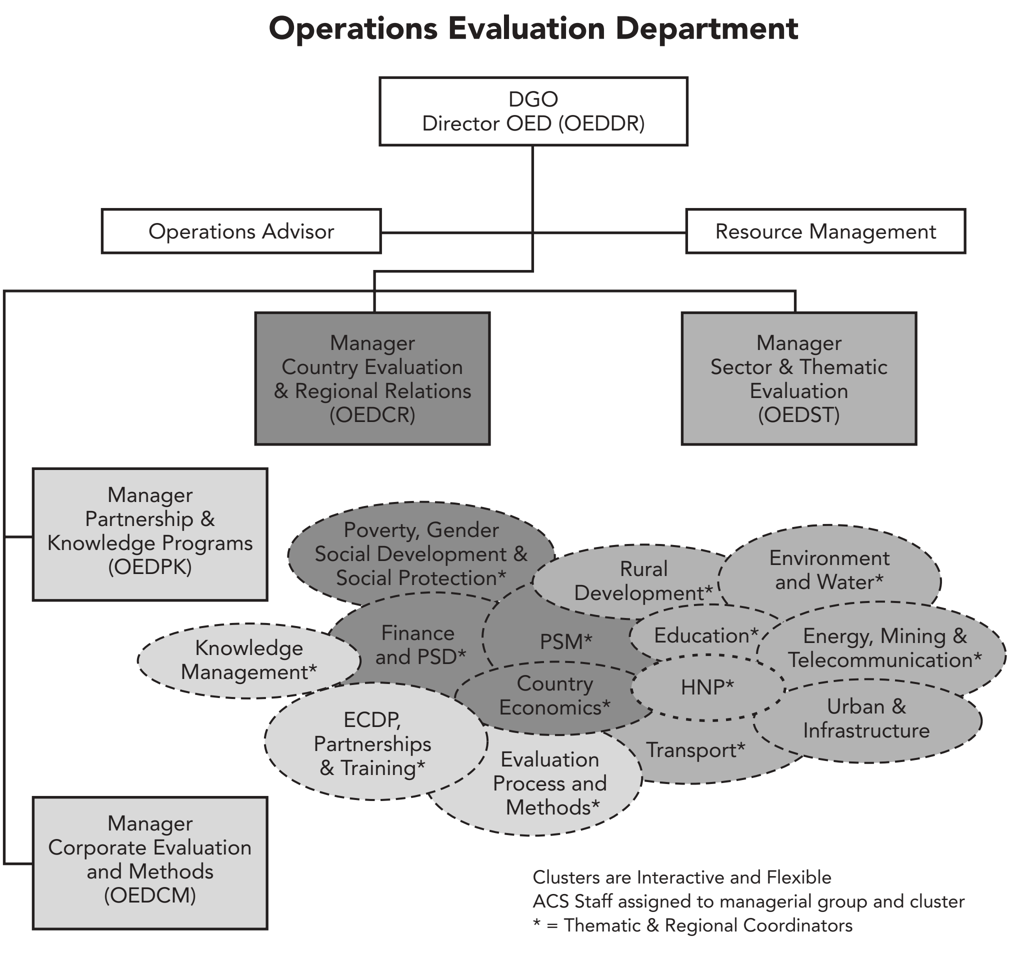
A very different organization was needed to deliver on our new strategy. Building on the work of the four staff groups, the new OED emerged in June 1997 after a three-day "work out" with staff and two managers. The new organizational chart (figure 7.1) reflected the goals of leadership at all levels; sector coordinators were meant to be global leaders, linking with the Bank's new networks and nurturing global policy networks. Regional coordinators would need deep Regional knowledge. We wanted delayering and teamwork across OED in recognition of the importance of knowledge sharing and the multi-sectoral approach inherent in the new vision of the Wolfensohn Bank.

Products and Processes

We started refocusing our products so that each project audit linked into a country and/or sector evaluation, which was linked into our major thematic evaluations, which were building blocks for the ARDE. And we also focused on linkages across program years to get a more efficient use of our resources and to ensure greater teamwork across OED.

As I have covered earlier, we changed the way we did evaluations. New activities included one-stop reviews instead of serial management sign-off for draft reports, entry/exit workshops, global consultations with civil society, and the creation of advisory

Figure 7.1. OED Organization Chart, 1997



groups for our large studies to bring the private sector, donor, NGO, and U.N. perspectives to bear on our analyses. We said to our new collaborators, "We're inviting you because we would like a broader range of perspectives on what constitutes development effectiveness. That said, the rules of the game are that you don't use your participation here as a lever for lobbying the Bank, and you don't use it as a communication line to your constituency. The quid pro quo of that is that you get to write your own section if you disagree with the findings of the evaluation."

Culture

In redefining OED, we understood the importance of culture. For example, we recognized we were not getting the full leverage from a professional relationship with the ACS staff. We supported two ACS-only retreats, and encouraged the ACS staff to take leadership in looking at what kind of new skills they would need and the rest of us would need in order to leverage ACS staff as players.

In nurturing a new culture, we were asking for a change in our staff's professional and personal lives. The role of the evaluators changed. They acquired new skills, new attitudes, new ways of thinking about work, about development, and about the role of the Bank. We were challenging OED staff to connect and to join, to provide rigor and systematic thinking in the fluid environment of development partnerships. We were inducing them to connect with the angry leaders of civil society with a view to breaking the impasse on such issues as the building of large dams, by providing credible evidence about tradeoffs, constraints, and achievements. We were asking them to join with others who have their own ideas about success and good performance. We were asking them to move from introverted research to extroverted facilitation—to guide systemic learning for effective policy dialogue and public engagement. And in making these deep personal changes, Bob modeled for all of us his own attempts in a very sharing and public way, and I think that made it easier for the rest of us to go deep inside to think about the changes that were necessary, if we were going to be a relevant and dynamic learning center for the Bank.

I think one of the toughest but most important cultural changes was the recognition that top leadership should be advisors. Here we were influenced by Terry Barnett, who helped us try to change from the command and control mode of leadership to a management-staff relationship based on providing and accepting advice.

We brought values to the forefront. Linda Dove and Salman Anees, who is here today, led a Values Working Group to define the values that needed to underlie our work with others and our relationships among ourselves. Once accepted by staff and management, the values were managed through communications, surveys, and training.

Staff groups developed all the major pieces of the renewal infrastructure. Values, as I said, was Linda Dove and Salman's group. On the accountability statements, Bob and I worked hard to redefine the DGO and director roles, Greg Ingram led the managers in clarifying their accountabilities, Becky Alvaro-Flores and Betty Bain developed the front office and ACS roles with staff, and the coordinators defined the task team leaders' accountabilities. Ruben Lamdany and Wendy Jarvie developed the criteria for excellence. We invited an external evaluator, Greg Jackson from George Washington University, to use these criteria to measure the quality of a randomly selected sample of our major studies, every two years. Patrick Grasso led the development of OED's knowledge management and staff training strategies. The new OED Strategic Management Framework was published as I left in 2000; it represents the renewal work of the entire department.

What also made the renewal a success was the confluence of a number of events and people. At the Board we had a very strong supporter in Jan Piercy, the U.S. Executive Director and chair of the Board Committee on Development Effectiveness at the time some of these changes came to fruition. Particularly on the forestry study, when an NGO used our consultative process on the penultimate draft for a media event—in a breach of trust—Jan was very supportive in countering the reaction from some senior managers who said we were providing sticks for the Bank to be beaten with. Another critical success factor was the new management team, which was fully engaged and supportive, starting with Roger Slade, then Wendy Jarvie, Osvaldo Feinstein, and Ruben Lamdany. Greg Ingram replaced Roger and brought tremendous talent and good humor when we needed it most. The staff participation, involvement, and dedication in a very demanding time was tremendous. And throughout, Betty Bain kept us strong and focused!

Box 7.5. Moving Forward

Partnership

- Stay the Course

Participation

- "Only OED gets it"... share lessons
- Nurture networking

Partnership

- Case study OED performance management

To give OED some quick advice for moving forward:

On partnership and joint evaluation, stay the course. I know it is time-consuming and it's hard to work with other agencies. The same evidence often leads to different conclusions, depending on one's evaluative and normative bases. Sometimes we may need to accept that we might not conclude with a joint product, but we can still celebrate that 80 percent of the process is shared, because the shared process reduces the burden on the client and increases triangulation.

On the International Development Evaluation Association (IPDET), and Partnerships and Knowledge Management (PK) Evaluation Capacity Development . . . again, stay the course, and ensure they don't compete. It's important that these initiatives have an institutional home, and that all three work together.

The president of a major NGO said, "Only OED gets it. Only OED really understands how to work with civil society."

On participation, once, shortly after I left OED, in a meeting with Jim Wolfensohn and the heads of major NGOs, the president of a major NGO said, "Only OED gets it. Only OED really understands how to work with civil society." It was then that I realized how much we'd learned in OED and how far we had all come. It would be helpful for you to chronicle OED's partnerships to help others in the Bank learn from OED what you've learned.

Use every opportunity to nurture networks. For the aid coordination study, we brought in the heads of aid coordination, knowing that people in these functions rarely have an opportunity to talk to one another. In organizing the workshop, we put emphasis on creating an international network of those individuals so they could provide support to one another after meeting through the workshop. Success is in the details: having nameplates so people remember each other's names and faces; ensuring e-mail address lists are prepared for the participants.

On managing for results, OED's forward pass is now being caught in Strategy & Resource Management, as we work with the Bank on reinventing the unit compacts as contracts for results between the vice presidents and the managing directors. This concept is rooted in OED's work on its unit compact as a contract with the Board; which is a best-practice example of a chart of accounts linked to outputs and outcome targets and indicators.

In particular, as I read Bob's last AROE, which used various methods to test OED's success relative to its Unit Compact, it was clear that OED is the unit in the Bank that is most advanced in understanding the role that results-based management can play in both improving performance and in demonstrating the effectiveness of the unit's contributions and influence. It would be extremely helpful to share the experience with those other Bank units that are grappling with what managing for results really means.

Thank you for the opportunity to return with my predecessor, Eberhard Köpp, and share the renewal story with you.

QUESTIONS AND ANSWERS

Patrick Grasso: Bob McNamara made the argument that if we had been able to get all our evaluation materials out much earlier, we would have stemmed a lot of the criticism of the Bank. Is that a fair statement?

Ms. McAllister: This isn't an uncontroversial view in the Bank, even now, because it is natural to try to position your organization in the best possible way. For example, you promise low and deliver high, versus promising high so as to create incentives; or you only tell what's good (or you try to change everything into a good story), so as to keep the heat on the outside; versus acting as a heat conductor, bringing criticism within and sharing self-criticism without—all in the effort to improve development and to show that any institution can learn. There are huge debates on this within the Bank and in other public and private institutions. But in this day and age it is very important for an institution to share what it is learning. And if you tell people what goes wrong, they'll believe you when you tell them what goes right.

Ridley Nelson: Given that one of our tasks in OED is rating, how do we know we're maintaining consistent standards? Performance standards have changed. Capacities have changed. Behind what we do, we should have some consistency of standards over time. Otherwise, the apparent ups and downs of Bank performance don't mean anything at all.

Ms. McAllister: Indeed, that's the job of meta-evaluations, and part of OED's role is to be constantly evaluating its evaluations to make sure that its standards have remained consistent. The other answer is that evaluators should be looking more and more towards outcomes and what's actually happening on the ground, rather than looking at quality at entry or quality of analysis. We should be making the client more powerful in evaluating the institution; we should measure our effectiveness in terms of clients' outcomes—targeting clients' needs and measuring whether they have been met. The Bank has to get better at defining whom we are trying to influence and why, what behavior we are hoping to change, and what indicators of behavioral change we will use. The closer we get to assessing outcomes and results on the ground, and to being able to measure outcomes, the easier it will be to judge whether or not our own standards of evaluation have changed.

PART II:

Relations with Board and Management

To be effective, OED must influence the actions of those responsible for Bank developing and carrying out Bank policies and programs. Crucially, this means getting its message across to the Bank's Board of Executive Directors, especially its Committee on Development Effectiveness, and senior Bank management. Only if these officials are convinced of the soundness of OED's evaluation findings and recommendations are they likely to use them to affect the Bank's development work. So the views of Board members and senior managers are critical to understanding the effectiveness of OED.

The papers in this section were presented at a major forum held in September 2002. The forum featured presentations by the then-current chair of the Committee on Development Effectiveness, a Bank managing director, and a panel including a former member of the Board, the Bank vice president responsible for developing operational policies for much of the 1990s, and the director of the Bank's Quality Assurance Group, responsible for real-time evaluation of ongoing projects and other Bank activities.

In broad overview, these papers highlight the importance of OED's independence as a critical attribute leading to its effectiveness in promoting accountability for and learning from Bank work. They also suggest that OED's effectiveness has been further enhanced in two ways. First, all of the forum speakers note that OED has adapted its work to meet new challenges. This includes constantly updating its methodology; moving from project evaluation to the higher plane of country, sector and global evaluation; and improving the timeliness and utility of its products over time.

Second, a number of institutional changes have helped to amplify OED's work in ways that make it more immediately relevant and useful. In this regard, several of the papers note the importance of the creation of the Committee on Development

Effectiveness in 1995, and of the Quality Assurance Group in 1996. Committee discussions of OED reports provide an opportunity for the Board, management, and OED to share the learning from evaluation and consider how it can be used to improve the outcomes of the Bank's work. The Quality Assurance Group is a key mechanism for applying the knowledge from evaluation to consideration of ongoing Bank work in a timely, consistent way.

The authors of these papers all discuss a number of challenges facing OED as it moves forward, including most importantly the contribution of evaluation to imbuing the Bank with a results-based focus on the outcomes and impacts of the Bank's work.

8

Promoting a Culture of Accountability and Learning

Pieter Stek
Former World Bank Executive Director and Chairman, Committee on Development Effectiveness

I've had the privilege of being a member of CODE for four years and chairing it for two years. This experience has made me marvel at the governance of the World Bank. It is as if Montesquieu had stood at OED's cradle 30 years ago. Surely, he would have helped to originate what we got at that time: a complete Trias Apolitica consisting of management—the executive arm—and the Board—the legislative and supervisory arm—both already in existence, but requiring for balance and performance an independent evaluator rather than an adjudicator.

That is where we in the Bank have a necessary difference with Montesquieu's Trias Politica, applied to the state, the best example of which is the governance structure of the United States. Within the Bank, OED cannot and should not have the power of an adjudicator in a system of checks and balances. It cannot, because neither the Board nor the shareholders behind it would ever relinquish the power vested in them by the Articles of Agreement. It should not, because an adjudicator with court powers would lack legitimacy and erode shareholder support. More concretely, our system is more than one of checks and balances. At its best, it is one of mutually reinforcing interaction between the composite parts of the Trias.

Our system is one in which management, knowing that a process of objective and independent analysis takes place, has an extra incentive to promote effective development. It is one in which the Board and CODE are made aware of what counts for effectiveness, and are protected against fads. Indeed, they have an instrument that enables them to supervise, and it is one in which OED is stimulated to be relevant in order to be respected and listened to.

Our Trias at its best functions in this way. Each of the three bodies has a role in instilling and reflecting the culture of accountability and learning that is crucial in the institution. First and foremost, there is the president and his management team. Their example is essential. Under this president, faced as the Bank was with the challenges of mediocre quality, lack of system in approach to development, and lack of ear for the client, major forward strides have been made. OED has directly given management the incentive for much better quality control and for creating the institution built for this purpose, the Quality Assurance Group. OED did not bring in COSO, but it latched on very effectively to that framework, particularly the effectiveness and efficiency window of controls, and the management of non-financial risk. Moreover, OED has pressed, together with Controllers, for systematic self-evaluation by country and sector management.

If we do not learn from OED's analysis and digest it, we will ourselves be responsible for the choice of wrong priorities, and ultimately for inadequate performance.

For the Board and CODE to reflect a culture of accountability and learning, let alone to instill it, is difficult. OED is our essential helpmeet. If we do not learn from OED's analysis and digest it, we will ourselves be responsible for the choice of wrong priorities, and ultimately for inadequate performance. Some of us are more subject to political pressures than others. Political pressures are not by definition bad. However, they are a source of inconsistency and instability, and at times lack of quality. Some are powerful but ephemeral. In fact, being steered by political pressures is politicization, and I do not apologize for the tautology. OED should function as a built-in stabilizer, as nourishment for our conscience if we have one, as our weapon against arbitrariness, and discrimination between clients, which ultimately leads to cynicism. OED helps us to steer a sensible middle course between perfectionism in strategies and policies and operating loosely. OED is an instrument for helping the Bank to work at reasonable cost to the client in contributing to the client's development.

Let me digress briefly to give an impression of the extraordinary productivity of OED over the last six years. Together with QAG, OED can claim credit for the vast improvement in project quality. The *Annual Reviews of Development Effectiveness* from OED and the *Annual Reviews of Project Performance* from QAG reflect and have inspired essential changes, raising awareness of what counts toward success and suggesting the tools for achieving it. At the same time, OED has critically evaluated its own evaluation techniques and those that management applies in self-evaluation. In 1998, OED argued forcefully that the Bank should adopt results-based management. A further impetus to this has now been given by the central position of the Millennium Development Goals, and by the compact of Monterrey. The same may be said of the Comprehensive Development Framework and that other Trias that the CDF has established, of ownership, partnership, and participation. All of these create opportunities to increase development effectiveness, yet at the same time they also create challenges of measurement and attribution. We can without exaggeration state that OED has been leaping forward to the higher plane of evaluation that is essential for coping with these challenges. It has developed a methodology for evaluating country programs that involves triangulation from the top down, bottom up, and from the side. This methodology is reflected in the country assistance evaluations, which are thought-provoking and seen as tough because the rating of Bank performance is linked to outcomes rather than intentions alone. The sector reviews have acquired great depth, and have served to inspire the sectors in the Bank that have benefited from such a review to deepen their strategies and at the same time to give much greater attention to implementation. Forestry is an example. So are cross-cutting issues like the environment and gender strategies. Then there are the process reviews. Ongoing at the moment is a wide-ranging review of the Bank's role in the provision of global public goods.

Besides its own evaluative work, OED has formed strategic partnerships for evaluating development effectiveness with major bilaterals, the Netherlands, Norway, Switzerland, and the United Kingdom. On the side of the recipient countries, it has thrown itself into evaluation capacity building and knowledge dissemination and outreach.

Finally, OED has given great attention to its presentation. Not only does it write well, but it has created the *Fast Track Briefs,* which capture what is most relevant in the evaluations.

Let me also speak of the personal role of the DGO. He's been a major force, indeed the causa movens, in most of what I have described. Personally he has driven the extension of the trias of IFC and MIGA with the fully-fledged operation of IFC's Operations Evaluation Group and recently MIGA's Operations Evaluation Unit. Bob Picciotto has personally contributed to evaluation literature and has played a key role in strengthening the Bank Group's evaluation methods. Only two weeks ago, his brain child, the International Development Evaluation Association, IDEAS, came to fruition; this is a professional association focusing on evaluation in developing countries in partnership with UNDP. Bob has also been instrumental in creating the Evaluation Coordination Group, consisting of the evaluation units of the major multilateral development banks and now the International Monetary Fund, too. He has contributed to the DAC Working Group on Aid Evaluation and the U.N. Interagency Working Group on Aid Evaluation. His contribution in all these fields has been high-powered, fundamental, and pragmatic.

