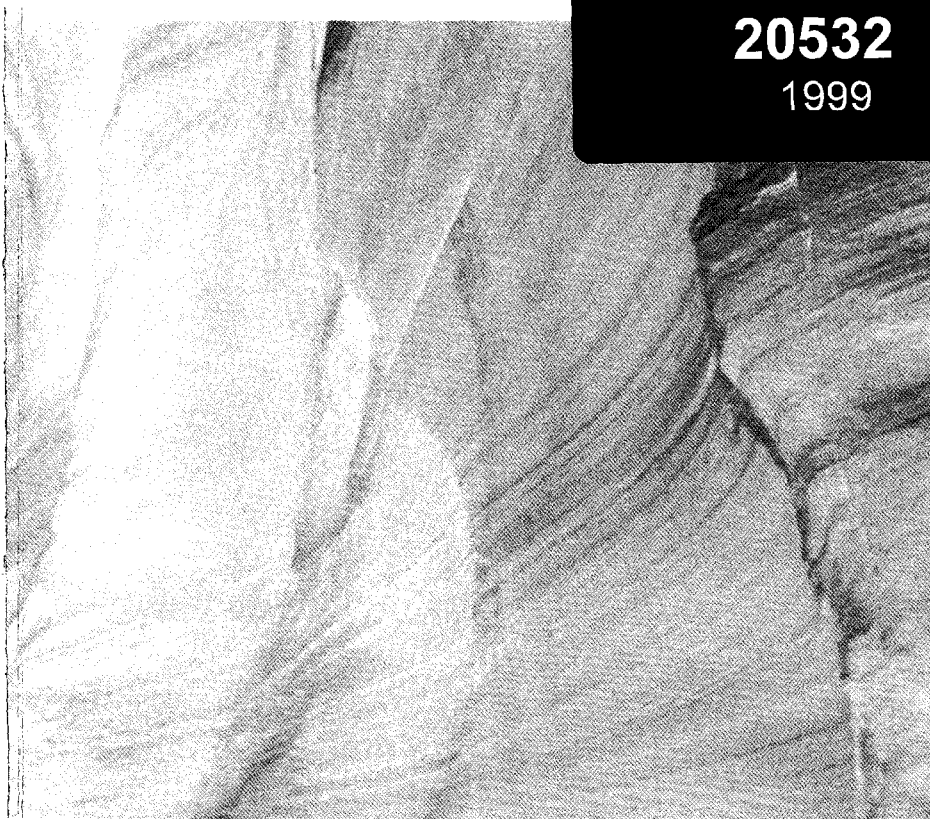




Public Sector Performance— The Critical Role of Evaluation

Selected Proceedings from
a World Bank Seminar

20532
1999



WORLD BANK OPERATIONS EVALUATION DEPARTMENT
EVALUATION CAPACITY DEVELOPMENT



OED
ECD

Public Sector Performance— The Critical Role of Evaluation

Selected Proceedings from
a World Bank Seminar

Keith Mackay
Editor

The World Bank
Washington, D.C.



This Proceedings is part of the OED Proceedings Series produced by the Operations Evaluation Department, Partnerships and Knowledge (OEDPK) group of the World Bank. The views expressed here should not be attributed to the World Bank or its affiliated organizations.

Contents

v	Abbreviations/Acronyms	
vii	Acknowledgments	
ix	Public Sector Performance—The Critical Role of Evaluation	Keith Mackay
1	PART 1: The Role of Evaluation in Development	
3	Why Bother About ECD?	Robert Picciotto
7	The Role of Evaluation	Mark Baird
13	PART 2: The Missing Link in Good Governance	
15	Evaluation Capacity and the Public Sector	David Shand
19	PART 3: Experience of Developed Countries	
21	The Development of Australia's Evaluation System	Keith Mackay
55	Evaluation in the Federal Government of Canada	Stan Divorski
59	Comments	Ray C. Rist
62	Comments	Frans Leeuw
65	PART 4: Experience of Developing Countries	
67	Lessons from Chile	Mario Marcel
72	Indonesia's National Evaluation System	Alain Barbarie
85	Evaluation Capacity Development in Zimbabwe	Stephen Brushett
94	Comments	Eduardo Wiesner
97	PART 5: Two Perspectives	
99	A View from a World Bank Network	Cheryl Gray
101	A View from USAID	Gerald Britan
105	PART 6: Where Do We Go from Here?	
107	Overview and Conclusions	Elizabeth McAllister
112	Recommended Reading	
113	List of Authors and Discussants	