I have spoken of how management and the Board gain effectiveness by using OED to full advantage. Now to OED itself. How can this third body of the Trias fulfill its role at the level of excellence required for effectively instilling and reflecting a culture of accountability and learning? I already said that OED needs to be relevant to be respected. It not only needs to be right, but to be right about the right things. It needs to be seen as engaged on the same mission as the Bank by being an integral part of it and not an island unto itself. OED enables us to avoid mistakes by drawing the right lessons from the past. It is a normative historian. If OED had to be depicted by an artist as a body with a human face, that face would display tough love. That was, in fact, the quality that was regarded as indispensable by many members of OED when I interviewed them on the requisite qualities of Bob's successor. In my time, OED has performed outstandingly.

This is where I do want to say a few words on OED's much vaunted independence. We all know that independence is a relative concept. We can protect and ring-fence personal, functional, financial, and instrumental independence in ways that have been done pretty effectively for OED and its head, the DGO. Reporting to the Board through CODE is essential. So is the impossibility for the DGO to go back into the Bank. So is the disclosure of its products, et cetera, et cetera. What can never be legislated is the independence of the personality of the DGO. Over 10 years, this Bank has been fortunate—whether by serendipity or, I would hope, elective affinity—to have had a DGO who was not only independent in the formal sense, but in the essential sense of never flinching from presenting home truths, eye openers to those who are unblinkered, whose power of analysis ranks with the very best, whose power of judgment ranks him among the wise of this world, whose power of presentation makes evaluations, while at times unpalatable, at least recognizable as pills that have to be swallowed on a time-release basis. Whose management and sense of humor have created a buoyant spirit and a true esprit de corps in OED, which is a feat of leadership that itself deserves a medal, and who is conspicuous for his modesty, to end with an oxymoron.

OED enables us to avoid mistakes by drawing the right lessons from the past.

This independence of the personality plus, we must hope, a few of the other qualities is what we must wish for from the next DGO, for the benefit of the Trias. May he, that successor, also have a few qualities of his own like Bob Picciotto, whose love is the toughest and who is unique.

9

Reflections on the 30th Anniversary of OED

Shengman Zhang
World Bank Managing Director

On this anniversary of OED there is a lot to celebrate: 30 years of impressive achievements and innovation, of exploring and pioneering. Today, OED is not just a key player in the Multilateral Development Banks' Evaluation Group, but it is a leader in the whole development evaluation community.

Since OED started its first project performance audit in 1973, it has continuously evolved and moved on to new challenges. Indeed, so much so that I sometimes wonder if there is anything we do in the Bank Group that is not evaluated! But the Bank Group's evaluators have clearly risen to the challenges. They have adapted their approaches and made lasting contributions to our work and the work of the development community.

The history of evaluation in the Bank is continuous rather than discrete; many of the things that we describe, using today's words, in fact had their roots back in the 1990s, 1980s, and indeed the 1970s. But since 1995, OED has been engaging in what it has modestly called a renewal. Some of the results:

- The feedback loop is much shorter. Indeed, I think we all benefit from the *Fast Track Briefs* that Bob Picciotto and OED send us.
- There has been an ever-greater focus on helping to build evaluation capacity, within OED and OEG, among Bank staff (the capacity for self-evaluation), and among our client countries as well.
- The department has been a pioneer in investing in knowledge. Despite the changes and turnovers within the department, staff are well informed about its focus and priorities and about the approaches that they are using. As one of those within management who benefits most from OED results, I really can vouch for that.
- A great effort is being made to build external partnerships, not only with other MDBs, but indeed in the entire evaluation community, including bilateral agencies, and in both developed and developing countries. An international association for development evaluation has been created, and I think that's just fantastic.

- In rising to new challenges, the department is continuing to search for ways to better manage evaluation results.

As an independent unit reporting to the Board, OED has played a unique and really important role. From our point of view in management, that role has not always been convenient. But OED was and is always a trusted advisor to the board, and I would also say, on behalf of management, that it is a respected counterpart of management as well. Indeed, the formal governance structure aside, we've always been able to work together very well for the common goal that we want to achieve: to learn and to improve our contribution to the difficult job of development. Today, as the world demands even more results from our work, evaluation and the contribution of OED seem even more important.

As we celebrate the achievements on this 30th anniversary, we should also recognize the challenges that face the Bank Group and OED.

Thanks directly to OED's work, over the past six or seven years the institution as a whole has moved from a focus on inputs to a focus on quality, and more recently, a focus on results.

It was OED that first started the notion of results-based management within the Bank. (I used to argue with Bob that though this was a great notion it was two years ahead of its time. I remember pleading, first let's get our input management right, so we know at least what we put in, and then let's get some of the input and process quality indicators right!) OED played a pioneering role in stressing the importance of moving to results-based management. Much is now expected of us and I hope that we can make steady progress in this regard, because it is, at the end of the day, results that matter.

It was OED that first started the notion of results-based management within the Bank.

We face a big challenge, however, on the measurement of results. With Joanne Salop's help and with the input from OED we have made a start, but only a start. We are at least setting up a framework, but to reorient the institution's policies, processes, inputs, approaches, staff mindset, and working relations with clients will be enormously difficult. To help us to learn better and more quickly, again I see an important role for OED.

We also need to see how we can collaborate better with clients so that we can better track progress not just for individual countries but across countries, and not just in one sector but across sectors, particularly as we pursue more of the comprehensive PRSP-type approaches. Development is a multisector, multifaceted, interlinked subject. We can't really achieve major progress in one area if we leave others completely behind. This interconnectedness raises a challenge not only from the point of view of practice, but also of evaluation methodology.

To conclude, it is only appropriate to pay a special tribute to those who have led the department over its three decades of existence. (I am so pleased to see some of the former colleagues and DGOs here with us.) Bob Picciotto, of course, has been the DGO for a third of OED's existence. Like his predecessors, Bob embodies the best of our great institution and its staff, in his unwavering commitment to development and to the highest professional standards. It is qualities like his that continue to make our institution respected, liked, and sought out in so many ways. Personally, over the past 10 years, I have learned a lot from him as a professional, but I have also come to know him very well as a wonderful gentleman. Even when we disagreed, we somehow managed to work together very well. And I think that says a lot about the quality of Bob as a leader as well as about the staff and the managers of OED, and about the greatness of the institution of which we are all a part.

10

Panel Discussion: Perspectives on Evaluation in the Bank Today

Prem Garg
Director, Quality Assurance Group, World Bank

Joanne Salop
Vice President and Senior Advisor, World Bank

Jan Piercy
Former United States Executive Director, World Bank

Prem Garg: Having been part of the evaluation family for the past seven years or so, let me say at the outset that the job of an evaluator is not easy. No matter how well one does one's job, somebody is almost invariably unhappy with what one has to say. Often, too, the message gets equated with the messenger. So the temptation is there to take the easy way out and to tell people what they want to hear, rather than what they need to hear if evaluation is to be the springboard for future improvements. Conversely, our well-meaning efforts to maintain our independence and objectivity as evaluators can sometimes lead us to ignore legitimate questions about the robustness of what we have to tell our clients. And that, of course, can undermine the clients' acceptance of our message, as well as any follow-up that needs to take place. In the worst-case scenario, it can mean that evaluators end up giving messages of questionable validity and prescriptions that may backfire in the future.

So a key challenge for any evaluator is to listen to the client, but to listen without trying to please people for the sake of pleasing them. How one arrives at the right balance is very much a matter of judgment. Much depends on the authorizing environment within which evaluators work. This is a question of organizational culture, and most importantly of the leadership of evaluation institutions.

When I look at OED's track record, and I do interact with our partners in multilateral and bilateral institutions, I believe OED has been uniquely successful in providing thoughtful, relevant lessons for practitioners as well as credible measures for tracing accountabilities. Indeed, OED has been a pacesetter in the evaluation field, and we can all be proud of the major strides made under OED's leadership and most recently under Bob Picciotto's leadership in professionalizing the field of evaluation.

Let me note two other major contributions of OED from my vantage point. First, rather than resting on its laurels, OED has been very proactive in keeping up with the changing times and adapting its evaluations to the evolving paradigm of development. This is most aptly captured, if I can borrow Bob Picciotto's phrase, by moving evaluation to a higher plane—that is, shifting the evaluation agenda to a more strategic level; whereas the unit of account for evaluation used to be the individual investment project, evaluations now also cover whole programs, and the agenda has been broadened to cover new themes such as governance, corruption, and the Bank's global policy programs. Indeed, OED has now become a key part of the Bank's policy-building architecture. It is a constructive voice heard in many Bank discussions around the changing agenda. It is a voice of support for stronger client relationships, a voice for better partnerships, and a voice for a more participatory style in economic development.

OED has now become a key part of the Bank's policy-building architecture.

Second, from the QAG perspective, OED has been an exceptionally valuable repository of empirically vetted knowledge about the process of development. I see OED's imprint all over what we do in the Quality Assurance Group, whether in the framework we use for defining risky projects, the framework we use for assessing operational quality, or any of our newer product lines. Our job would have been immensely more difficult had we not been able to tap into the vast storehouse of wisdom provided by OED evaluations.

Looking ahead, I would like to highlight two challenges, one on the external front and one internally.

On the external front, we as evaluators in the Bank have a major problem in that the way we report evaluation results does not resonate with our shareholders and our stakeholders. Much of our reporting is perceived to be too abstract, too process oriented, and too far removed from the concerns of shareholders and civil society. The recent convergence of the development agenda around the Millennium Development Goals provides a focal point both for the Bank and OED to bridge the linguistic and substantive gaps that affect our relationships with our outside clients. To be fair, OED has already been pushing in that direction for the past several years with its pleas for results-based management. Joanne Salop's work over the past few months on results-based management should provide the momentum we need to move forward in this important area in the coming years.

Internally, despite all the progress the Bank has made so far in building an internal culture of quality, I believe we still face a challenge in persuading our staff and our managers about the value added by our evaluation work. Rightly or wrongly, many staff and managers remain skeptical about whether our evaluations focus on the right things, whether our evaluation instruments are robust, and whether the benefit of our evaluation work outweighs the cost. The challenge we face is to build greater ownership and buy-in by staff, both in the substance of what we evaluate and in how we do our evaluations.

• • •

Joanne Salop: From where I sit, the Bank has improved a lot, operationally, over the past ten years. Certainly I'm a vested interest here. But OED's evaluations, the OED numbers, and the QAG numbers show that the Bank has improved the quality of its operations. I think that there is no debate about this and it is a tremendous achievement. I think it's also quite important for the world, because the Bank plays an important role in helping countries achieve better development outcomes.

Three players—OED, the Board, and management—have all made critical contributions to the Bank's operational quality improvement. Way back when, OED was the voice in the wilderness complaining about the overall performance scores of our products. But by the time of the Wapenhans Task Force report in the early 1990s, we looked very much to the OED results as a critical ingredient. OED really played a very important role and continues to do so in signaling what's right, what's wrong.

Second, the Board, and in particular, the Committee on Development Effectiveness, and its predecessor the Joint Audit Committee, played a critical role. They listened to OED, they bugged management, and continued to follow up.

Third, without the changes made by management, you would have had OED signaling the problems, the Board being worried about the problems, but no action. Here I also need to mention the role of QAG. QAG works for management, so they're part of the management team, but on the other hand, they're independent of line management. QAG has made an enormous contribution by independently pursuing in real time the kind of evaluation issues that OED pioneered. To improve the quality of operations—particularly if you're dealing with a staff that is decentralized, not just outside Washington, but even within headquarters outside the central units in the Regions—you need to have consensus about what makes up quality, and you need to have some numbers in real time for managing across units, and again, that's where QAG's contribution has been very important.

What's next, for the Bank and for OED? Having made major progress on the quality agenda, we now need to make the transition to managing for results. That's what I'm working on. I sincerely believe that now that we have operational quality under control, we have the delivery systems under control, we can really focus on results. In the transition to results-based management, just as in the improvement of operational quality, we are still going to need strong contributions from OED, the Board, and management, including QAG.

In fact, the change in focus from operational quality to results implies something of a change in the balance of contributions from OED and QAG. When we were stepping up the work on quality in the mid-1990s, a big issue was that the OED numbers on performance came too late in the overall cycle to be really helpful for management purposes. (By the time OED issued a report on a project, it was maybe eight years after appraisal, and by that time the responsible staff had moved on.) From the point of view of managing for operational quality, it was very important for us to get evaluative information, even if not a full-blown evaluation, earlier in the game. QAG's work makes it possible to get evaluation assessments in real time while staff are still on the job.

Today, now that we're focusing on results and on country outcomes, OED's longer-term perspective is actually advantageous, because these results/outcomes only come about in the longer term. We shall be looking to OED much more to give us answers on how effective our country assistance strategies, our projects, etcetera are in actually influencing the course of country outcomes.

I also think it would be useful, going forward, for us to be thinking with open minds about the division of labor among QAG, OED, the Internal Audit Department, and the Inspection Panel. Each of these institutions was created to fill a particular role, a particular need, at a particular time. Clearly the functions of all four are essential, and must continue, but it may now be timely to step back, particularly in the context of the overall shift to the results focus, to see whether there is virtue in re-dividing the lines among these groups.

• • •

Jan Piercy: For two reasons, I am delighted to be the third to speak.

The first is that you have already heard from people I admired greatly during my time in the Bank. Most of the Bank's executive directors come and go rather quickly, given their two-year terms of appointment. As you enter as a new director, you are truly dependent on being able to reach for a context, a history, to help you make the right judgments as you consider the loans, policies, and programs that are brought to the Board for approval. It is thanks to the continued improvement over the last eight or nine years that I as a director felt I was much better able to do my job. So I wanted to tell you my appreciation for the work that those up here as well as those in the audience have done.

Second, I am pleased to be the final speaker because the Bank's Board is sometimes seen as a Board that proclaims or dictates, but we also listen. With respect to evaluation, it is a significant part of the Board's role to listen, and then to look forward. I was struck how many times we would go from a meeting of the Committee on Development Effectiveness into a debate the next day of a proposed policy or country assistance strategy or loan, and I would hear colleagues around the table refer back to some insight from the day before.

It is a significant part of the Board's role to listen, and then to look forward.

Those discussions in the Committee on Development Effectiveness were for me some of the greatest learning moments of my time on the Board. They were not about history alone, they were all through the lens of, "What have we learned, what do we take from this, where do we go with this, what do we do now, what do we next, how do we do this better?"

The introduction of QAG, which provides, in a sense, self-evaluation, allowed OED to expand its focus, to become somewhat more timely, and to move to a different scale. At the same time, we evaluated the audit function and thought about how that could work holistically, and how audit in the sense of financial accountability could be integrated with audit in the sense of measuring development results and effectiveness. Today, with that body of experience, we're poised at a juncture where it will be possible to take another look. Some social psychologists and others opine that there is an analogy between the life cycle of an individual and the life cycle of organizations. When I first heard this, I thought it was a little too cute by half, but after observing organizations of different ages, I now believe there is some truth to this. If you apply the idea to OED, 30 is an age when you've got much of your education already invested, you have vigor and energy and perhaps the beginnings of wisdom, and a lot more can be expected from you. I hope that those in this room who are an ongoing part of management, of staff, and of the evaluation function in the Bank will say, now we're 30, we've learned a great deal, we're empowered, we know what is needed, we're working in ever-greater partnership than before, so now let's truly attain the maturation and reach of our function. We have an important base now to build on, let's go forward.

As I think about how to characterize the function of OED, I keep coming back to the notion of a balance beam. I say this because OED indeed makes an effort to do justice to serious work by operational colleagues in the Bank, whether it's looking at a project, or at the entirety of the Bank's engagement with a country or a sector. Now—in an institution that has quite an exceptional level of intellect, dedication, and creativity—imagine what it is like when you must very quickly, with a very small team, and a modest budget, try to make a thorough and fair assessment of the results of years of engagement with a given borrowing country. This is almost an impossible task. So you need to become ever better at defining how you're going to do that evaluation. I think we've achieved good success, but we should acknowledge that it's inherently impossible to perfect this. We always have to keep striving, and I am delighted to see the vice president for Human Resources in the audience today, because I believe that in the course of a Bank career, we should build in a rotation through OED or QAG. Having such experience would better inform the work that each of us does, whether it is looking at results prospectively, in the course of being a manager to achieve outcomes, or, after the fact, evaluating the impact and success of the work.

Very often, institutions in the development arena are criticized without due regard for the dire conditions they address.

I spoke last week with Robert McNamara because it was during his presidency that OED was created, and he said that the creation of the evaluation function was one of the developments in which he took the greatest pride. He believes that this phenomenon of honest self-appraisal and looking hard at what fell short as well as what was achieved is the telling point about the Bank. And as Prem and as Joanne spoke, I wished that some newspaper reporters and congressional staff were in this room, because to understand the genuineness of the Bank's search for improvement is to come away with a different understanding of the Bank. Very often, institutions in the development arena are criticized without due regard for the dire conditions they address. We're criticized because poverty has not been more deeply reduced, because education hasn't reached all, and yet we need to look at the other side of the equation, and that is what our open self-assessment significantly contributes to doing.

I'd like to just give you the Board perspective on OED from my seven years as U.S. executive director. It's very humbling to speak as just one executive director, so imagine, if you will, with me in this conversation Eveline Herfkens, who as dean of the Board, was one of the people who were instrumental in creating the Committee on Development Effectiveness; Ruth Jacoby, who became the first CODE chairman; Surendra Singh, who became the second chairman; and Pieter Stek, who now chairs CODE and came as chairman after I left. For each of us, the job of CODE chair has taken a very significant part of our time. It demands of us that we operate not as representatives of our own particular constituency but on behalf of the institutional whole and—more broadly than that—on behalf of what counts for development. As Bob Picciotto has observed, at its best, the Board transcends the interests of particular member countries. And it is in the context of the evaluation discussions that I think the Board most often achieved that very important perspective.

When the Committee on Development Effectiveness was created in 1995, it operationalized for the first time the relationship between the Board and the independent evaluation function. Up to that point, OED "reported to the Board," but what did that mean in a Board of 24 executive directors, with no executive committee? The creation of CODE gave us a nexus and a point at which we could engage with the evaluation function.

It also increased the Board's own accountability. One of the developments since the Committee was created is the practice of discussing in detail the formation of the OED work plan for the year. In the process the Board has to confront very tough tradeoffs and then stand behind its choices. Again, because most directors are gone by the time of completion of the particular operations they have voted into life, it is important for them to focus on the other side of the ledger, to look at what the outcomes were of earlier operations, and then apply those thoughts to today's decisionmaking. When the Committee on Development Effectiveness began to operate, it really created that engagement and strengthened directors' awareness of their accountability.

The creation of CODE gave us a nexus and a point at which we could engage with the evaluation function.