Abbreviations and Acronyms

Bappenas	National Development Planning Agency, Government of Indonesia
CAR	Country Assistance Review
CAS	Country Assistance Strategy
CDIE	Center for Development Information and Evaluation (USAID)
CODE	Committee on Development Effectiveness (World Bank)
DAC	Development Assistance Committee (OECD)
DEC	Development Economics&Chief Economist Vice Presidency (World Bank)
DoF	Department of Finance (Australia)
ECD	Evaluation Capacity Development
EDI	Economic Development Institute (World Bank)
ESW	Economic and Sector Work
GOI	Government of Indonesia
GPRA	Government Performance and Results Act (US government)
IBRD	International Bank for Reconstruction and Development
IDRC	International Development Research Centre (Canada)
LLC	Learning and Leadership Center (World Bank)
M&E	Monitoring and evaluation
NGO	Nongovernment organization
OECD	Organisation for Economic Cooperation and Development
OED	Operations Evaluation Department
PERL	Public Expenditure Reform Loan
PREM	Poverty Reduction and Economic Management Network (World Bank)
QAG	Quality Assurance Group (World Bank)
Quango	Quasi-autonomous nongovernment organization
RBM	Results-based management
SDC	Swiss Agency for Development and Cooperation
TA	Technical assistance
USAID	US Agency for International Development
WB	World Bank

Acknowledgments

The joint OED-LLC seminar was organized under the auspices of LLC, PREM and OED. The seminar team consisted of Keith Mackay (task manager, OEDPK), Kathy Peterson, and Erika Marquina (LLC), with additional support from Patty Rodriguez (OEDPK).

This proceeding was produced in the Partnerships and Knowledge Group (OEDPK) by the Dissemination and Outreach Unit. The unit is directed by Elizabeth Campbell-Pagé, Task Manager, and includes Caroline McEuen (editor), Kathy Strauss and Lunn Lestina (desktop design and layout), and Juicy Qureishi-Huq (administrative assistance).

Director-General, Operations Evaluation Department: *Robert Picciotto*

Director, Operations Evaluation Department: *Elizabeth McAllister*

Manager, Partnerships & Knowledge Programs: *Oswaldo Feinstein*

Task Manager: *Keith Mackay*

Public Sector Performance— The Critical Role of Evaluation

Selected proceedings from a World Bank seminar

Editor's Preface

The development community and the World Bank recognize the importance of sound governance and the need to support countries' capacity building and institutional development. To encourage debate and explore the links between public sector performance and the role of evaluation, the World Bank organized a seminar focusing on the experience in developing national evaluation systems, or *evaluation capacity development (ECD)*, as an aid to better governance, in Washington DC in April 1998.

The purpose of this publication is to present selected papers plus the transcripts of key seminar proceedings. This selection represents only a subset of the broad and rich range of issues covered. A main focus of this volume is national experiences in developing evaluation systems. By preserving these experiences, OED intends to contribute to the growing "library" of evaluation case studies—in other words to document lessons concerning what worked, what did not, and why. These lessons will be shared with World Bank staff, other development agencies, and developing country governments. To further support this objective, the World Bank is also launching a separate series of short ECD papers, the main focus of which is country case studies.

The Priority for Developing Evaluation Capacity

Evaluation potential can be better understood by recognizing the importance of economic governance and a sound public sector to national economic competitiveness—markets reward countries able to manage and screen public expenditures, and evaluation offers a tool to help do that.

Robert Picciotto argues that evaluation is in many ways central to the effectiveness of development assistance:

- that the development assistance community has turned to results-based management at the project, country and global levels and that this approach requires solid institutional capacity in countries;

-
- that partnerships and coalitions among development agencies to help support country programs and institutions also require a common approach to evaluation and assessment; and
 - that our growing need to demonstrate the effectiveness of development interventions to the electorates of industrial democracies is evident.