Some of those meetings were quite lively, and there were real debates among directors and with management about how to proceed, particularly when it came to the point of saying, "If these are the insights, and we agree on these lessons from a particular OED study, what shall we include in the ledger of steps that management will now take to implement these recommendations?" Management would say, "We agree with you, these are all desirable, even necessary, improvements, but we can't deliver on them all." One moment came, in fact, when we had to take an entire ledger, which had gotten quite large, and say, "We want to commit to what can be delivered. So, pragmatically, what can we do?" The vice presidency that Joanne Salop was in during this period was essentially retooled to focus on how the Bank would deliver on such improvements. I think that some of the progress the Bank achieved in operational quality owes a lot to this iterative, if messy, process of deciding, "What are we going to do, how are we going to sequence it, and though tradeoffs are needed, how can we really continue to improve as a result of what we're learning through evaluation?"

Looking at the tremendous number of changes in OED evaluation over the last decade, I would like to focus on just two continuing themes. One is a successful commitment to integrating capacity building, collaboration, and real coordination into evaluation. This is certainly in the spirit of leadership from Jim Wolfensohn that has permeated the entire institution. But I have been very excited to see a transformation in how evaluation is conducted, so that more evaluations are now being done with other institutions. Often these are organizations in client countries, so that the work contributes to the development of indigenous evaluation capacity. Some evaluations are done jointly with other departments in the Bank. These partnerships serve to increase country ownership of evaluation outcomes and therefore to encourage greater follow-through on the results.

Another theme is the increase in participatory studies in which the conduct of an evaluation has itself made a contribution to development. For example, the gender evaluation, in which OED looked at the extent to which the Bank had actually followed its own policies with respect to gender, involved a series of Regional workshops, each with different sets of organizational partners. I had the privilege of attending some of those workshops, and you could tell that those workshops themselves provided a tremendous contribution for convening and substantive discussion for people who didn't often have the opportunity to meet.

I'd like to leave you with the image of evaluation as not simply or even primarily an after-the-fact assessment of what has occurred, but as a dynamic exercise in which we're learning from what *has* happened, at the same time sparking further thinking about what *is* happening, and shaping what *will* happen.

PART III:

Relations with Clients and Partners

Today's approach to development assistance is much more collaborative than when OED opened its doors. Donor agencies acknowledge the need to coordinate their contributions to programs that themselves are designed collaboratively by multiple stakeholders. Information sharing, ownership of programs by their intended beneficiaries, transparency, and accountability for results are seen to be qualities at the forefront of successful development. Not just development assistance agencies, but also those who evaluate their work need to cultivate these qualities. As the papers in this section point out, OED has become increasingly outward-looking and systematic in its relationships with clients and partners, and in its professional exchanges with other parts of the evaluation community.

Strong capacity for monitoring and evaluation is a key element in sound governance. OED's efforts to support the development of evaluation capacity in Bank borrower countries began very early in its history, and the road since has been rocky, but with notable and lasting successes along the way. OED now plays an active catalytic role, across a wide range of borrower countries, in helping build evaluation capacity both through collaborative work with their evaluation staff and through training and professional support, and in supporting Bank programs for better monitoring and evaluation systems.

Partners provide vital intellectual as well as financial support for OED, and today working through partnerships is a key part of the department's strategy. The origins of some of OED's current partnerships go back 20 years or more, while others are quite recent. OED connects to the evaluation organizations of other development assistance agencies through formal partnership agreements, through taking part in joint evaluations, and through proactive participation in international working groups on aid evaluation.

11

Evaluation Capacity Development

Keith Mackay
Senior Evaluation Officer, OEDPK, and
Coordinator for Evaluation Capacity Development

OED has been a persistent and often influential proponent of evaluation capacity development both in the Bank and in borrower countries. From its earliest days, it supported the strengthening or establishment of borrower governments' evaluation systems. Reporting to the Bank's Board in 1976, OED noted that:

> The more active involvement of borrowers in evaluating the Bank's contribution to their development effort will improve the quality of evaluation findings, and may also help borrower governments establish more effective evaluation systems of their own.

By 1977, the signs were auspicious. Borrowers had all responded positively to invitations to become more actively involved in Bank evaluations, and their involvement was expected to benefit both the management of ongoing projects and the design of future projects. By 1978, governments of some 25 to 30 Bank member countries had expressed interest in establishing some form of domestic evaluation system, and that year, the Bank began formally requiring borrowers to prepare completion reports on Bank operations.[1] Regional meetings[2] were held with borrower countries to discuss the desirability of establishing and strengthening evaluation systems for all projects, and in 1979 OED noted that:

> There would appear to be ample basis for concluding that the Bank's effort to enlarge the borrowers' role in the Bank's evaluation activities should be sustained. The executive directors and the president have consistently supported and encouraged such an effort.

The DGO and the director of OED continued to advocate this type of capacity building in a number of forums. Staff on secondment from member governments were welcomed, as a means of disseminating OED evaluation systems and methodologies. And the Economic Development Institute (precursor to the World Bank Institute) was offering monitoring and evaluation (M&E) as an integral part of most of its training courses on project management.

By 1980, however, OED was sounding a cautionary note. The expressions of interest by member government officials had not been matched by action. To the Board, OED revealed that:

> Thought had once been given to reviewing briefly in this and succeeding annual reports where the Bank's major borrowers stand in institutionalizing some kind of national project evaluation function. This idea has been dropped . . . principally because of the substantial diversity of views about what are reasonable country goals in this regard. There are senior borrower government and Bank staff who believe that solving pressing implementation problems must take priority and that, until this is done, even limited attention to evaluation cannot be justified in administrations chronically short of qualified staff. The fact is that the subject is not yet a matter of systematic Bank operational dialogue with member governments, not even with those which are now or will soon be engaged with Bank staff in preparing numbers of project completion reports.

Subsequent reports to the Board echoed the same views; most governments lacked the political will to have a fully functioning and independent central evaluation function. Even so, OED continued to describe the Bank's efforts to promote evaluation in borrowing countries, and to list countries that had assumed responsibility for preparation of completion reports on Bank projects.

OED Strengthens Its Focus on ECD

The appointment of Yves Rovani as DGO in March 1986 strengthened OED's focus on helping countries build M&E systems. Mr. Rovani believed that:

> Helping countries to develop their own capacity for evaluation is just as important a task for the DGO as promoting self-evaluation in the Bank. Both tasks are key to the effective process of learning and enhancing operational quality.

From 1986 onwards, the formal list of OED's five principal functions included the following: "help the Bank encourage and assist member countries to develop their own operations evaluation systems," while the DGO's own formal mandate included "encouraging and assisting developing member countries to build effective monitoring and evaluation associations, capacities, and systems."

Consistent with this, in 1987 Rovani instituted the Evaluation Capability Development Program (ECDP) within OED,[3] to promote better M&E by countries of ongoing projects, and better completion reporting on projects of the Bank and other assistance agencies. The new program was tasked with strengthening networks of evaluators in developing countries; assessing countries' interest in evaluation and their evaluation capacity–building requirements; and organizing evaluation seminars and workshops to build awareness and, in association with EDI, provide evaluation training; providing short-term training within OED for borrower staff; and organizing pilot programs in a small number of countries. To manage the program, the DGO nominated an OED staff member as its (part-time) coordinator.

In 1991, the Board accepted the DGO's recommendation that the Bank establish a taskforce to develop a support strategy for ECD.

Partnerships with multilateral and bilateral donors, such as UNDP, helped provide funding and other support. The OECD Development Assistance Committee's Expert Group on Aid Evaluation, one of the most vocal supporters of the ECDP's mission, sponsored various regional seminars for this purpose and the regional banks also provided valuable support, as did some bilateral donors, including DANIDA and USAID.

OED's own resources for ECD were limited, so assistance was targeted to countries that met the following selection criteria: willingness to appoint an experienced task manager for ECD, allocating adequate staff resources for the work, and having a steady prospective flow of evaluation work that would benefit from regular OED support. Countries that received assistance in this period included Argentina, Brazil, China, Colombia, Indonesia, Morocco, the Philippines, Sri Lanka, Thailand, Tunisia, and Zimbabwe (box 11.1).

Bank Adopts a Support Strategy for ECD

Meanwhile, OED continued to urge Bank management to promote the development of medium- to long-term technical assistance programs to strengthen borrowers' evaluation capabilities. In 1991, the Board accepted the DGO's recommendation that the Bank establish a task force to develop a support strategy for ECD.

The Task Force for ECD reported in 1994 and its recommendations became the Bank's ECD strategy. It was recognized that feedback from M&E played an important role in improving government performance and public sector management, so assistance to ECD was to be undertaken as an integral part of the Bank's country portfolio management work and support for public sector management reform. The long-term

Box 11.1: OED's Catalytic Role in ECD—Colombia

After an evaluation by OED of the Bank's experience in Colombia's power sector, in 1990 the Minister for Finance asked the DGO for assistance in developing an evaluation system for the Colombian public sector. At the DGO's suggestion, the minister and the head of the National Planning Department (NPD) designated a senior Colombian adviser, Eduardo Wiesner, to design an approach for establishing such a capacity. Wiesner's study, which was assisted by information and advice from OED (see chapter by Yves Rovani in this volume), led the Colombian authorities to conclude that to create a public sector evaluation function would require a change in the law. In 1991 the Colombian Constitution was being rewritten, and, seizing this opportunity, an article was drafted with the support of members of the Constitutional Assembly, making public sector evaluation mandatory and conferring on NPD the responsibility for establishing a national evaluation system.

In late 1991, the NPD consulted with present and past public sector officials and private sector representatives as to alternative organizational approaches; evaluation experience arising from large controversial public sector projects; and appropriate regulations for the evaluation function. Participants included members of the Constitutional Assembly, the Controller General of the Republic, the Minister for Finance, the head of NPD, the Attorney General, and other sector ministers. Also participating were the Deputy Controller General of the U.S., the DGO, and other international officials.

NPD then established an evaluation function. In September 1993, the World Bank approved a Public Financial Management Project, which incorporated a *public policy evaluation* component. This had the goal of verifying that public resources had been well programmed and honestly and efficiently spent, and that public policies generating resource mobilization and use were sound and represented the best possible use of public resources for development, including special concerns such as equity and environmental protection.

In 1997, Colombia formally established a *national system for evaluation of results of public sector performance* (SINERGIA). This system has been implemented in all 16 sector ministries and in 170 public entities. It has focused on the setting of strategic objectives, performance indicators, and performance targets. Some ex post evaluations have also been commissioned. By 2001, and based on SINERGIA, the planning and finance ministries had started to enter into *efficiency agreements* with specific ministries and public entities, to help guide budget funding commitments.

In February 2001, the Bank approved a second Public Financial Management Project, which contained a component designed to strengthen SINERGIA in a number of ways, such as implementing an incentive model for sector ministries, linked to public policy outcomes, and increasing the priority placed by sector ministries on conducting evaluations of their performance and that of territorial and decentralized entities.

goals of the strategy were to mainstream ECD within countries and to strengthen M&E in the Bank. The Joint Audit Committee, endorsing the strategy, requested a progress report from management.

In the ensuing years, OED repeatedly noted to the Board that the Bank was making too little progress in implementing the strategy. In 1996, for example, it reported that:

> Though some progress has been made in alleviating key capacity constraints, most public sector studies do little to nurture demand or seek to persuade borrowers about the practical utility of incorporating evaluation practices into the design and management of public expenditure programs.

And in 1997 it noted that, "Unfortunately, there has been no take-off of operational activities and no progress report from management about its intentions."

OED's Continuing Role in ECD

Given the continuing lack of attention from the Bank, OED intensified its support. OED's renewal program, introduced in 1997 under DGO Robert Picciotto, confirmed the department's catalytic role in ECD. In 1999 the department recruited a full-time staff member in this area. It devoted more time and effort to persuading key senior Bank staff about the merits of ECD, and persuading the evaluation offices of other multilateral and bilateral agencies—such as the African Development Bank, the Swiss Agency for Development and Cooperation, the Norwegian Foreign Affairs Ministry, the Development Bank of Southern Africa, and the UNDP—to actively support ECD efforts.[4] These partners jointly sponsored international seminars[5] for senior government officials and donors to raise awareness of the potential role of M&E in sound governance, strengthen interest in ECD, and help initiate ECD efforts in countries such as Ghana and Uganda.

OED also played a leading role in highlighting ECD at international evaluation conferences, such as the annual meetings of the American and European evaluation associations, and at the regular meetings of the evaluation offices of multilateral donors (the Evaluation Cooperation Group) and bilateral donors (the OECD Development Assistance Committee's Working Party on Aid Evaluation).

A related strand of the ECD work was educational in nature—the preparation and dissemination of ECD resource material such as seminar proceedings, diagnostic guides, country case studies, and creation of an ECD Web site.[6]

Bank M&E Improvement Program

In 1999, management responded to OED's continuing calls for an ECD progress report as well as a persistent concern about the quality of M&E conducted by the Bank.[7] It set up a working group on Bank M&E, with broad-based representation from the Regions, networks, and key central units including OED. The Working Group highlighted a set of interrelated weaknesses in M&E common to the Bank and borrower countries:

- Poor incentives for good M&E
- Diffuse accountability, because of unclear roles and responsibilities both within the Bank, and between the Bank and borrowers
- Weak capacity for M&E.

The Working Group highlighted a close relationship between the quality of the Bank's project, sector, and country-focused M&E, and borrowers' ability to conduct and utilize M&E.[8] It recommended a Bank M&E Improvement Program that would include ECD pilots in eight countries and two sectors, with a view to mainstreaming ECD in Bank-supported operations starting in FY04. This recommendation can be interpreted as a shift from a focus on project-level M&E toward a concern with M&E systems at the country and sector levels.

The Working Group highlighted a set of interrelated weaknesses in M&E common to the Bank and borrower countries.

In September 2000, CODE agreed with the recommendations put forward by management, and the program was then launched. Like OED, however, CODE expressed concerns that a pilot approach would not be sufficient. The Committee noted that "the cost of not having good M&E systems could be very significant for the borrower, resulting in expensive restructuring of projects or even cancellation of projects."

New Pressure to Show Development Results

Several recent developments have heightened the pressure on countries, the Bank, and donor agencies to demonstrate the results of development spending. The Comprehensive Development Framework (CDF) initiative, followed by the strong results-orientation of Poverty Reduction Strategy Papers (PRSPs) and by the linking of PRSPs to the Heavily Indebted Poor Countries (HIPC) debt relief initiative—which will cost donors some $29 billion in present value terms—have emphasized a country-focused, results-oriented approach to development. About 60 countries are preparing PRSPs. PRSPs in turn are linked to the Millennium Development Goals (MDGs), which emphasize measurable indicators and results. With both the CDF and the PRSPs, the strong emphasis on results provides an environment where M&E has a potentially central role to play.

The need for solutions was underlined by a 2001 OED finding that weak M&E contributes to poor project performance in 38 percent of Bank projects rated unsatisfactory or highly unsatisfactory.[9] At the March 2002 Monterrey Conference on Financing for Development, the heads of the five multilateral development banks issued a joint statement stressing the need for better M&E and for related capacity building:[10]

> The quality of country systems for measuring and monitoring results is important for the challenge before us. This puts a premium on our capacity building support for public sector management, statistical development, and monitoring and evaluation systems, which are important in their own right for underpinning countries' accountabilities to their people and results agreements and compacts with donors. Our agencies have programs of support for evaluation and statistical capacity building, aiming to help countries improve their measurement, monitoring and management for results—whether at the project level or at the agency or ministry level.

Following the Monterrey Conference, the Bank has recommitted itself strongly to providing clear evidence of the results of its work, partly through more focused attention to M&E within the Bank and via support for ECD in client countries. And in early 2002 it created another task force. Headed by a Bank vice president, the task force is developing an action plan for better measuring, monitoring, and managing development results, with a strong focus on progress toward the Millennium Development Goals. The recommendations are likely to focus again on M&E systems and capacities of both the Bank and borrowers.[11]

OED's Role in ECD Today

During all these developments, OED has maintained its support and advocacy for ECD. Its own support is based on the subsidiarity principle—leaving Bank country teams to take the primary responsibility—and is designed to be catalytic, by prompting and supporting the involvement of the Bank's Regions, supported by central units such as Operations Policy & Country Services (OPCS), the World Bank Institute (WBI), and the Development Economics Vice Presidency (DEC):

- OPCS is responsible for implementing the M&E Improvement Program, including providing support for pilot countries (with Bank country teams), as well as for a growing number of other countries involved in ECD.

- WBI provides M&E training programs for Bank staff and borrowers.
- DEC supports research and statistical capacity building in borrower countries.

OED itself tries to ensure that the Bank's ECD work is of high quality—relevant, effective, efficient, sustainable—and contributes to borrowers' institutional development.

OED's own resources for ECD are limited (currently to two full-time staff members plus consultants), but it receives valuable support from partners. OED has established formal ECD partnership arrangements with, for example, the evaluation offices of UNDP, the Swiss Agency for Development and Cooperation, the ministries of foreign affairs of the Netherlands and Norwegian governments, and the U.K. Department for International Development. OED receives funds from its bilateral partners to support its ECD work.

For ECD, OED provides:

- M&E training such as the annual International Program for Development Evaluation Training (IPDET) and the Introduction to Program Evaluation (IPE). IPDET is a 2–4 week residential course on development evaluation that attracts more than 100 participants a year. It was started in 2001 in partnership with Carleton University, Ottawa. OED's ECD partners fund a large number of scholarships for officials from borrower governments and leaders of national evaluation associations to participate in IPDET. IPE provides an introduction to basic M&E concepts and skills; it is delivered by video conference to sites around the world. More than 1,000 persons have been trained in 34 countries since 1998—Bank staff are included.
- Support for the creation in September 2002 of the International Development Evaluation Association (IDEAS), in collaboration with the evaluation office of the UNDP, and again with the support of OED's other ECD partners. IDEAS focuses on the evaluation needs of developing countries and transition economies. It seeks to strengthen the development dimension of national evaluation associations; promote ECD; inspire professional excellence in development evaluation; and provide global knowledge, support, and connectivity to development evaluation professionals.

In playing its catalytic role, OED takes every opportunity to highlight the importance of ECD. For example, its *2002 Annual Report on Evaluation Capacity Development* took stock of ECD activities underway across the Bank, and found that by March 2002, the Bank was supporting ECD in at least 21 countries, 12 of which are PRSP countries. Stressing the priority for mainstreaming ECD, the report raised issues for Bank management:

- The need to clarify within the Bank what mainstreaming would entail, including when ECD should be mainstreamed
- Confusion evident among Bank staff and borrowers about what M&E comprises, and about how to institutionalize M&E in a sustainable manner
- Inconsistent messages in Bank policies and guidelines on M&E
- ECD coordination and knowledge sharing within the Bank and externally
- The resources needed to support M&E and ECD.

In considering OED's report, the Board's Informal Subcommittee on Development Effectiveness remarked on the strong priority for ECD and reiterated its full support. It expressed concern over the limited progress since the introduction of the strategy in 1994, and urged management to determine the key constraints on ECD.

Conclusions

As OED has always appreciated, stronger M&E capacity in borrower countries can facilitate Bank M&E, but for countries themselves it is also a key element in sound governance. It provides access to reliable country data systems and to the work of the country's own evaluators. Strong M&E capacity, developed over the medium to long run, will make it possible for donors to rely on recipients to self-evaluate aid projects, subject to appropriate quality assurance, and it can also increase country ownership of evaluation results, increasing the probability that these results will be used.

Being the independent evaluation arm of the World Bank provides OED with a particular comparative advantage in the area of ECD.

OED has built a catalytic role for itself by promoting ECD, by supporting Bank country programs for improved M&E systems, and by prodding Bank management to mainstream ECD. Being the independent evaluation arm of the World Bank provides OED with a particular comparative advantage in the area of ECD, because of its evaluation expertise and its links with the evaluation offices of the multilateral development banks and bilateral donors.

The current emphasis on results and country ownership provides the best opportunity for ECD so far. Whether that opportunity will at long last translate into the mainstreaming of ECD remains to be seen. The growing number of countries with which the Bank is working to strengthen or build M&E systems provides some indication that the mainstreaming of ECD might actually become a reality.

Notes

1. One reason for this (the other was budgetary) was to encourage borrowers to strengthen their M&E systems to better support efficient development and to help the Bank be of better service to them. OED also hoped that borrower involvement in evaluation would help support OED's independent evaluation of the Bank's work. Due to continuing problems with the quality of completion reports prepared by borrowers, this requirement was abolished in July 1989. Bank project staff resumed responsibility for preparing completion reports; some input from borrowers was also expected.

2. Abidjan in March 1978, Nairobi in February 1979, and Kuala Lumpur in May 1980.

3. Renamed the Evaluation Capacity Development Program in 1993.

4. R. Picciotto, "The New Challenge of Evaluation Capacity Development," *Evaluation Cooperation Group Newsletter,* 1999, 1(1), 1–2.

5. Washington D.C. in 1998, Abidjan in 1998, Johannesburg in 2000, and The Hague in 2001.

6. http://www.worldbank.org/evaluation/ecd/

7. In 1999, for example, the Bank's Quality Assurance Group found that less than half of Bank projects had an appropriate M&E system in place.

8. OED has rated borrowers' provision of M&E data as substantial or high in only 36 percent of operations exiting in FY01.

9. This estimate is based on a sample of 45 Bank projects with unsatisfactory or highly unsatisfactory outcomes. (*2001 Annual Review of Development Effectiveness,* 20 December 2001, World Bank Operations Evaluation Department, para 3.28).

10. Statement issued in Monterrey by the heads of the five multilateral development banks, *On Measuring, Monitoring and Managing for Development Results,* 20 March 2002.

11. An emphasis on results has been characterized as another, third stage in the evolution in the Bank's priorities: the first stage constituted the particular emphasis on loan volumes when the Bank was first created; the Wapenhans task force recommendations in 1992 were a response to a growing concern over the worsening performance of Bank projects, and this led to the second stage—an increased focus on the quality of project design and implementation. Another recommendation was that OED continue to assist borrowers to strengthen their evaluation capacities, in the context of public sector management reform programs.

12

Evolution of OED's Evaluation Partnerships

Osvaldo Feinstein
Manager, Partnerships and Knowledge Management, OEDPK

Rema N. Balasundaram
Evaluation Officer, OEDPK

Partnerships involve collaborative relationships under mutually agreed objectives with evaluation units of other development agencies. Evaluation partnerships provide OED with:

- Additional technical expertise and in-country knowledge
- Increased involvement of key stakeholders in the evaluation process
- Increased dissemination of evaluation findings and recommendations
- Ways to improve donor coordination
- Increased access to financial resources.

Informal Partnerships and Experience-Sharing

During its first two decades, OED began developing informal partnerships with the evaluation departments of other international financial institutions and bilateral development agencies, and with agencies in Bank client countries. By the early 1980s, OED was undertaking joint evaluations with cofinanciers of projects, such as France's Caisse Centrale (for three oil palm projects in the Ivory Coast) and Germany's Federal Ministry of Economic Cooperation (for a phosphate fertilizer project in Morocco).[1]

Staff from client country governments, interested in learning about OED's structure and ways of working and in exchanging views on evaluation, began visiting OED in 1979, and over the years since then, staff from bilateral aid organizations, the regional banks, the Economic Commission for Africa, the European Union, the OPEC and Kuwait Funds, the Commonwealth Secretariat, and UNESCO, among other organizations, have exchanged visits with OED to share information on the work of their agencies and on the Bank's operations evaluation system. OED staff and managers have also exchanged views and experiences at international conferences and workshops on evaluation.

Participation in International Evaluation Groups

OED has long played a pivotal role in international evaluation working groups. It began participating in the meetings of the OECD Development Assistance Committee in 1979, in order to share evaluation experience with DAC members and multilateral development banks. The DAC Working Party on Aid Evaluation (WP-EV) was created in 1982, and now consists of 29 representatives from OECD member countries and multilateral development agencies.[2] OED became an observer member of the Working Party in 1982. It has participated in various steering committee groups and has played a key role in the evaluation of programs promoting participatory development and good governance; evaluation rating systems; country program evaluation; the development of a glossary of key terms in evaluation and results-based management, and evaluation capacity building. OED continues to play an important role in the WP-EV.

OED has long played a pivotal role in international evaluation working groups.

In response to the Development Committee's 1995 Report on Multilateral Development Banks, OED has worked jointly with evaluation units in other multilateral development banks to harmonize MDB evaluation criteria and practices. In October 1995, OED joined other MDB evaluation heads to form the Evaluation Cooperation Group. The chairmanship of the Group rotates annually, and is currently with the Bank's director-general, Operations Evaluation. The Group works to strengthen cooperation among evaluators and harmonize evaluation methodologies, and it has facilitated the dissemination of evaluations through its electronic newsletter, "Evaluation Insight."

OED has been a member of the United Nations Inter-agency Working Group (IAWG) on Evaluation since its inception in 1984. The IAWG provides a forum for the discussion of evaluation issues within the UN system.

OED Conferences and Seminars on Evaluation

Organization of conferences and seminars on evaluation has been one important way through which OED has promoted evaluation partnerships and networking among development evaluators. OED organized a first seminar on evaluation experiences in 1991. That event brought together evaluation staff of most of the bilateral aid agencies and regional development banks with representatives of developing countries active in evaluation or interested in developing evaluation capacity: Brazil, China, Colombia, Mexico, Morocco, and Thailand.

Three years later, OED started a series of Evaluation and Development Conferences. Evaluation practitioners, academics, and evaluation managers and officials from borrowing countries and other development agencies participated, along with Bank managers and staff, to explore evaluation methods, instruments, and processes in relation to the demands of the Bank's development agenda. Subsequent conferences have concentrated on the Institutional Dimension (1997),[3] Evaluation and Poverty Reduction (1999),[4] and Partnerships and Evaluation (2001). The 2001 Conference on Partnerships and Evaluation brought together academics, bilateral organizations, multilateral organizations, Bank staff, nongovernmental organizations, and the development evaluation community to discuss the foundations of partnerships, their workings in practice, approaches to their evaluation, and lessons of experience for carrying out successful partnerships. The proceedings of this conference are being published[5] and may serve as a basis for improving the design, implementation, and evaluation of partnerships in evaluation and in other development areas.

Enhancement of OED's Evaluation Partnerships

Development assistance since the late 1990s has emphasized sectorwide approaches and multi-donor assistance efforts with common goals. In this context, effective evaluations require systematic collaboration among donors and increased borrower country participation in evaluation processes—promoting shared learning to improve development effectiveness.

OED's 1997 renewal strategy, described earlier in this volume, paved the way for strengthening and enhancing the department's evaluation partnerships and for institutionalizing the partnership dimension in its work. Investment in knowledge and partnerships became one of OED's five strategic objectives and, in order to support this new emphasis, a new unit was created in OED to deal with partnerships, knowledge management, and evaluation capacity development. The unit became fully operational in March 1998.

Since then the number of formal evaluation partnerships has grown significantly (table 12.1).

Most of OED's partnerships with external agencies involve funds transfers to OED (of those listed in table 12.1, the only exception is the partnership with UNDP's Evaluation Office). The purposes of the partnerships endowed with funds are indicated in table 12.2.

Table 12.1. Formal OED Partnerships Since 1997

Year	OED Evaluation Partnership with
1997	Swiss Development Cooperation
1998	Swiss Development Cooperation and UNDP's Evaluation Office
1999	Swiss Development Cooperation, UNDP's Evaluation Office, Norway
2000	Swiss Development Cooperation (2nd. Phase), UNDP, Norway, Netherlands
2002	Swiss Development Cooperation, UNDP, Norway, Netherlands, UK Department for International Development

In addition, a $2 million dollar Multi-donor Trust Fund for evaluation of the Comprehensive Development Framework was set up in 2001 by OED and the Bank's Development Economics Vice Presidency, with 10 donor organizations. OED has set a ceiling of 15 percent in the volume of trust funds that it will allow in relation to total Bank budget funds used for evaluation activities. This ceiling allows OED to properly execute its fiduciary responsibility over these funds and to avoid dependence on external funding.

A precursor to these arrangements was the Technical Cooperation Program Agreement signed in 1988 with the Canadian International Development Agency (CIDA), for one million Canadian dollars for three years. That program provided OED with additional resources to increase the analytical depth of studies in its work program, focus on the human dimension of development projects and their impact on the environment, and strengthen dissemination and feedback.

Partnerships without funds include relationships with the German Technical Cooperation Agency (GTZ); the Evaluation Office of the UN Development Program; the Evaluation Cooperation Group, comprised of the evaluation offices of multilateral development banks; the DAC Working Party on Aid Evaluation, comprised of the evaluation offices of bilateral cooperation agencies; and the UN Inter-agency Working Group on Evaluation.

OED works collaboratively with the Operations Evaluation Group of the International Finance Corporation and the recently established independent evaluation unit of MIGA (all three report to the DGO). It also works closely with the IMF's new independent evaluation office to evaluate initiatives in which both the Bank and the Fund are involved (such as the poverty reduction strategy papers).

Table 12.2. Purposes of OED's Evaluation Partnerships

	Partnership			
	SDC–OED	**MFA/ Norway–OED**	**MFA/ Netherlands–OED**	**DFID–OED**
Year of agreement	1997 1st Phase 2000 2nd Phase 2002 Amendment	1998	2000	2002
Purpose	• Promote learning from evaluations • Provide additionality to the evaluation system through deeper poverty focus and increased utilization of evaluation findings • Support evaluation activities in the areas of health, education, gender, rural poverty, environment, PSD, water, energy, small and medium-size enterprises, social development, urban	• Support evaluations in the following areas: social development, economic development, post conflict, governance, knowledge management, environment, NRM, and gender • Support evaluation capacity development, including training	• Strengthen monitoring and evaluation capacity, particularly in PRSP countries • Support International Training in Evaluation • Support country-led evaluations • Build Regional Centers of Excellence in Evaluation	• Contribute to impact evaluation activities • Support development evaluation training

Joint Evaluations[6]

OED is also involved in ad hoc partnerships to carry out joint evaluations with bilateral and multilateral organizations. OED's partners in joint evaluations are most often bilateral and multilateral agencies, but occasionally they include recipient countries, NGOs, or other institutions. Some examples follow.

Project evaluation. In Jordan, the Bank-supported Education VII project included grants from the European Union, DFID, Japan, USAID, and UNDP.[7] Financing for school construction and hardware was provided by the World Bank and by Japan's Overseas Economic Cooperation Fund, which is now part of Japan's Bank for International Cooperation (JBIC). JBIC conducted a performance assessment jointly with OED. Each organization focused its evaluation on the parts of the project most closely linked to its own funding.

Country assistance evaluation. OED and the African Development Bank recently completed a joint country assistance evaluation that covered Lesotho's economic and social development in the 1990s.[8] The partnership augmented the country assistance evaluation capacity of the African Development Bank and led to further collaboration with OED. The African Development Bank is currently leading a joint evaluation with OED of the Rwanda country program. OED is carrying out a country assistance evaluation for Jordan jointly with another multilateral partner, the Islamic Development Bank (IsDB). The evaluation builds on a joint mission and joint analyses. Because this is the first country-level evaluation of the IsDB, the OED team leader has provided advice to the IsDB and helped the institution to develop its evaluation.

Sector evaluation. OED participated in the multi-donor review of Ghana's road/transport subsector strategy by the government of Ghana, under the leadership of Danida and in partnership with DFID, the EU, the African Development Bank, Japan, the Netherlands, and Germany. All the partners shared the cost of the evaluation, and most (including OED) participated in an oversight steering committee chaired by the government. Completed in 2000, the review was carried out by consultants managed by Danida.

Process evaluation. The joint process evaluation of the Comprehensive Development Framework, referred to above, is being conducted according to CDF principles and will be completed in the spring of 2003. Its governance structure is complex because it includes a wide range of participants. Decisions are made by a Management Group including two representatives of CDF countries, one representative from a bilateral evaluation unit, and two representatives from the World Bank. The Management Group is advised by a Steering Committee with 25 members representing bilateral donors, international financial institutions, NGOs, and research institutes. The evaluation builds on six country case studies and four thematic reviews.

Are Evaluation Partnerships Working?

During 1999, OED and the Swiss Agency for Development and Cooperation carried out an "appreciative enquiry" to assess the first phase of their evaluation partnership and explore how to increase its efficiency and effectiveness. The partnership was found to be meeting its goals, although better exchange of information and reporting between the partners was needed. As a result of the evaluation, the parties now use specific working agreements that define the expectations and inputs of all stakeholders, the timing of work, and the specific objectives of each collaborative project.

OED staff who have participated in evaluation partnerships generally find them to be relevant to their work, especially because partnerships provide means and incentives to achieve broader evaluation objectives, such as borrower and beneficiary participation in evaluation and improved methodologies. OED staff see evaluation partnerships as helping to make their work more effective by providing finance, technical inputs, country knowledge and experience, local contacts, and recommendations for consultants. Several also find that partnerships enable them to create more useful evaluation products and processes.

OED management has increased the information on partnerships, using different channels to disseminate it.

However, OED staff have expressed concern about inefficiencies associated with partnerships. As barriers to good partnerships, some OED staff pointed out the transaction costs of partnering activities, communication breakdowns, lack of sufficient information about particular partnerships, and lack of a full understanding of the partner's perception of the partnership. OED management has increased the information on partnerships, using different channels to disseminate it (including a Web sub-site on OED partnerships, which has been considered "best practice" in a recent independent review of Bank's Web sites). A key challenge for OED is to design and implement an effective and efficient system for monitoring and evaluating its evaluation partnerships.

Notes

1. OED, *Annual Review of Operations Evaluation* (Washington, DC, 1982).

2. Australia, Austria, Belgium, Canada, Denmark, European Commission, Finland, France, Germany, Ireland, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States; World Bank, Asian Development Bank, African Development Bank, Inter-American Development Bank, European Bank for Reconstruction and Development, UN Development Program, and International Monetary Fund.

3. Picciotto, Robert, and Eduardo Wiesner, eds., *Evaluation and Development The Institutional Dimension* (New Brunswick, NJ: Transaction Publishers for the World Bank, 1998).

4. Feinstein, Osvaldo N., and Robert Picciotto, eds., *Evaluation and Poverty Reduction,* World Bank Series on Evaluation and Development, Volume 3 (New Brunswick, NJ: Transactions Publishers for the World Bank, 2001).

5. Liebenthal, Andres, Osvaldo Feinstein, and Gregory K. Ingram, eds., *Evaluation & Development: The Partnership Dimension* (New Brunswick, NJ: Transaction Publishers for the World Bank, forthcoming).

6. Lessons learned from the experience of partnerships for joint evaluation are summarized in Osvaldo Feinstein and Gregory K. Ingram, "Lessons Learned from World Bank

Experiences in Joint Evaluations," paper presented to the Development Assistance Committee Working Party on Aid Evaluation Workshop, Paris, March 25–26, 2003.

7. See Operations Evaluation Department (2000), *Partnership for Education in Jordan,* Précis No. 193.

8. Operations Evaluation Department, *Lesotho: Development in a Challenging Environment* (Abidjan and Washington DC: World Bank, 2002).

PART IV:

Challenges of Development Evaluation

OED has played a major role in the evolution of development evaluation as a discipline. The contributions in this section argue that, going forward, development evaluation will need to keep pace with rapidly evolving ideas in the development business. In particular, it will need to confront the real incentives governing the behavior of aid donors and recipients (many of which may work against the efficient and effective use of aid), and the fact that aid flows alone are not large enough to make the needed impact on poverty. Recognizing that developed countries' policies on trade, migration, financial flows, and the environment have effects on poor countries and poor people that are many times as great as those of aid, development evaluation will need new units of account, and new methods, structures, and processes. As a meta-discipline, it will need to combine the special strengths of each of the knowledge disciplines working in development: accounting and auditing, management consulting, policy research, and advocacy by voluntary agencies. And it will need to be managed through partnerships and networks including governments, private organizations, and civil society. The International Development Evaluation Association, which OED has helped to bring to life, will need to play a key role in supporting the new discipline and in mainstreaming the new evaluation approaches that are needed.

13

The Global Challenge of Development Evaluation

Robert Picciotto
Director-General, Operations Evaluation, 1992–2002

I am happy to be back with friends and colleagues to explore what lies in store for development evaluation. In September, I talked about the renewal of OED over the past decade. But the past is not always prologue and it is time for a new start under new leadership.

The Development Environment

Development evaluation is rambunctious, energetic, and filled with hope. It is the youngest discipline of a young profession—the new kid on the block. It has established its identity as a discipline and acquired a reputation for rigor and openness.

But it lives in a tough neighborhood—at the intersection of two highly contested fields. Its early successes do not guarantee its future prospects. Indeed, its place in the sun will not be secure unless it can sustain its relevance and demonstrate its value within the tough, competitive professional arena of public accountability and organizational learning.

Thus, any speculation about the future of development evaluation must start with an examination of changes already under way in the authorizing environment. What are the prospects of the development enterprise? What force field will influence development evaluation activities?

The main strategic reality to be acknowledged is the commonality of challenges facing developed and developing countries.

The main strategic reality to be acknowledged is the commonality of challenges facing developed and developing countries. Their fates are inextricably interlinked. Development can no longer be conceived as something that happens "out there."

The problems of the global poor—lack of security, opportunity, and representation; a fractured and inequitable social order; and a stressed physical environment—have become everyone's problems. In the words of Kofi Annan:

> We are all influenced by the same tides of political, social, and technological change. Pollution, organized crime, and the proliferation of deadly weapons likewise show little regard for the niceties of borders; they are problems without passports.

Development evaluation is building on sound foundations. Its track record at the project and country levels has been rewarding. The new discipline has had a notable impact on development practice. Development effectiveness is now securely locked into the agenda of development agencies; Millennium Development Goals have been endorsed by the entire development community, and the themes of coherence, partnership, ownership, and results orientation are on everyone's lips.

Unfortunately, actual practice has lagged behind the rhetoric of consensus.

Unfortunately, actual practice has lagged behind the rhetoric of consensus. First, lessons drawn are not necessarily lessons learned. Second, basic changes in aid policies and practices have yet to be made across the system. Third, developing countries have not set up fiduciary processes that give adequate confidence to taxpayers in developed countries that public expenditures are properly managed. Fourth, the development paradigm has not kept up with the challenge of globalization.