Developing national evaluation capacities is a means for ensuring that evaluation findings are available to assist countries in three areas. First, evaluation findings can be an important input for government decisionmaking and prioritization, particularly in the budget process. Second, evaluation assists managers by revealing the performance of ongoing activities at the project, program or sector levels—it is therefore a management tool which leads to learning and improvement in the future (i.e., results-based management). Similarly, evaluation results can also be used to assess the performance of organizations and institutional reform processes. Third, evaluation data contribute to accountability mechanisms, whereby managers and governments can be held accountable for the performance of their activities. As David Shand explains, there may well exist trade-offs between these three uses of evaluation findings.

The concept of performance encompasses the efficiency of a project or activity—the ability to undertake an activity at the minimum cost possible. It also includes effectiveness—whether the objectives set for the activity are being achieved.

There are many types of evaluation tools, which can be used in a variety of ways. These tools are related—in that they deal with the concept of performance—but they can lead to confusion, exacerbated by the different terminology employed by evaluation practitioners. Regardless of the terms used: ongoing monitoring and performance information; project and program evaluation—*ex ante*, ongoing/formative and *ex post*/summative; performance (or value-for-money) audits; financial auditing, they all address performance measurement. This broad spectrum of performance measurement activities is also known by other generic labels, such as monitoring and evaluation (M&E). Unless otherwise stated, the term ‘evaluation’ is used in this volume.

Experts from different backgrounds—economists (who have traditionally focused on project evaluation), program evaluators (who typically have a broad social science background) and auditors (who in the past have emphasized financial compliance)—tend to use different concepts but often with similar or even identical nomenclature. It is little wonder that those new to evaluation are often confused by the panoply of concepts and jargon. This volume does not address the problem of terminology in any depth, but the reader is alerted to its existence and is advised to carefully interpret the messages on which these papers focus.

Performance measurement is a valuable exercise not least because it provides an opportunity and a framework for asking fundamental questions such as: What are you trying to achieve?; What does “success” look like?; How will you know if or when you’ve achieved it? As Mark Baird emphasizes in his paper,

No public sector can afford to overlook the importance of clearly defining its objectives and priorities, assessing performance against well-defined benchmarks, and changing the bureaucratic culture into one that stresses client service and achievement of results . . . Rather than an imposed requirement of donor agencies, evaluation now becomes a key instrument of good governance and institutional development within our client countries. We all have a responsibility to make sure this function is nurtured and supported, as it has been within our own institutions.

The support that the development of evaluation capacity offers to broader governance, institutional development and public sector reform is often not fully appreciated. Links and commonalities are seen in the areas of:

- budgetary financial management systems, which includes financial reporting and auditing;
- intergovernmental fiscal relations, and the extent to which they encompass a focus on performance;
- commercialization and the private sector delivery of public services. For the private sector to be successful in the delivery of public services, governments should have a clear understanding of program objectives, and they should undertake ex ante, ongoing and ex post assessments of performance;
- formulation of customer service standards by service delivery agencies, and monitoring the extent to which these standards are actually achieved;
- civil service reform, including personnel performance, management and appraisal—recognizing that individual performance is reflected, to some extent, in project or program performance;
- civil service policy advice, which should encompass existing evaluation findings or seek to commission new ones;
- participation and the ‘voice’ of civil society, which incorporates the views and expectations of ordinary citizens concerning the performance of government activities; and
- anti-corruption efforts, particularly to improve financial management systems and performance reporting, strengthen watchdog agencies, and achieve greater transparency in policymaking and implementation.

Country Experience

Developed and developing countries alike are accumulating a growing volume of experience with national evaluation systems. These proceedings present experiences in Chile, Indonesia, Zimbabwe, Australia and Canada.

The experience of developed countries illustrates the potential links between national evaluation capacity and good governance, reflecting the opportunities and difficulties in achieving cultural change in a government—winning hearts and minds is a slow business. It also underscores the different dimensions that must be developed to achieve a robust national evaluation system in the areas of demand, supply, and information infrastructure.