Aid Is Not Enough

A priority task for development evaluation will be to keep decisionmakers' feet to the fire (in developed and developing countries alike) and highlight the gap between promises and actions, means and ends, clarity and obfuscation. Conversely, to achieve sustainable progress, chronic optimism about the capacity of aid to deal with all the ills that affect the developing world will have to be relinquished in favor of a more comprehensive (and realistic) evaluation framework.

Such a framework should encompass not only governance, policies, and institutions in developing countries, but also the major transmission belts of globalization anchored in the policies of rich countries. Beyond aid, these include trade, migration, intellectual property, private finance, and the environment. Policy and institutions matter at the global level just as they do at the country and local levels. This has fundamental implications for development evaluation that have yet to be drawn.

With some exceptions, objective assessment of global policy is largely missing from development evaluations. Rich countries' policies are currently treated as exogenous in most development evaluations. Yet they are man-made constructs rather than laws of nature. They will have to be internalized and perceived for what they are: critical drivers of sustainable poverty reduction that are not immutable and can be adjusted.

In assessing project and country performance—for example, in agriculture—rich countries' policy distortions are currently not explicitly factored in. Hence, the responsibility for mediocre results is all too often attributed exclusively to developing countries and aid agencies rather than shared among all the actors responsible for the global policy framework within which development operations take place. In agriculture, the result is a reduction in aid flows just at a time when they should be raised.

This illustration confirms that the focus of policy analysis once exclusively directed to the south will have to shift toward the north. The unit of account that had gradually moved from the project level to the higher plane of the country assistance strategy will have to move up once again—to the global level.

Only then will the thorny problem of attribution in performance assessments be resolved credibly. Aid matters, of course (especially for the poorest, aid-dependent countries), but it is only part of the rich mix of policies that create the ultimate outcomes tracked by the Millennium Development Goals.

New Methods, Structures, and Processes

In other words, development evaluation should be globalized. This will involve new evaluation methods, structures, and processes. To avoid evaluation bombardment, flexible, tailor-made, collaborative evaluation systems and networks will have to be fashioned to combine legitimacy with rigor and efficiency.

Thus, joint evaluations must evolve toward multilateral forms while ensuring that the accountability of individual agencies is properly assessed in parallel. Only then will it be possible to distinguish income from performance. Only then will development evaluation put the country at the center of country assistance reviews and the global development community at the center of sector/thematic evaluations. In parallel, partial evaluation partnerships will have to be supplanted by system-wide networks.

Thus, the "global compact" view of the development process has implications not only for the methods of development evaluation but also for its structures. These will need to be adapted to a development business increasingly driven by global concerns and managed through partnerships and networks encompassing government, the private sector, and the civil society. The International Development Evaluation Network will have a major role to play in helping to set up such networks.

Joint responsibility for outcomes implies shared objectives, distinct accountabilities, and reciprocal obligations. If evaluation is to be judged by its results, then the processes through which evaluations are conducted will have to be owned by all parties and designed with the same amount of care as methodologies. Hence, greater creativity and innovation will have to be channeled into the design of joint evaluations so that accountability is served, transparency is assured, and all parties respect the independence and legitimacy of the process.

Joint responsibility for outcomes implies shared objectives, distinct accountabilities, and reciprocal obligations.

Fortunately, the building blocks for a new evaluation architecture exist. Capacity building resources are in place. They must now be reconfigured to make the structures of evaluation consistent with the creation of a level playing field for all development actors.

Learning from the Competition

In fulfilling its mandate, the development evaluation profession has a distinctive role to play, but it enjoys no monopoly with respect to its goals, methods, or skills. It must compete for resources and for the attention of development decisionmakers. Other professional groups with a global reach are far more powerful, extensive, and solidly entrenched.

Development evaluation (the new kid on the block) will have to use "smarts" to resist competitive forays, escape bullying, and protect a space it can call its own. To this end, it cannot afford to ignore four major professional groups active in the global sphere: (1) accounting and auditing; (2) management consulting; (3) policy research; and (4) voluntary agencies. In fact, it can play a useful role in linking them up to achieve results.

The accounting and auditing industry is especially strong and well resourced. Its prospects have been set back by recent corporate scandals. But it is recovering through a combination of modest reform and effective public relations. Deeply grounded in the global market economy, it is endowed with a far more cohesive, loyal, and disciplined membership than the freewheeling and somewhat disputatious evaluation profession. Its value-for-money audits and its comprehensive audits have long encroached into the evaluation domain. It is now muscling into the areas of environmental monitoring and corporate social responsibility reporting.

The global consulting industry is also a tough competitor. It has grown relentlessly over the past three decades through shrewd adaptation to the new opportunities offered by globalization, outsourcing, and the new information technologies. Recent excesses in the corporate world by major clients of consultancy firms (Enron and the like) have dimmed the luster of the profession and reduced the size of the consultancy market. Accordingly, consulting firms are in the process of downsizing and reorienting their practice by stressing the delivery of specialized services, experimenting with results-oriented contracts, and marketing themselves as business partners rather than as a passive channel of technocratic advice to management. They can also be expected to move beyond the private sector and to emphasize the public sector and the developing world as their frontier markets.

Policy research by think tanks and official agencies is another formidable contender for resources and influence. It is rooted in economics, the queen of the social sciences. It enjoys credibility with a wide range of policymakers. Recently, it has begun to address development effectiveness issues and it is championing evidence-based policy analysis and randomized social experiments. In an effort to add value to society, it is moving away from abstract doctrines and quantitative models. As a result, it is trespassing across disciplinary boundaries and providing leadership to the other social sciences.

Finally, the advent of a voluntary sector interconnected globally through the new information technologies has transformed the development business. Hundreds of NGOs have burst onto the development scene to deliver services, empower local organizations, and lobby for selected causes. They are subjecting development organizations, operations, processes, and policies to external verification and relentless advocacy. They are also teaming up with industry groups to subject private corporations to independent verification and social, environmental, and ethical scrutiny.

Reaching Out

Development evaluation cannot expect to overcome such behemoths through confrontation. Instead, it should demonstrate to each of these groups that they have much to gain from teaming up with each other—through evaluation. All four groups are grappling with issues of independence and conflicts of interest that development evaluation has had much experience with. None of the four groups enjoys the breadth of functional coverage that evaluation—the meta-discipline—does.

Accordingly, as an integral part of its partnership strategy, development evaluation should build bridges toward all four groups, reach out to their adherents, emulate their good practices, and help create value for them and for their clients.

There is little doubt that auditors' narrow focus on compliance can be broadened profitably through professional evaluations of the rules themselves. Similarly, models built by development researchers can be tested against the evidence through evaluation. Equally, the credibility of consultancy advice and the legitimacy of NGOs can be strengthened through more systematic resort to independent evaluation. Conversely, development evaluators have a great deal to learn by connecting to all four professional groups.

Specifically, development evaluators can emulate the listening skills of auditors and adopt their more rigorous documentation practices. They can benefit from the experience of their research colleagues through more careful handling of data, more rigorous use of statistical methods, and more effective use of peer reviews. They can acquire a better understanding of how to build client capacity by interacting with their consultancy colleagues. Finally, they can improve their outreach to disadvantaged groups and enhance their dissemination practices by adopting the proven methods of elite NGOs.

What then is to be done?

The distinctive strength of OED is that it is independent, committed to discovering the truth through excellence in evaluation, and focused on achieving results. Its distinguished track record gives it convening power and legitimacy. Conversely, the greatest risk faced by OED is complacency.

To manage this risk, OED should avoid comparing itself to other evaluation agencies. Instead, it should benchmark its independence to that demonstrated by the most respected auditing firms. It should compare the rigor of its methods in relation to that displayed by policy research organizations that are at the cutting edge. It should assess the quality of its client relations in relation to those nurtured by the best consulting firms. Finally, it should measure the quality of its participatory processes against the record of the most accomplished NGOs.

In addition, OED should leverage what auditors, researchers, consultants, and NGOs do by designing a distinctive evaluation strategy grounded in subsidiarity and complementarity. In a nutshell, the new development evaluation discipline should conceive of itself as a meta-profession combining all the knowledge disciplines working on development. It should create connectivity among them by promoting rigorous and fair performance assessments and principled learning partnerships. This falls well within its overarching mandate of accountability, learning, and capacity building.

Let me summarize. The development environment has changed. The ambitious Millennium Development Goals are not matched by sufficient resources. The development agenda needs to be expanded beyond aid to incorporate all global public policies. Hence, global policy reform should become a central focus of development evaluation, and a new evaluation strategy is needed to move to this higher plane.

New units of account, new methods, new structures, and new processes will have to be fashioned. Toward a new global evaluation architecture, a new phase of the OED renewal should be launched and new linkages forged with other global players—in the accounting and auditing profession, the development research networks, the consulting industry, and the voluntary sector.

The development evaluation profession at large will not adopt such an approach unless OED takes the lead. This means that you must start now and stick to the task until it has been mainstreamed. In this context, a strategic alliance between OED and IDEAS would help greatly in pushing the new development agenda forward.

If OED decides to adopt this course, it will need a substantial overhaul of its programs and processes. An alternative does exist, of course: for OED to go on doing what it has been doing successfully for the past ten years. This option is respectable and comfortable. Unfortunately, it is irrelevant.

14

Broadening the Constituency for Evaluation

Sulley Gariba
President of the International Development Evaluation Association

I come from Ghana, and I've been working in evaluation for almost half my life. As the new kid on the block, I will be very brief and straight to the point.

The good news is that a new set of players has come into the fold, into the science of evaluation, into the science of looking back and projecting forward. Over the years we have achieved success in what Bob Picciotto would call the mainstreaming of evaluation. More and more agencies are questioning what they are doing, and at the same time documenting the monumental development results that have been achieved, and using that to learn lessons and to create synergies that move the frontiers of development forward.

I want to chronicle three successive perspectives on development evaluation. The first has to do with the provision of public accountability, addressing the fundamental questions of value for money and where we're spending our hard-earned income and taxes. The central questions of accountability are clear now, but not so the questions of who is counting, what are we counting, and to whom are we accountable? In particular, evidence is emerging about the fuzziness of to whom we're accountable.

The second perspective has to do with making a case for more resources. Increasingly, officials in developing countries understand development evaluation really as a means to gather the information with which to make a case for more resources.

A third perspective, which is increasingly common, is the perspective of development evaluation as a matter of survival. It is on this third perspective that I wish to focus. Evaluation, if you come right down to it, is about institutionalizing cynicism, raising questions about the parameters of change, constantly asking, Why did this happen, how did it happen, to what effect is it happening? For hundreds of millions of people, the reality is about how the other half dies; the fundamental question has to do with their next meal. To them, evaluation is about the reality of their survival. Many millions of cynics about development effectiveness are out there, marching the streets daily, and they are asking the same fundamental questions that we as development evaluators are now codifying into a methodology for analyzing results. Thus in the discipline of evaluation, the disciples are increasing in number. Many of them have growing doubts about development effectiveness and, in particular, aid effectiveness.

Cast in this context, our tools of the trade are indeed in recession. Our tools may be enough to measure the results of, say, World Bank lending or of a particular project or program or set of policies, but they are not nearly enough to deliver the answers that millions of people are asking for.

Faced with this avalanche of potential converts to evaluation, we confront three important challenges.

The first is about convergence, between, on the one hand, people who specialize in development evaluation as a rigorous methodology that has to be learned and mastered, that is coherent, and that is science-based and logical, and, on the other hand, people who are meant to be served by development, recognizing that we all are interested in essentially the same questions.

The second challenge we face is to retool, bringing together what we already know and practice into synergy with the views of people who feel the pinch and pain of poverty. We need to engage more broadly in participatory evaluations with people in poverty, along with people who make the policies that affect the lives of the poor.

We need to engage more broadly in participatory evaluations with people in poverty, along with people who make the policies that affect the lives of the poor.

The third challenge, which has been talked about very frequently, is that we need to do a different kind of evaluation capacity development: not simply to equip developing countries and their professionals with technical skills for evaluation but also, maybe even more fundamentally, to unlearn some of the orthodox beliefs that we have about development evaluation, so they can blend culture with passion and with science. For example, I'd like to take a few of my colleagues at OED and other places back to my village to live in the context of that open society in which so many factors of life blend into each other, and in which the basic pursuit of livelihoods can turn a person's whole mind around. Without a baseline, people can really tell you how they feel today, how they feel about the development interventions that have gone on around them for decades, and may have actually made their lives worse. There is a lot about evaluation capacity development that we need to both learn and unlearn.

In that context, the International Development Evaluation Association has come into being at just the right moment—the moment when we're trying to build a global organization of development evaluators, colleagues who can begin to reinterpret the Millennium Development Goals, not simply as goals that we have set for making aid more effective, but as goals that will have a larger constituency behind them. That constituency of people who are out there questioning and critiquing our very intentions should be brought in to help formulate the ways to achieve the Millennium Goals. In this way, development evaluation can truly win the constituency that it deserves.

15

Politics, Aid, and Development Evaluation

Eduardo Wiesner
Former Minister of Finance, Colombia, and former Executive Director, World Bank

What are the critical challenges that development evaluation should address more intensively in the years to come?

Without pretending to be original on a subject that has recently received particular attention at the Monterrey and other global conferences, I would like to focus my remarks around what I would call the institutional failure or near-failure of the ODA community and multilateral institutions to induce in the developing world, and particularly in the poorest countries, sufficiently locally owned policies effective in delivering sustainable growth and distributive justice.

To me this is the most relevant priority for development evaluation in a world in which the relative importance of evaluating specific projects is declining. Policy-based lending and other modalities of non-project lending have become the main vehicles for transferring resources. But have these vehicles been effective in delivering their much-wanted development results and outcomes?

If they have not, why not? Within this vast and complex theme, which, of course, I am not the first to address, I shall offer only a few comments on three specific questions that we need to revisit:

- What is the nature of the problem we are facing?
- Could it possibly be that in many cases excessive or permissive financing is made available when a given problem becomes acute in the developing world?
- What enhanced role is there for development evaluation to play, inter alia bridging the gap between the supply of resources on the one hand and the demand for better results and outcomes on the other?

From my comments around these three interdependent questions, I will draw some preliminary conclusions and summarize the corresponding policy implications for development evaluation.

My basic message is that there is an insufficiently well understood collective-action problem in the ODA and multilateral environments. There may well be an institutional failure of the kind identified by Douglass North, in the sense that at the aggregate level—that is the macro-institutional multilateral arrangements—some wrong incentives may be at work. These could be engendering what Professor Dixit would call a commons problem, with multiple principals and agents, where conflicting objectives and asymmetric information pose daunting challenges for incentive theory.

Some wrong incentives may be at work.

These wrong incentives severely restrict the capacity of development evaluation to be effective or relevant. In brief, results and outcomes do not seem to match expected effectiveness. This is the typical situation where there is a latent demand for enhanced evaluation. This is the challenge now for development evaluation: to be more effective in its intermediating role between inputs and observed results.

Let me explain in more detail my first question. The problem we are facing is one of incentives and of the institutional framework in which they are embedded. For evaluation to be effective, it must be able to identify the specific origins of the restrictions that limit its capacity to intermediate between, on the one hand, the supply of resources for projects or alleged policy advice and, on the other, the results or outcomes that are sought.

In this respect, development evaluation must discern the wrong incentives and the unintended consequences of the right incentives. Among the varied sources of restrictions that development evaluation may face in the ODA and multilateral environments, two deserve particular consideration: first, political markets and political economy restrictions, and second, power considerations and state interests. These closely interdependent markets are much more incomplete and opaque than economic markets, and yet key bilateral and multilateral aid and credit choices are there nurtured and made.

Often these multilateral public choices override private market choices and even technical or efficiency and equity-based choices. Power and political considerations may explain why about 70 percent of foreign aid is still spent bilaterally, even though studies show that multilateral organizations reach the poor more accurately.

This means that it can happen that foreign policy concerns will be reluctant to yield to lessons from purely technical evaluations. If it is politically correct to support some countries or projects independent of what evaluations may advise, these political concerns may have the upper hand.

These political economic realities lead me to my second, almost implausible, question. Is there excessive or permissive financing from ODA and multilateral institutions?

This is a difficult question to answer because one would need to define what excessive means. That is, excessive with regard to what? In principle, it should be with regard to the demand for development resources to facilitate the implementation of reform in developing countries.

Under this approach, if that demand is real and growing in the sense that aid is actually buying reforms in developing countries, then there could hardly ever be an excessive supply of funds. They would just flow in a market-like fashion from the ODA community toward the reforming countries. In brief, excessive or permissive financing is that which exceeds what would be warranted by the quality of policy reforms it actually induces.

Another reason why this is a tough question to handle is because it is known that ODA flows and financing for development objectives are unstable. However, they are unstable in the aggregate, and my remarks about possible permissive financing are directed to segments or mini-markets within that aggregate. If one looks at the history of bailouts under the Heavily Indebted Poor Countries Initiative and numerous other examples of refinancing, restructuring, and outright debt forgiveness, the question of whether financing is permissive for specific cases, countries, and groups still has to be considered.

In essence, there could be two answers to this question. The first is that, yes, financing was excessive because it did not generate sufficient adjustment, and a difficult situation became worse and turned into a crisis. Since sufficient adjustment did not take place, clearly the financing was excessive; it may even have contributed to the final collapse as reforms were postponed as long as there was soft financing available. The second possible answer is that, no, there was no excessive financing; on the contrary, financing was insufficient since it failed to prevent the final unraveling of events.

Under both of these possible answers, we face a familiar tautological problem: each side could proclaim itself right, in having predicted the result that confirmed its own premise—but after the fact, and we know that what really matters is having predictive power before the fact. Under both of these possible answers, as Ann Krueger has rightly observed, the World Bank and the International Monetary Fund "are perhaps bound to be criticized regardless of what they do." And yet, under the second, one has to wonder, at what point would the supply side of financing finally say no? If it continues to provide financing, will this not be excessive at some point? (And whether by saying no, suppliers can avoid or fall into a moral hazard situation is another topic, which for reasons of brevity I cannot discuss now.)

At what point would the supply side of financing finally say no?

If a good portion of the financing flows not so much to buy real reforms but in response to other political or power considerations, then why have evaluations? If the results of evaluations do not substantially change access to resources, again, why have evaluations?

Why have them beyond some formality that gives the impression that they are taken seriously? If countries in difficulties receive financing that substitutes for locally owned financing, which has to come from above the line, again, why have evaluations?

In other words, I suggest it may be precisely because there has been permissive financing that reform results are so lackadaisical and that development evaluation has been so severely challenged.

The problem of verifiability emerges as one that permeates the evaluations and the lessons that can be drawn.

However, according to a professor whom we all respect, Dani Rodrik, my argument that financing is excessive or permissive would be wrong, because multilateral institutions do try to protect their lending from political demands. This is tantamount to saying that borrowing countries do try to adopt reform policies in exchange for loans and resources.

This possible criticism could be answered in two ways. First, all multilateral lending has guaranteed repayments through the use of very strong and enforceable legal and political instruments. After all, national governments guarantee multilateral lending and virtually 100 percent is repaid. But second, and this is the real issue, a policy-based loan (bilateral or multilateral) may be repaid, and meet its ex ante conditionality, yet leave many pending questions about how much lasting reform has been achieved. This can happen for many reasons. One of them may be that when the loan operation was configured, those in charge of the design were subject to the wrong collective-action incentives. They may have found themselves under terrible dilemmas and conflicting interests, facing problems of adverse selection, moral hazard, and asymmetric information.

In the end, the problem of verifiability emerges as one that permeates the evaluations and the lessons that can be drawn. In brief, it can happen and has happened that loans have been granted as the result largely of political pressures and multilateral public choices, rather than of recommendations from detached economic analysis. This is not plainly wrong or reproachable. But what really matters is whether those policy and power choices are themselves the subject of independent evaluations and whether those who make those choices are accountable for the results.

So, to turn to my third question, what enhanced role should development evaluation play? Within a broad context in which the supply of development finance is subject to closer scrutiny, and the results of the use of the resources appear to have so much margin for improvement, the potential role for development evaluation seems to be immense.

Few situations can reveal this more eloquently than the remarks made by Paul O'Neill, America's Treasury Secretary, when in mid-2001, he stated: "We have too little to show for the hundreds of billions of dollars spent in the developing world. It is clear that we have a major repair and restoration problem before us." This call for enhanced effectiveness in the use of aid and credit resources for development is, in fact, a call for development evaluation to play a much greater role in bridging the gap between results and policies. It is a demand for development evaluation to be more effective and relevant—a daunting challenge and a formidable task in a world in which incentives, interests, information, perceptions, and even emotions are difficult to align in pursuit of the collective interest.

Development evaluation should support inter-disciplinary research on political power and economic markets.

Within that evolving environment, I suggest that development evaluation should, first, focus on identifying what incentives for collective action are present in the fast and massive growth of policy-based lending, and whose interests they serve. This is not an easy task. Second, it should focus on evaluating the entry and exit quality of the individual policy and other operations that are configured as counterparts for emergency or sectoral lending that is really undertaken as balance of payments support. Third, it should evaluate the macro-institutional arrangements through which different multilateral institutions establish conditionalities and agree to financing for problems that often the private markets are no longer willing to deal with. More generally, development evaluation should support interdisciplinary research on political power and economic markets. And the work of independent commissions like the Meltzer one should be supported and more fully utilized.

Although we all know that the IMF, the World Bank, the regional development banks, and bilateral agencies are strengthening their work in development evaluation, these endeavors need to be more closely examined from the perspective of their combined results, and in terms of the learning that results for these institutions and for developing countries. Those of us busy in the field of development evaluation should go to the primary stakeholders from the supply side and inform the taxpayers in Helsinki, Palo Alto, and London that their taxes may not be really buying the type of international public goods or development public goods that they think they are buying. We should also approach the no less important stakeholders on the demand side, in the developing world, and inform them that results are less than desirable because development evaluation is constrained unless there is real demand for reform.

New research is needed. Long-term development evaluation research programs must be strengthened. Institutions like the International Development Evaluation Association have a major role to play in this respect, not so much in conducting research directly, but making sure that what we know is used and that we clearly identify what we do not know.

Let me summarize my conclusions.

First, the unit of analysis of development evaluation is shifting away from projects and toward support for sectors, policies, and wide-reaching social and political reforms. The transfer instruments for this support are difficult to measure and their design, results, and outcomes are difficult to evaluate.

Second, in the institutional governance structures of multilateral and bilateral credit, the wrong collective action incentives may be at work. These incentives or their intended consequences may lead to permissive or excessive financing that responds more to political and power interests than to technical efficiency or equity objectives.

Development evaluation should focus on identifying wrong incentives for collective action and should inform supply and demand stakeholders about their consequences.

Third, if financing is permissive, there is little room for demand-driven development evaluations: if a good portion of the financing flows not so much to buy real reforms, but in response to other political or power considerations, there is no substantive reason why evaluations should take place. The dearth of demand-driven development evaluations may explain the lackadaisical results of foreign aid and bilateral and multilateral lending for policy restructuring or for balance of payments crises. Although the markets end up being the final evaluators and the adjudicators of costs and benefits, they do so largely anonymously and peremptorily. Under these circumstances, policy corrections or the lack of them can have high social costs and political accountabilities can be diffused, which is probably the highest cost of all.

What policy implications should one draw from these conclusions? The key policy implication for development evaluation is that it should focus on identifying wrong incentives for collective action and should inform supply and demand stakeholders about their consequences.

Another key implication is that development evaluation should focus on helping policymakers make the right choices about what to evaluate and under what circumstances. When a technical recommendation is overruled by political considerations, development evaluation should be there to assess the experience and to assure that learning takes place.

Independent evaluations by high-level commissions under the auspices of the principals of the system as a whole should be established. Commissions like that of Alan Meltzer and others are very important.

I am aware that I have left out many answers to the questions I have posed and that I have gone out on a limb in some of my arguments. Also, I have left unrecognized laudable research work by highly regarded professionals and institutions. But I trust these omissions will be treated with some benevolence—hopefully with the same benevolence that was given to those proposals of 30 years ago that resulted in the World Bank's establishment of OED. These were finely polished and took root, and look now, how far we have advanced.

QUESTIONS AND ANSWERS

Salah El Serafy: I have worked for OED on several occasions. I am very skeptical about widening project evaluation into development evaluation. In project evaluation there are certain triggers, certain benchmarks that start the process of evaluation, and you evaluate against various defined targets, and the process is fairly easy. When it comes to development evaluation, development is an elusive concept with so many dimensions, and evaluators lack defined targets, or a clear idea of intentions, against which to evaluate. Is there any intention in the Bank to define development ex ante according to certain benchmarks?

Mr. Gariba: The whole notion of the elusiveness of development is a pertinent one, and so is the danger in extrapolating from project evaluation to the evaluation of development itself and of development effectiveness. But if one assumes that the current paradigm of the Comprehensive Development Framework—and with it the poverty reduction approaches that are country-led and country-owned—defines the vision of development, then one ought think about how to scale up development evaluation, in terms of methodology and inclusiveness, so as to look at the effectiveness of the goals that developing countries are setting themselves. This entails looking at notions of accountability for the kinds of policies that are implied in this paradigm shift. So even the politics of the perceived paradigm shift need to be included as one of the parameters of what we're evaluating. Thus we're not just evaluating the end result or stated goals, but *also the very processes* of defining those goals. And are we in the final analysis prepared to engage in the politics behind these policies and the goals that are set?

Mr. Picciotto: Basically Salah is saying that we need to have benchmarks, specific goals against which to judge results. Much of the money for development now is going for policy reform and program lending and things of that kind, and those are presumed to have benchmarks; they have conditions that are presumed to deliver results; they have tacit theories behind them. The difficulty of development evaluation is to explicate those theories: first of all figure out what was the implicit theory and then assess whether the theory holds or not, and if not, why not. This is where evaluation differs from auditing in that we have to go back to the goals themselves and to how the particular policy inputs are supposed to deliver results.

One of the problems we are facing in this area is really that economists, macroeconomists more than microeconomists, have tended to take for granted that certain policy inputs will deliver results. And then I think the difference between economists, particularly macroeconomists, and evaluators is that evaluators give more importance to context.

I think evaluation should bring in the other social sciences, along with economists, to try to illuminate what actually happened, and, say, whether the lending was permissive without intending to be. Also, given the fact that the World Bank has moved its own role beyond projects to help coordinate programs, there is really no option but try to figure out what this means for evaluation. Simply to say we should stick to projects really means consigning evaluation to irrelevance.

The Millennium Development Goals have the tacit hypothesis that aid can deliver poverty reduction, but in fact this tacit hypothesis just doesn't hold. Poverty reduction is a result of a large number of factors that are very context-dependent, and that need to be elucidated up front and assessed objectively. I hope this is what eventually the Monterrey Consensus is going to lead to.

16

The Challenges of Development Evaluation: Some Conclusions

Gregory K. Ingram
Director-General, Operations Evaluation

The foregoing chapters make clear that development evaluation has emerged as an important sub-field within the evaluation profession. Relative to other kinds of evaluation, what is different about development evaluation? I suggest there are two key dimensions.

The first is context. Development evaluation typically occurs in very demanding circumstances. Data are scarce, and data quality is often suspect. Baseline information is rarely available. The capacity of local institutions to carry out evaluation is often weak, as is the capacity of the institutions that then need to apply evaluation lessons in the client countries. And cultural differences can be a stumbling block. Offsetting this, there is often a hunger for relevant lessons of experience.

> There is often a hunger for relevant lessons of experience.

The second major dimension of difference is the changing concept of development. Evolution in our thinking has led to different kinds of development interventions over time, and new interventions often require new evaluation approaches. For example, evaluation has expanded its concept of capital from productive capital to human capital to social capital. Central planning approaches have been replaced by market-based approaches that emphasize international trade and the private sector. Macroeconomic stability and appropriate policy frameworks are now thought to be necessities for growth. Poverty reduction has become the major objective with new instruments in education, health, and social safety nets. Government ownership and citizen participation are key elements of strategies that now focus on countries rather than on projects or sectors. And partnerships are an important means of doing business. And our strategies now have a new focus on results as embodied in the Millennium Development Goals.

The dynamic of development thinking has kept development evaluators moving rapidly to keep up. Of course, evaluation of development experience has helped promote some of the new thinking about development, and it has been instrumental in fostering a greater focus on the results from development programs.

Annexes

Annex A. Reflections from Former and Current OED Staff

None of the efforts of OED's management during these first 30 years could have been effective without the hard work and dedication of the department's staff, including evaluators and support staff. As part of the retrospective, a number of staff were interviewed on their experiences in OED. These interviews focused on the nature of the work, what the staff members gained from their experience in OED, and what they thought the effect of that work was on the Bank and the development community more generally. These interviews were the basis for a retrospective videotape first presented at a forum in September 2002, and subsequently used at a number of events to increase awareness of how OED had contributed to the Bank's work since 1973.

This section presents excerpted portions of the interviews with seven current and former OED staff members. Collectively, these staff cover virtually the full period of OED history, from the 1970s until today, and represent a wide range of skills and perspectives. Their stories provide rich detail on how OED works and how it connects to the rest of the Bank and the development community more broadly.

Perhaps more importantly, these reflections give us insight into the way that working in OED has influenced the professional growth and development of the staff themselves. The perspective gained from evaluation work and the way it can change the thinking and actions of those who undertake it are strikingly captured by the staff member who said, referring to her time in OED: "Learning about the Bank and what makes it effective and what makes borrowers effective is . . . at the very core of the Bank's mission."

On Joining OED

Susan Stout
Former Operations Adviser

I joined OED with some trepidation. Its reputation—for many of us who joined the Bank in the early 1980s—was as a kind of retirement home. This is where you went at the end of the day to review projects after they were completed. It was not the front edge of the wedge. It was also considered quite a conservative part of the Bank, unwilling to change the way it did business.

Yet working for OED became the most satisfactory job I have had in 20 years at the Bank. First, it is the only place I know of in the Bank where you can get an opportunity to understand what the Bank has accomplished. And that is an enormous sum.

Second, not many people know how interesting learning from the past can be. One of the disparaging comments from my friends, when I told them I was thinking of joining OED, was "Oh, that's just history." And, now I come back aggressively and say that evaluation is the core discipline. Good history is fundamental to learning. Evaluation is learning. Learning about the Bank and what makes it effective and what makes borrowers effective is, in my view, at the very core of the Bank's mission.

Third, I found the responsiveness of the Bank's management and its staff to the lessons of evaluation very encouraging. Our lessons from the health study I worked on were not particularly pretty: we found lots of things that the Bank was not doing well in health, nutrition, and population. One could easily anticipate that people doing the day-to-day work would react very negatively to those findings. It turned out to be exactly the reverse. I found that the task managers, sector managers, and even the leadership of the sector welcomed the clarity and honesty of OED's treatment of the issues. The study was done objectively and with the best methodology we could think of, and they appreciated its clarity. That reinforced my own confidence in professionalism and commitment to the effectiveness of evaluation.

On the Shifting Emphasis of the Bank's Lending

Gladstone Bonnick
Former Principal Evaluation Officer

When I joined the Bank in 1978, sector lending and sector programs were receiving a lot of attention. The emphasis of the economic work was mainly at the micro level, focused on seeing that appropriate sector policies were in place. After that, there was a shift toward structural adjustment lending, where macro considerations become important. And here the Bank was primarily lending on the basis of a joint effort with the IMF, in which the IMF would look at the macro side of things, and the Bank would continue to look at the micro side of things.

This led OED into considerations regarding the sustainability of programs. Especially with regard to structural adjustment lending, we found that many programs were started, money was disbursed, and then later on, policies were reversed, and hence the effectiveness of the programs was compromised. So we were concerned that policies once put in place should be sustained. This meant that the policies should not just be owned by the government in office, but should in a sense also represent the ideas that possible alternative governments might have regarding the country's development strategy.

In evaluating these programs, OED had to develop its sensitivity. When you look at the policy dialogue, you need to look not just at the technical content of that dialogue, but also at how the dialogue has been managed and the extent to which diplomatic sophistication has been built into the dialogue. And I think that OED itself has been concerned to foster a dialogue among clients and indeed among donors, with an appreciation of the need to keep on evaluating the work that is being done, to see what lessons can be learned and how that work can be improved. So I think that there has been a very good experience there for OED.

On the Experience of Working in OED

Lynette N. Kwaku
Former Executive Assistant

What I learned from OED: OED presented to me the other facet of the Bank, to show the completeness and the thoroughness with which the Bank conducted its operations. Until my assignment in OED in 1984, I had been with Operations for a number of years, and I knew only the loan process and the project cycle, but I was not aware of this other arm because I do not think we had any interest, really, in feedback from OED. But when I started to work with the Operations Evaluation Department, I realized that this is a very important aspect of Bank work.

Based on my experience, having had the opportunity to be exposed to OED, I think it is a positive thing for all staff to rotate around the Bank—to get to see all of the aspects of the Bank's work, how everything is interrelated, and that each unit has a unique and important role to play, especially in our relations with borrowers and donors.

On Country Assistance Evaluations

Rene I. Vandendries
Lead Evaluation Officer

Shortly after I joined OED in 1988, the Bank's emphasis started to shift from individual projects to a country as a whole, and we started paying much more attention to country assistance strategies. OED started its country assistance evaluations (CAEs) in 1995, and I have been involved heavily since the beginning.

What I found out right away is that these evaluations help you see the forest. What we had been looking at earlier was the trees—the individual projects—and what the country evaluations taught us very quickly was that the forest can be quite different from what the individual trees tell you. Producing the country assistance evaluations has been a challenge, particularly as regards methodology, but that has improved over time, and I am sure we shall come up with an excellent methodology in the very near future.

A very important aspect of CAEs is that you need to know who your clients are. These clients are of different kinds. We have the Board as the main client, and my experience has been they have been always very receptive of CAEs. Another client is the government, who would like to learn from what we are learning about the country and what the Bank is doing in the country. Governments, on the whole, have been also very receptive; and, in some cases, actually, the government gets involved in writing the evaluation with us. Bank management is, of course, another client, and there we have occasionally had some frictions because it is very natural that everybody is a little vulnerable to criticism. But I have learned over time that even though management may be unhappy with what you have to say at a given moment in time, two or three years later, things change. I have seen management being upset originally and then two years later they said, "Well, you know, you were right, after all, and you did us a good service." So that to me attests to the sustainability, the durability, of the country assistance evaluations.

On the Importance of Partnerships and Stakeholder Involvement

Andres Liebenthal
Lead Evaluation Officer

In the mid-1990s, the Bank was facing a lot of controversy in its dam-related projects. The controversy arose from the growing awareness of the environmental and social impacts associated with dam projects. In relation to resettlement, in particular, these projects had a lot of repercussions—not just in terms of discussion, but of attempts to stop the projects. As a result the Bank decided that it would be good if OED could evaluate actual results in a consistent framework, looking at the benefits of the projects versus the costs and whether the projects were justified or not, even taking those environmental and social aspects into account. I took over that study when I joined OED.

OED's report was well received. It concluded that most dam projects financed by the Bank had produced very good benefits, but that some of them had not met what are currently known as resettlement and environmental guidelines. The Board suggested that in light of the controversy, some further evaluation on dams would benefit from cooperation with other stakeholders. So a workshop in Gland, Switzerland, was organized in partnership with the International Union for the Conservation of Nature to bring together some of the leading proponents and opponents (from industry, NGOs, governments, and the environment community) in a neutral framework and setting. This workshop led eventually to the creation of the World Commission on Dams—with the mandate of evaluating the global experience with large dams, and whose report presented a list of recommendations on the ways dams, should be built, consulting those to be affected, and addressing social and environmental impacts.

What I got out of this particular study was a sense of the importance of working in partnership beyond the Bank. For issues beyond the technical and strictly economic, it is very important to take into account the interests of other stakeholders and to find efficient ways to meet their concerns. From this experience, we certainly learned a lot about how to build such partnerships, what to expect from them, how to manage them, and also how to apply lessons in other work that we have been doing in relation to, say, forest policy and the extractive industries.

On the Evaluation of Global Programs

Uma J. Lele
Senior Adviser

There was a lot of interest in the Bank's Board about the growth of global programs, and perhaps questions about how well the Bank was managing its strategy at the global level. OED's review of the Bank's approach to global programs was prompted by these concerns, and partly stimulated by the Forestry Review I had conducted earlier, which explored the divergence between global objectives and local objectives.

The scope of this study is huge. When we looked at partnerships, we found an inventory of more than 200, and then selected 70 programs in all networks and all sectors to do an overview. And now we are doing reviews of 26 global programs, which is quite a large exercise.

What have we learned from the study? The first phase looked at the Bank's internal management, and at the strategic and programmatic management by the Bank of its strategy on globalization. We made a number of recommendations on how the Bank could manage this global strategy more effectively. Our initial discussions with the Bank's management created a lot of waves; our findings did not go down that well, because people did not want to hear that maybe there is some scope for improving the way the Bank is doing business.

But I am told that there is a big difference now. There is quite a lot of thinking among management about how to improve internal management, and at a recent briefing on the (2002) Johannesburg meeting, the Bank's president was saying that the Bank is realizing that it needs to get its act together on partnerships if it is going to be a better partner with institutions outside. I was very pleased to hear that, because the first time I made a presentation to management at Corporate Day, the reaction was not that positive. It has been very satisfying to see that this globalization review has contributed to something key in the Bank.

Annex B. OED: The First 30 Years

A Chrestomathy

The fiscal year 2003 (beginning July 2002) marks the 30th year of the establishment of the Operations Evaluation Department at the World Bank. This note briefly traces—through a select "chrestomathy"*— the evolution of the evaluation function at the Bank, some of the emergent challenges, and the institution-wide response to these challenges.

The 1970s

THE BEGINNINGS

I feel that, with the growth of the Bank Group, the time has come to carry our systematic auditing procedures a stage further. I have therefore decided to establish a unit whose sole responsibility will be to evaluate the contribution of our operations to the development of member countries. The principal task of this unit will be to review past lending operations with the central objective of establishing whether the actual benefits of the completed projects are in accordance with those expected at the time of appraisal and, in case of divergence, the reasons. I believe that this initiative will have two principal benefits. It will contribute to the formulation of policy and procedures by enabling us to learn more from our past experience. And it will give us a degree of conviction about the impact of what we are doing which we could obtain in no other way.