The main *precondition* for developing a national evaluation system is country demand—an evaluation system cannot be effectively foisted on an unwilling government. There are particular risks if the impetus for an evaluation system is donor-driven; this is not to say that donors cannot take the lead in “selling” the merits of evaluation systems to countries, but rather, that unless and until countries accept the strength of such arguments, or reach their own conclusions about the merits of evaluation, an evaluation system is unlikely to be sustainable.

As Stephen Brushett notes in his paper, two building blocks for effective demand include sensitizing key stakeholders to the need for and benefits from evaluation, and building awareness of suitable techniques and approaches.

Experience tells us that the main barriers to developing evaluation systems in developing countries are: poor demand and ownership in countries; lack of a culture of accountability (often related to ethics or corruption); absence of evaluation, accounting, or auditing skills—there is a need to develop the supply of these skills and systems to match demand as and when it grows; poor quality of financial and other performance information, and of accounting/auditing standards and systems; lack of evaluation feedback mechanisms into decisionmaking processes; and the need for greater efforts to develop evaluation systems’ capacity for sustainability. (These issues are discussed in The World Bank, *Evaluation Capacity Development*, Report of the Task Force, Washington DC, 1994.)

What is not yet clear is whether a government or ministry needs some minimum level of overall capability before an evaluation system can realistically be contemplated. But we do know that developing an evaluation system should not be viewed as a stand-alone activity—it

would be unrealistic to attempt to simply “bolt on” an evaluation system to an existing structure of governance if the institutional framework and incentives do not support it. On the other hand, if the framework and incentives are insufficient, then this is a strong argument for ensuring that the development of an evaluation system is part of a broader initiative to develop governance. This approach recognizes the strong synergies between performance measurement/evaluation and performance management.

Success Factors

Experience suggests that a number of success factors (discussed in greater depth in the country papers) contribute to the development of an evaluation system—developing a system should be pursued only if many of these already exist or if there are reasonable prospects for creating them. One example of a success factor is the role played by a ‘champion’ ministry or agency in supporting, encouraging and pushing the development of an evaluation system. Best results are seen with powerful and influential lead agencies, such as finance or planning ministries (as in Australia and Canada), or a national audit office. Indonesia found it valuable to have the support of such central agencies, as well as a network of committed supporters in ministries. An explicit and high-profile evaluation strategy can also be effective in selling the message, as can the support of individual ministers or the Cabinet as a whole—particularly via ministerial or presidential decrees.

Sustained government commitment is also important. An evaluation system cannot be developed overnight; indeed experience indicates that it can take at least a decade at the whole-of-government level to embed such a system in a sustainable manner, to develop necessary skills, and set up civil service structures, systems and “ownership” to make full use of evaluation findings. The counterpart of the need for sustained government support is the need for sustained support from development assistance agencies—Alain Barbarie emphasizes the value of ongoing, active and visible support of the World Bank in Indonesia.

A whole-of-government approach (e.g., Chile, Indonesia, Canada, Australia) has advantages in terms of achieving momentum and helping to ensure that laggard ministries endeavor to keep up with leading ministries. This approach might be especially feasible if a major series of reforms is being contemplated, such as major changes to public expenditure management (i.e. budgetary processes and decisionmaking) in response to budgetary imperatives.

But a whole-of-government approach might be unrealistic in a number of developing countries—it can be difficult to achieve. In these countries a more modest approach

might be to start with an initial focus on ongoing performance monitoring (in particular sectors or ministries in order to create a demonstration effect), and then seek to apply the approach to other sectors/ministries and to other performance measurement tools (such as evaluation) as opportunities are found or created.

This type of sequencing implicitly depends on perceptions of satisfactory benefits vis-à-vis the costs of the initial, more modest approach. Chile took such an incremental approach, following an initial focus on performance information, which led to questions about program outcomes and causality. These types of questions can only be answered by summative program evaluations.

Perhaps the key message here is the need to tailor approaches for developing evaluation capacity to suit the circumstances and opportunities in different countries. It is definitely not the case that “one-size-fits-all.”