Administrative Circular: *Establishment of an Operations Evaluation Unit,*
signed by Robert S. McNamara, President, September 2, 1970.

*The word "chrestomathy" was formed and first used in Greek, from the words *khrestos,* useful, and *mathein,* to know, hence useful learning. Its most frequent sense these days is that of a selection of passages designed to help in learning a language. The most popular title is that of *A Mencken Chrestomathy,* a selection of the choicest writings of the American journalist and writer on language, the late H.L. Mencken.

Source: Michael Quinion, *World Wide Words* (1996).

A distinction may usefully be drawn between "Audit" and "Evaluation." In the work of the Operations Evaluation Division "Audit" has come to mean comparison between the targets and projections contained in the project appraisal reports on the one hand and actual developments on the other, in order to see whether or not appraisal objectives were attained; if not, why not and if so, was it due to the project? "Evaluation," on the other hand, has come to mean a deeper analysis of the contribution to development made by the project and by the Bank in its support of the project, with a view to seeing not only whether the project met its assigned objectives but also whether these objectives were themselves correct in retrospect and whether they might have been improved in any way.

Memorandum by the President to the Executive Directors:
Future Work Program in Operations Evaluation, July 20, 1972.

INDEPENDENCE OF THE EVALUATION FUNCTION

The unit responsible for this work has always had a high degree of independence from management, and it has grown in stature and experience its independence has been increasingly formalized most recently by making it a separate department in July 1973 under a Vice President without other direct operating responsibilities and giving it a formal link with the Joint Audit Committee of the Executive Directors in November 1973.

Report of the Joint Audit Committee for 1973, October 25, 1973
(Document R73–243/1); and *Terms of Reference of Joint Audit Committee,*
November 21, 1973 (Document SecM73–704).

I think that the time has now come for the growing link between the Executive Directors and the Operations Evaluation Unit to be recognized formally by drawing lines on our Organization Chart directly linking the unit, and the Vice President in charge of it, to the Executive Directors as well as to the President. I also propose that, following the retirement in 1975 of the Vice President presently responsible, and in any case not later than December 31, 1975, a title of Director-general, more indicative of independence from the normal operations of the Bank, should be instituted for the manager of this function. Appointees would hold office for renewable terms of five years, be removable only by the Executive Directors and be ineligible for subsequent appointment or reappointment to the Bank Group staff except in unusual circumstances. The basic work program of the unit would be reviewed and approved periodically by the Executive

Directors and all reports of the Operations Evaluation Unit would be sent simultaneously to the Executive Directors and to the President, without any interference by the latter with their content.

Memorandum by the President to the Executive Directors:
Operations Evaluation, September 27, 1974.

THE EVOLVING OPERATIONS EVALUATION SYSTEM

Project Performance Audits will be prepared on all projects in developing countries about one year after completion of loan/credit disbursement (i.e., typically 4 to 8 years after approval by the Board of Directors). The feasible sophistication of this enquiry, one year after completion of Bank/IDA disbursements, will vary considerably among projects, depending on the nature of the objectives sought, the sector to which the project belongs and the duration of disbursements.

Memorandum by the President to the Executive Directors:
Operations Evaluation and Project Performance Audits, April 13, 1973.

This report is a summary review of the first 50 projects covered by Project Performance Audit Reports which were issued between October 1972 and March 1975. The overall average economic rate of return to the $3,200 million of investment is now estimated at about 15%, and this figure is not very dependent on the changes in relative prices that have occurred in the last few years. This is considered a good performance on the part of the Bank as a development lending institution, especially in view of the difficult and changing environments, natural, political and economic, in which many of the projects were implemented.

First Annual Review of Project Performance Audit Results,
September 26, 1975.

In accordance with the basic principle that, for most effective learning and impact on future operations, evaluation should be carried out by those directly involved in the operational side of the Bank's work, most of the evaluation activities are of a decentralized and self-evaluative nature.

Operations Evaluation, World Bank Standards and Procedures,
June 1976.

It has long been Bank practice to require progress reports from its borrowers. This is a regular part of the supervision process. In the past, these reports have focused primarily on the typically clearly specified and monitorable activities for which project monies are being disbursed (such as the construction of a road, an irrigation facility, a power plant or a school, or the commitment of monies by financial intermediaries to sub-borrowers), and related institution building. They have not focused as consistently on the impact of the projects on the society and on the intended beneficiaries; especially where, as in the fields of agriculture and rural development, education, population and urban development, special efforts are needed to establish a measure of the situation the project is intended to alter if the changes effected by the project are to be observed.

Operations Evaluation—Annual Report, October 13, 1976.

The 1980s

FROM PROJECTS TO POLICIES

Policy-based lending is appropriate in the right circumstances, whether it is called program lending, as it was called in the 1960s, or structural adjustment lending, as it is called in the 1980s. But my own view is that policy-based lending is not the answer to all development problems; there is a role for intelligently conceived and well designed investment project lending. But most important is the role of intelligently conceived and managed country lending programs. These lending programs are for me the key operational instrument of the Bank; that is, the aggregation of the structural adjustment loans and the sector loans and the project loans which, all together, can be designed to focus on an agenda of issues with the countries. The real Bank lever, as I see it, is neither an SAL nor a project, but the set of activities called the country program.

Transcript of interview with Mervyn L. Weiner
(first DGO, 1975–84), July 1986.

One of the most important services the Bank can provide is to ensure that the process of policy reform is "internalized" in the country as quickly as possible, so that the reform program is designed by the country itself and integrated into its long-term development program.

"Lessons Learned and Implications," in *OED, Structural Adjustment Lending: A First Review of Experience,* September 24, 1986.

Structural Adjustment Loans (SALs) are rather difficult to evaluate because structural adjustment involves the interaction of many variables about which we still have much to learn. The first SALs were implemented in a very difficult world economic environment. SALs were a rapidly evolving product and their evaluation was like focusing on a moving target.

"All about OED," Interview with Yves Rovani,
The Bank's World, December 1986.

BANK-COUNTRY RELATIONS (AND SPECIAL) STUDIES

The Bank's successful mediation efforts in the 1960 Indus Waters Treaty between India and Pakistan has awarded the Bank the benefit of doubt ever since. [In Pakistan], the boundary line between the givers and receivers of aid became a fluid one. When necessary, the Planning Commission and the Finance Ministry could seek the alliance of the [Aid to Pakistan} consortium in persuading other departments to accept a change in policy. On other occasions, the World Bank, the USA, or some other members of the consortium could seek the alliance of the Government of Pakistan in order to impress its own view on the consortium as a whole.

OED, *World Bank in Pakistan: Review of a Relationship,* 1986.

OED IN THE REORGANIZED BANK

The mandate and the independence of the Operations Evaluation staff have been reaffirmed by the reorganization. Internally, our independence helps ensure the choice of subjects, access to information and candor of reporting necessary for accessing the accountability and for drawing fully on the benefits of experience. Externally, it contributes to the credibility of the Bank's achievements by demonstrating publicly its openness and willingness to be the first to recognize its own shortcomings. The increased challenge is to revitalize: evaluation, the process by which we account for results and draw lessons from experience; dissemination, the way we share the experience across the Bank; and feedback, by which experience is reflected into new operations.

Yves Rovani, "The Bank's World: An Editorial," January 1988.

The 1990s

ENSURING QUALITY AT PROJECT ENTRY

To improve upon the Bank's portfolio performance management, the Task Force advances six principal recommendations and a comprehensive program of measures for their implementation: (a) introduce the concept of country portfolio performance management linked to the Bank's core business practices; (b) provide for country portfolio restructuring in adjusting countries including the reallocation of undisbursed balances of loans/credits; (c) improve the quality of projects entering the portfolio through better project analysis, appraisal methodology, and contractual arrangements; (d) define the Bank's role in, and improve its practices of project performance management; (e) enhance OED's role as an instrument of independent accountability and refocus ex post evaluation on sustainable development impact; and create an internal environment supportive of better portfolio performance management.

Effective Implementation: Key to Development Impact,
Report of the Portfolio Management Task Force
(the "Wapenhans Report"), September 22, 1992.

SHIFTS IN EVALUATION CRITERIA

Conceptually, the Bank's evaluation criteria have evolved to reflect the shifts in development paradigms over the decades. Gradually, the economic evaluation of projects became far more comprehensive and multidisciplinary than before. The main "unit of account" in the Bank's operational work shifted from the project to the country. The full mix of lending and non-lending services that make up country assistance programs has become the dominant preoccupation of development policymakers. Sustainability of projects and the institutional development impact of projects began to be systematically rated beginning in the early 1990s. Priority attention began to be been given to the relevance (in addition to efficacy and efficiency) of project goals reflecting country assistance priorities.*

Robert Picciotto, "Evaluation in the World Bank: Antecedents, Methods,
and Instruments," from *The International Atlas of Evaluation,* 2002.

* An overall development effectiveness index has recently been introduced to reflect the combined consideration of outcomes, sustainability, and institutional development impact.

REAL-TIME EVALUATIONS

Nearly fifty years of experience indicates that the vast majority of the Bank's operations do in fact comply with our own policies and procedures. The Inspection Panel provides a safety net in exceptional cases where those high standards are not met-and in that sense, offers a link between the Bank and the people affected by the operations we help finance.

World Bank News Release: "World Bank Establishes Independent Inspection Panel," September 23, 1993.

The Bank's Strategic Compact established a "stretch goal" of zero defects in Bank workmanship. To that end, [Quality Assurance Group] QAG was established in 1996 with a mandate to increase accountability by conducting real-time assessments of the quality of the Bank's performance in its major product lines. An additional objective of QAG is to enhance learning and identify best practice, so as to catalyze changes in policies, programs and procedures.

QAG Web site.

PARTNERSHIPS AND EVALUATION CAPACITY DEVELOPMENT

Poverty reduction was the theme of the third Biennial World Bank Conference on Evaluation and Development, held in June 1999. The conference was sponsored jointly by two Word Bank departments, OED and the World Bank Institute. It brought together more than 250 representatives of the international development evaluation community, including bilateral and multilateral donor agencies and civil society. Conference participants discussed lessons learned from evaluations of poverty reduction programs; ways to advance the methodological frontier; partnerships, ownership, and participation in evaluation. The participants also explored challenges for the evaluation of poverty reduction programs, both in methodology and in setting substantive priorities.

Evaluation and Poverty Reduction, World Bank Series on Evaluation and Development (Volume 3).

The main precondition for developing a national evaluation system is country demand. . . . There are particular risks if the impetus for an evaluation system is donor-driven. . . . Unless and until countries reach their own conclusions about the merits of evaluation, an evaluation system is unlikely to be sustainable.

Public Sector Performance—The Critical Role of Evaluation, Selected Proceedings from a World Bank seminar, 1998.

The 2000s

TOWARD A COMPREHENSIVE DEVELOPMENT AGENDA, AND AN INCREASINGLY COMPLEX MANDATE

As it seeks to implement an increasingly crowded agenda, the Bank must learn to reconcile "client ownership" of activities with the Bank's corporate priorities; to adapt global knowledge to local conditions; to balance support for countries with strong enabling environments with support to countries with poor enabling environments but a high incidence of poverty; and to be efficient and selective in implementing a holistic vision of development, making judicious use of partnerships.

Annual Review of Development Effectiveness, 2000.

The Bank has also taken upon itself the goals of generating and disseminating development knowledge, engaging its partners in global public policy debates, and dealing with global development challenges that cannot be tackled effectively at the country level. These institutional responses have given rise to tensions and trade-offs among goals.

Today's interview with Robert Picciotto, February 22, 2001.

SHARING KNOWLEDGE ON DEVELOPMENT EVALUATION

The impetus for the program originates in a growing global demand for professional evaluation of development policies, strategies, programs, and projects. A rapidly changing environment, increasing public pressures for accountability and results, as well as an emerging consensus behind more comprehensive development strategies underlie the keen interest in acquiring up-to-date evaluation knowledge and skills within the development community.*

Today's interview with Ray Rist (Senior Evaluation Officer), July 23, 2001.

We the participants of the Preparatory Meeting on IDEAS that was jointly sponsored by UNDP Evaluation Office and the World Bank Operations Evaluation Department, representing evaluators and development practitioners, mostly from the developing world,

---

* July 2001 saw the launch of an annual summer program for executive training in development evaluation, the International Training Program in Development Evaluation Training (IPDET), an effort of OED in collaboration with Carleton University in Ottawa. Approximately 130 participants from 41 countries attended the first offering of this annual event.

declare their commitment and support towards the establishment of an International Development Association (IDEAS). We believe that IDEAS would serve as an effective platform advocating development evaluation as a key tool for managing for results, promoting transparency and accountability in decision-making and governance processes, and facilitating knowledge learning on evaluation and development at the global level.

London Declaration on the Establishment of IDEAS, 7–8 May 2001.

Annex C. Institutional Changes for Independent Evaluation at the World Bank: A Chronology (1970–2002)

Before 1969

- Bank auditing consists of a three-person internal auditing unit; work is carried out by external auditors.

1969

- Internal Audit becomes an independent office reporting to a vice president.

1970

MAY

- Joint Bank-International Finance Corporation (IFC) Audit Committee (JAC) is established to discuss, among other things, the scope and results of Internal Audit and its relations to external auditors. JAC aims to "strengthen the assurance of the integrity of the financial statements as a realistic reflection of the Bank Group's operations."

SEPTEMBER 2

- President Robert McNamara creates Operations Evaluation Unit within the Programming and Budgeting Department (PBD), under John H. Adler, Director, with responsibility for evaluating the contribution of the Bank's operations to the development of member countries.

  The unit numbers five professionals, and makes extensive use of consultants as well as receiving staff support from elsewhere in the Bank.

Initial, experimental work concentrates on a comprehensive evaluation of the Bank's 20-year-long development work in Colombia and a more summary evaluation of lending to seven important electric power borrowers in Latin America, Asia, and Africa.

1971

- The Operations Evaluation Unit is converted to the Operations Evaluation Division in the PBD.

  The Division has a small staff of six professionals (mostly new graduates from the Young Professionals Program) and several research assistants.

1972

- Two main reports, *Operations Evaluation Report: Electric Power* and *Bank Operations in Colombia—An Evaluation* are distributed to the Executive Directors (EDs) for review and discussion.

OCTOBER 20

- First project audit reports on highway, and power and telecommunications lending to Costa Rica are produced by the Operations Evaluation Division. The primary purpose of this 'series audit' is to determine, by brief analysis and by short discussion with those who have been involved, the extent to which the basic project objectives and loan covenants were fulfilled, reasons for significant deviations, and implications regarding the efficacy of Bank action. To the extent possible, lessons are sought as to how the lending activity might have been improved.

1973

- Project performance audits begin for all projects one year after completion of loan/credit disbursement (typically four to eight years after project approval by the Board).
- In-depth studies begin of selected portions of operations—for example, Development Finance Companies studies.

JULY 1

- The Operations Evaluation Department is established, along with the Internal Audit Department, under the supervision of Mohammad Shoaib, a vice president without operational responsibilities.
- Christopher R. Willoughby is appointed first director of OED. One chief evaluation officer and two senior evaluation officers dealing with the sectors of agriculture and industry, respectively, are appointed.

NOVEMBER

- OED is placed under the overview of JAC, which consists of six executive directors.

DECEMBER

- An amendment to the Foreign Assistance Act by the U.S. Congress encourages the establishment of an independent review and evaluation for the World Bank.

1974

- Key performance indicators—critical variables used to measure implementation progress and development—are first proposed in the 1974 Operational Memorandum on Project Supervision.
- A monitoring unit is created within the Agriculture and Rural Development Department (AGR). As the first Bank unit formed with a mandate to support monitoring and evaluation, it is responsible for producing global performance indicators on rural poverty and for assisting regional staff in institutional capacity building.
- OED issues the first Concordance of OED Reports.

JANUARY

- Management policy reviews begin, with the *Delays in Loan Effectiveness* and *Delays in Project Preparation* reports.

SEPTEMBER

- Establishment of the position of director-general, Operations Evaluation (DGO), is proposed by President Robert McNamara with a view to ensuring independent evaluation of projects and programs. Request is made that the new DGO with the rank of vice president "hold office for renewable terms of five years, be removable only by the executive directors and be ineligible for subsequent appointment or reappointment to the Bank Group staff except in unusual circumstances." The DGO is to be appointed by the Board from a short list of nominees proposed by the World Bank President.

OCTOBER

- The Board decides that OED will report directly to the Board with an administrative link to the president.

1975

- JAC establishes a subcommittee to review project evaluation reports and to assess the adequacy of the project evaluation system in the Bank.

SEPTEMBER 26

- First *Annual Review of Project Performance Audit Results* covering 50 projects is released.

OCTOBER 1

- Mervyn L. Weiner is appointed first DGO. Upon the appointment of the DGO, Vice President Shoaib relinquishes the responsibility of supervising OED.

1976

- New senior evaluation officer positions are established within OED.
- First *Annual Report on Operations Evaluation* is released.
- Completion reporting by operational staff becomes mandatory for all completed projects, with all being subject to audit by OED and then submitted by the DGO to the Board.

- More than 70 percent of OED resources are devoted to project auditing; about 25 percent are used for evaluation studies and policy reviews.

JUNE

- Shiv S. Kapur is appointed second director of OED.
- *Standards and Procedures for Operations Evaluation* are developed. These provide—for the first time—that the voice of operational as well as OED staff is to be protected in all OED reporting to the Board.

1977

- The Bank issues the first directive on Project Monitoring and Evaluation: *Operational Manual Statement (OMS) 3.55,* which recommends (but does not require) that all projects include some form of both monitoring and evaluation.
- OED begins to formally invite borrowers' comments on all draft completion reports on their projects, with the understanding that borrowers' comments, like comments of operational staff, will be included in all final reports to the Board.

JUNE

- General guidelines for preparing *Project Completion Reports* (*PCRs*) are developed by OED and operational staff.

1979

- Inclusion of key performance indicators in all Bank reports is mandated by *OMS 3.50.*
- AGR's Monitoring Unit is renamed the Monitoring and Evaluation Unit (AGRME). The unit begins disseminating "lessons learned" notes, based on PCRs and OED reports to all operational staff; it is the only central sectoral unit to undertake such feedback to its staff.
- OED begins to participate in Development Assistance Committee (DAC) of the Organisation for Economic Co-operation and Development (OECD) Evaluation Committee meetings, to exchange evaluation experience with participating DAC members and multilateral development banks.

JUNE

- A new series of impact evaluation studies is released. The primary purpose of the studies is to take a second look at completed projects on a selective basis and get a more complete view of their socio-economic impact.

AUGUST

- Second edition of *Standards and Procedures for Operations Evaluation* is released.

1980

- The DGO organizes Regional seminars in Nairobi and Kuala Lumpur to inform borrowers in these Regions about the Bank's evaluation system and invite them to participate in OED's activities in their countries with a view to establishing evaluation systems of their own.
- In pursuit of the same goal, the DGO is invited by the Comptroller General of the United States to attend a meeting of the International Association of Supreme Audit Institutions (INTOSAI) as an observer; from 1980 to 1982, the DGO visits selected auditors general and planning directors of member countries.

1981

- A.W. Clausen becomes sixth president of the World Bank.

1982

- OED establishes senior positions, to have three chief evaluation officers and six senior evaluation officers.

1983

APRIL

- Selective auditing system is introduced.

1984

- Operations Evaluation Unit is established within the Development Department of IFC.

JUNE

- Shiv S. Kapur is appointed second DGO.

AUGUST

- Yukinori Watanabe is appointed third director of OED.

1985

- Working Group on M&E is formed by AGRME with active staff participation from other sectors.

SEPTEMBER

- *Annual Review of Project Performance Audit Results* is replaced by *Annual Review of Project Performance Results.*

1986

- Barber B. Conable becomes seventh president of the World Bank.

JANUARY 27

- With *World Bank and Sri Lanka: Review of a Relationship 1960–1984*, OED launches a new series of reviews to evaluate the interactions between the Bank and selected borrowing countries at different levels and over extended periods of time.

MAY

- Yves Rovani is appointed third DGO.
- *Procedural Guidelines for OED Staff* is developed.

AUGUST

- Ram K. Chopra is appointed fourth director of OED.

1987

- OED Publications Committee is established with two primary purposes: to propose criteria for publication of OED material and to counsel the DGO regarding the annual publication of a few selected studies subject to the agreement of the JAC and/or the Board.

- Evaluation Capability Development (ECD) Program is created to assist client countries with central capacity building for ex-post development evaluation and to advise and support implementing agencies in preparing evaluation reports, including those of Bank-supported projects.

OCTOBER

- *Directory of OED Reports,* catalogued by sector, Region, and report number, is issued and disseminated to Bank staff.

DECEMBER

- A task force, chaired by Policy, Research and External Affairs (PRE), with the participation of OED and Operations, is formed to review means to enhance the effectiveness of internal dissemination and feedback processes in the reorganized Bank.

1988

- A three-year Technical Cooperation Program is set up between OED and the Canadian International Development Agency (CIDA). The program provides additional resources for OED to enhance the analytical depth of, and add new insights to, ongoing OED studies.

1989

- *Annual Review of Project Performance Results* is replaced by *Annual Review of Evaluation Results.*

- Two *OMSs* on key performance indicators (KPI) and monitoring and evaluation (M&E) are reissued as *Operational Directives (OD). OD 13.05* mandates that KPI be reintroduced into all supervision reports; *OD 10.70* mandates that M&E systems be included in all Bank-funded projects.

MAY

- *OD 13.60: Dissemination and Utilization of OED Findings* is issued. Reflecting the recommendations of the Task Force on Dissemination and Utilization of OED Findings, this directive mandates that Bank staff seek out OED findings, disseminate them to others concerned, and take the findings into account in future activities.

1990

MAY

- A revised version of *Procedural Guidelines for Operations Evaluation Staff* is produced.

JULY

- Hans-Eberhard Köpp is appointed fifth director of OED.
- First issue of *Précis* is released. OED *Précis* are intended to help inform Bank staff of new evaluation findings and recommendations for the Bank's current and future work programs.

AUGUST

- OED develops, in collaboration with the Information, Technology, and Facilities Department, the Evaluation Text Base (ETB), which contains summaries of all the Bank's ex-post evaluation reports, along with basic information on all completed operations in a computerized system. ETB, an innovative text base with bibliographic and text search capacities, is intended to make OED's work more service-oriented by facilitating easier access to lessons of evaluation for Bank staff members.

1991

- Lewis T. Preston becomes eighth president of the World Bank.
- OED's Annual Review Database is expanded and now contains information on the operational performance of all completed lending operations, including completion reports, OED audit reports, and performance ratings by OED.

1992

MAY

- *The Future of Evaluation in the World Bank: Mandate, Mission, and Resources* is prepared by Yves Rovani, DGO, for discussion at the request of JAC.

SEPTEMBER 1

- Robert Picciotto is appointed fourth DGO.

SEPTEMBER 22

- *Effective Implementation: Key to Development Impact* (the "Wapenhans Report") is produced by the Bank's Portfolio Management Task Force (PMTF). As one of its six principal recommendations to improve the Bank's portfolio performance management, PMTF proposes that OED enhance its role of independent accountability and focus on sustainable development impact. The report was prepared in response to OED's annual review of the previous year, which showed a worsening decline in portfolio performance and, implicitly, in the Bank's portfolio management.

NOVEMBER 3

- The Wapenhans Report is transmitted to the Board.

1993

- The Bank-wide Task Force on Evaluation Capacity Development (ECD) is established. The goals of the Task Force are to determine the priority for ECD in public sector management, review past experience of ECD, assess Bank efforts in relation to ECD, determine possible country strategies for successful ECD, and define generic Bank strategies for supporting ECD.

APRIL 13

- *Report on Portfolio Management and the Evaluation Function* is produced by JAC.

JULY 8

- *Next Steps—A Program of Actions* is approved by the Board, based on management's proposals on how to implement the recommendations of PMTF.

AUGUST 26

- The Bank's disclosure policy is endorsed by the Board; the policy subsequently (in 1994) provides for disclosure of OED *Précis,* OED's *Annual Review of Evaluation Results,* and selected OED evaluation studies.

SEPTEMBER 1

- First issue of *Lessons and Practices* is released. Drawn from completed Bank and external evaluations, the series provides recommendations and good practice guidelines for development practitioners, synthesizing what has and has not worked from past development operations.

SEPTEMBER 22

- An independent Inspection Panel is established by the Board to ensure that Bank operations adhere to the institution's operational policies and procedures. The panel is to consist of three members of different nationalities from Bank member countries, who will be nominated by the president and appointed by the Board. The nominees must not be World Bank staff members. The panel will review and investigate requests made by affected parties or an ED, and will report its findings to the Board.

1994

- OED launches its Web site.

FEBRUARY

- A Bank-wide task force is established with the participation of OED to review and assess the dissemination of lessons learned from experience within the Bank and outside.

APRIL

- The initial three Inspection Panel members (one full-time chairman and two part-time members) are appointed by the Board.

MAY 16

- *OED Review* is produced by Organization and Business Practices Department (OBP).

JUNE 30

- *An Overview of Monitoring and Evaluation in the World Bank* is released by OED with focus on the 20-year history of involvement with the establishment and use of project M&E systems.

JULY

- *Operational Policy (OP) and Bank Procedure (BP) 13.55: Guidelines for Implementation Completion Reporting* become effective and replace the *OP, BP, and Good Practice (GP) 13.55* of April 1994. Under the new guidelines, *ICRs* will be transmitted directly to the Board by the Regions. OED will prepare an evaluative note on each *ICR* assessing its quality and circulate it to regional management for comment and feedback before producing the final version.

JULY 5

- *Report of the Task Force on Evaluation Capacity Development (ECD)* is produced.

SEPTEMBER

- Committee on Development Effectiveness (CODE) is established as a standing committee of the Board.
- The committee reviews the work of the OED, management's responses thereon, and selected reports from management to satisfy the Bank Group's operations evaluation and determine whether self-evaluation activities are adequate and efficient.

SEPTEMBER 7

- Office of the Inspection Panel is opened.

NOVEMBER 1

- *Progress Report on the Implementation of Portfolio Management: Next Step—A Program of Actions* is released by the Bank. The report highlights and summarizes the progress and changes that have been made by the Bank in the area of recommended actions proposed in *Portfolio Management: Next Steps—A Program of Actions,* 1993.

NOVEMBER 7

- Revisions are made to the Bank's disclosure policy. The revised policy provides that evaluation documents be disclosed as presented to the Board and that publications based on evaluation studies incorporate management's responses to the studies and summaries of the JAC and/or Board discussions.

DECEMBER

- Ruth Jacoby becomes first chair of CODE.

DECEMBER 5–6

- The Bank's First Biennial Conference on Evaluation and Development is held.

1995

- James D. Wolfensohn becomes ninth president of the World Bank.

MARCH

- Francisco Aguirre-Sacasa is appointed sixth director of OED.

JUNE 1

- The first in a series of new studies (*Ghana: Country Assistance Review [CAR]*) is released. These studies seek to evaluate the relevance and efficacy of the Bank's overall country assistance strategy as well as the effectiveness of various Bank lending and nonlending instruments.

AUGUST

- Informal Sub-committee of CODE is established. The Sub-committee, among other things, examines the adequacy of a selected number of Project Audit Reports (PARs) submitted by OED to the Board and considers whether adequate follow-up action has been taken by OED and/or Management on recommendations made by the Sub-committee and endorsed by CODE.

OCTOBER

- In line with the revision of the operations evaluation mandate at the IFC, the Corporation's Operations Evaluation Unit is reorganized as a department, the Operations Evaluation Group (OEG), to provide for increased attention to the development impact of IFC operations. OEG transmits its report to the Board through the DGO. William Stevenson is appointed first director of OEG.

NOVEMBER

- Upon EDs' request, OED conducts a review of evaluation in the Economic Development Institute (EDI). The purpose of the review is to establish an evaluation mechanism in accordance with EDI's expanding role.

DECEMBER 29

- *Monitoring and Evaluation Plans in Staff Appraisal Reports issued in Fiscal Year 1995* is released by OED with a view to follow up on the findings from the OED's report, *An Overview of Monitoring and Evaluation in the World Bank*, issued in 1994.

1996

- OED's e-line (eline@worldbank.org) is initiated as part of OED's dissemination efforts.

FEBRUARY

- Evaluation Cooperation Group (ECG) is established by the heads of evaluation in the multilateral development banks. The purpose is to strengthen cooperation among evaluators and harmonize evaluation methodology in its member institutions: the World Bank Group, Inter-American Development Bank, Asian Development Bank, African Development Bank, European Investment Bank, and European Bank for Reconstruction and Development.
- Quality Assurance Group (QAG) is established as part of the Bank's renewal effort aimed at improving decentralized decisionmaking, better accountability, and a faster learning loop. QAG's evaluation of the Bank's performance is ex ante and does not duplicate OED's work.

1997

- OED reduces the ratio of completed projects audited in a given fiscal year from 40 percent to 25 percent.
- OED introduces *Country Assistance Notes (CANs)*—smaller in scope and produced more quickly than full *CARs*.
- OED introduces country sector impact studies to examine the long-term impact of Bank activities in particular sectors.

MARCH 3

- A partnership agreement is signed between OED and the Swiss Agency for Development and Cooperation (SDC). The purpose is to support evaluation activities related to making poverty alleviation more effective; strengthening evaluation capacity; eliciting participation, commitment, ownership, and empowerment of Bank clients and partners in developing countries; and strengthening their evaluation capacities.

APRIL 1–2

- The Bank's Second Biennial Conference on Evaluation and Development—the Institutional Dimension—is held.

MAY

- Elizabeth J. McAllister is appointed seventh director of OED.

SEPTEMBER

- Surendra Singh becomes second chair of CODE.

NOVEMBER 24

- *Annual Review of Evaluation Results* is replaced as *Annual Review of Development Effectiveness (ARDE).*

NOVEMBER 6

- OED releases the first *Fast Track Brief (FTB),* to provide timely and relevant evaluation findings on Bank operations to the president, Board, and senior management, before the reports are discussed by the Board.

DECEMBER

- Introduction to Program Evaluation (IPE) course starts as a classroom course for OED staff and evolves into a distance-learning course for Bank staff and clients involved in planning and implementing project and program evaluations.

1998

- The Oversight Committee consisting of senior staff from the Legal Department, Internal Audit, the Office of Professional Ethics, and Operational Core Services is created.
- Program evaluation workshops (PEWs) offering training in core M&E skills are initiated through the World Bank Institute. They focus on building local institutional capacity through training of trainers.

SEPTEMBER 30

- A partnership agreement is signed between OED and the Royal Ministry of Foreign Affairs of Norway. The objective is to strengthen the evaluation capacity of clients in developing countries (including NGOs) and that of Norway and the Bank.

OCTOBER

- OED Help Desk is created with the launch of OED's Knowledge Management Program.

NOVEMBER

- Jan Piercy becomes third chair of CODE.

NOVEMBER 16–19

- The Regional Seminar and Workshop on Monitoring and Evaluation Capacity Development in Africa is held in Abidjan. OED participates as a sponsor and co-publisher of the proceedings.

1999

- The Evaluation and Control Working Group (ECWG) is established to address "a lack of knowledge among staff members regarding the Bank's mandates and operational policies of evaluation and control functions, as well as frustration with the administrative aspects of responding to evaluations."
- *CAR* and *CAN* are replaced by *Country Assistance Evaluation (CAE).*

JUNE

- The Bank's Third Biennial Conference on Evaluation and Development—Evaluation and Poverty Reduction—is held.
- The first issue of ECG newsletter—*Evaluation Insights*—is released.

JULY

- New Guidelines for *ICR* preparation replace *OP/BP/GP 13.55* of 1994. As a result of the Bank-wide *ICR* Process Working Group of 1998, the changes are aimed mainly at strengthening the lesson-learning objectives of *ICRs* and streamlining the criteria, definition, and rating systems.

DECEMBER

- M&E working group is established to develop a comprehensive action plan for achieving sustainable improvements in M&E in Bank-financed projects.

2000

- Development Gateway is launched. OED participates in the Aid Effectiveness Web site.

MAY

- Gregory K. Ingram is appointed eighth director of OED.

JULY 13

- Partnership agreement is signed between OED and the Policy and Operations Evaluation Department (IOB) of the Ministry of Foreign Affairs (MFA) of the Netherlands. The purpose is to strengthen cooperation between the two departments in evaluation, with special emphasis on evaluation capacity development.

AUGUST

- Joint Task Force on Country Assistance Methodology is established by OED and the Bank's Operations Policy and Country Services (OPCS).

NOVEMBER

- Pieter Stek becomes fourth chair of CODE.

2001

- *Project Performance Audit Report* is replaced by *Project Performance Assessment Report.*

MAY

- London Declaration on the establishment of the International Development Evaluation Association (IDEAS) is adopted at a meeting sponsored by the United Nations Development Programme (UNDP) Evaluation Office and OED in partnership with the Department for International Development, United Kingdom. Anchored in the needs of evaluators in developing countries and transition economies, IDEAS is an initiative to form a global professional network with an emphasis on ECD.

JUNE-JULY

- First International Program for Development Evaluation Training (IPDET) is held in Ottawa under the co-sponsorship of OED and the Faculty of Public Affairs and Management at Carleton University. This four-week residential program is designed to meet the needs of evaluation and audit units of bilateral and multilateral development agencies and banks, governments, and evaluators working in development and nongovernmental organizations.

JULY 23–24

- The Bank's Fourth Biennial Conference on Evaluation and Development—the Partnership Dimension—is held.

AUGUST

- The Board endorses a newly revised disclosure policy effective January 2002. The updated policy provides for disclosure of the OED's performance assessment reports and process evaluations, as well as the management responses to these evaluations. Synthesis reports prepared by QAG on the quality of key activities (such as loan preparation, loan supervision, and economic and sector work) are also to be made publicly available.

OCTOBER

- OED establishes, for the first time, a multi-donor trust fund for carrying out a joint evaluation of the Comprehensive Development Framework (CDF) with eight development partners.

2002

JANUARY

- Multilateral Investment Guarantee Agency (MIGA)'s evaluation office is established. The office transmits its reports to the Board through the DGO. Aysegul Akin-Karasapan is appointed first director.

AUGUST

- First issue of the OED electronic newsletter—*e-Alert*—is released. As part of OED's dissemination efforts, *OED e-Alert* sends out news and announcements of OED's new products and activities on a monthly basis.

SEPTEMBER

- IDEAS is launched in Beijing, drawing development evaluators from more than 40 countries.

OCTOBER

- Gregory K. Ingram is appointed fifth DGO.

NOVEMBER

- Finn Jonck becomes fifth chair of CODE.

OED PUBLICATIONS

Study Series

2002 Annual Review of Development Effectiveness—Achieving Development Outcomes: The Millennium Challenge
Agricultural Extension: The Kenya Experience
Agricultural Extension and Research: Achievements and Problems in National Systems
Bangladesh: Progress Through Partnership
Bridging Troubled Waters: Assessing the World Bank Water Resources Strategy
Debt Relief for the Poorest: An OED Review of the HIPC Initiative
Developing Towns and Cities: Lessons from Brazil and the Philippines
The Drive to Partnership: Aid Coordination and the World Bank
Financial Sector Reform: A Review of World Bank Assistance
Financing the Global Benefits of Forests: The Bank's GEF Portfolio and the 1991 Forest Strategy and Its Implementation
Fiscal Management in Adjustment Lending
IDA's Partnership for Poverty Reduction
India: The Challenges of Development
India: The Dairy Revolution
Information Infrastructure: The World Bank Group's Experience
Investing in Health: Development Effectiveness in the Health, Nutrition, and Population Sector
Lesotho: Development in a Challenging Environment
Mainstreaming Gender in World Bank Lending: An Update
The Next Ascent: An Evaluation of the Aga Khan Rural Support Program, Pakistan
Nongovernmental Organizations in World Bank–Supported Projects: A Review
Paddy Irrigation and Water Management in Southeast Asia
Poland Country Assistance Review: Partnership in a Transition Economy
Poverty Reduction in the 1990s: An Evaluation of Strategy and Performance
Promoting Environmental Sustainability in Development
Reforming Agriculture: The World Bank Goes to Market
Social Funds: Assessing Effectiveness
Uganda: Policy, Participation, People
The World Bank's Experience with Post-Conflict Reconstruction
The World Bank's Forest Strategy: Striking the Right Balance
Zambia Country Assistance Review: Turning an Economy Around

Evaluation Country Case Series

Bosnia and Herzegovina: Post-Conflict Reconstruction
Brazil: Forests in the Balance: Challenges of Conservation with Development
Cameroon: Forest Sector Development in a Difficult Political Economy
China: From Afforestation to Poverty Alleviation and Natural Forest Management
Costa Rica: Forest Strategy and the Evolution of Land Use
El Salvador: Post-Conflict Reconstruction
India: Alleviating Poverty through Forest Development
Indonesia: The Challenges of World Bank Involvement in Forests
Uganda: Post-Conflict Reconstruction

Proceedings

Global Public Policies and Programs: Implications for Financing and Evaluation
Lessons of Fiscal Adjustment
Lesson from Urban Transport
Evaluating the Gender Impact of World Bank Assistance
Evaluation and Development: The Institutional Dimension (Transaction Publishers)
Evaluation and Poverty Reduction
Monitoring & Evaluation Capacity Development in Africa
Public Sector Performance—The Critical Role of Evaluation

Multilingual Editions

Allègement de la dette pour les plus pauvres : Examen OED de l'initiative PPTE
Appréciation de l'efficacité du développement :
L'évaluation à la Banque mondiale et à la Société financière internationale
Determinar la eficacia de las actividades de desarrollo :
La evaluación en el Banco Mundial y la Corporación Financiera Internacional
Côte d'Ivoire : Revue de l'aide de la Banque mondiale au pays
Filipinas: Crisis y oportunidades
Reconstruir a Economia de Moçambique

http://www.worldbank.org/oed

WORLD BANK OPERATIONS EVALUATION DEPARTMENT: THE FIRST 30 YEARS

THE WORLD BANK

THE WORLD BANK

ISBN 0-8213-5549-X
E-ISBN 0-8213-5550-3