The final lesson is that incentives are crucial to ensuring both that an evaluation system is developed and that evaluation findings are actually used. In working with national governments it is important for those providing technical assistance to understand the incentive frameworks in the country. This involves conducting an institutional diagnosis as a precursor to provision of advice, while also ensuring a close dialogue with the government (such an approach was followed in Zimbabwe). This diagnosis should also examine the extent of evaluation capacity in a government—it is often the case that governments over estimate the extent and quality of their evaluation activity.

Eduardo Wiesner emphasizes that existing incentives often work against evaluation, and that it is important to identify who stands to lose from having information available on the performance of government activities. Line ministries may perceive evaluation findings as a threat because of fears that finance or planning ministries will use the findings to reduce budget appropriations or demand accountability. This underscores the trade-off between central ministries using findings to assist budget allocations or for accountability, versus managers using the findings to improve performance.

Who should be responsible for measuring performance? Impartial outsiders, or expert insiders? The former approach stresses objectivity and independence, while the latter stresses expert knowledge and ownership of the evaluation results, which in turn, is likely to encourage learning by managers and their staff.

One answer could be to emphasize *learning* where civil service performance and

capabilities are believed to be reasonably good and the institutional environment is judged to be conducive to improvement. The other option is to emphasize *accountability* especially where performance is very poor or corruption is a problem.

In the latter case there are opportunities for measuring performance to expose poor performance of the civil service and thus to increase the pressures for a more responsive public sector. This was carried out in Bangalore, India, for example, where an NGO conducted surveys of citizen views concerning the quality of government services and the extent of corruption in delivery.

There is clearly scope for civil society (including NGOs) to play a role in performance measurement. There may also be opportunities for bilateral donors to promote and support such activities.

Implications for the World Bank and Other Donors

The development of evaluation capacity is a long-term proposition, not amenable to a simple, two- or three-year project approach. Thus sustained support and commitment, and the active involvement of the country government is necessary. Development agencies should view the provision of technical assistance as part of a sustained, ongoing partnership or dialogue with the government. The role of donors is important—donors can help or hinder the development of evaluation capacity.

Donors can help by providing technical assistance—their advice and sharing of lessons and best practice can be invaluable, as can providing funds for training and building evaluation systems. Their participation helps to build confidence within the government.

Donors can impede the development of evaluation capacity by excessive, conflicting or multiple donor requirements for evaluation. There is a particular danger that scarce country evaluation capacity will be diverted to satisfy donor requirements for the evaluation of development assistance activities, which might not necessarily align with the greatest areas of benefit from evaluation in the government.

New lending instruments, which provide multi-year funding for technical assistance activities, are likely to be better tailored to meeting country needs for long-term commitment and support in this area, as Cheryl Gray emphasizes.

Collectively, donors have a considerable and growing track record in supporting such

activities. The World Bank, for example, implemented a program to support evaluation capacity development in 1987. Other multilateral donors (such as the UNDP and the Asian Development Bank) and some bilateral donors, have also supported evaluation capacity development.

There is considerable scope for closer donor collaboration and partnership. This could take place via: sharing existing case studies (which serve as invaluable learning tools); identifying and analyzing countries with good or promising practices; developing approaches for undertaking country diagnoses; and supporting regional collaboration.

It is heartening that the seminar on which this volume is based has already led to a round-table meeting of a number of donors to discuss options and modalities for greater cooperation and collaboration. More effort to pursue these opportunities is needed, and this will provide a litmus test of the extent of commitment to this important aspect of governance.

Keith Mackay
The World Bank

PART 1:
The Role of Evaluation
in Development



Why Bother About ECD?

Robert Picciotto

ECD and Development

We are in a business which has been under stress for several years. It has been further shaken by East Asia's financial crisis. Although the dust has not settled yet, three lessons are already clear.

First, in an increasingly integrated global economy, policy convergence and sound macro fundamentals are not enough. Institutions matter quite as much. And fickle markets are wont to punish countries where reform-induced capital flows have put excessive demands on domestic institutions.

Second, the quality of public expenditures is as important as fiscal balance. Large-scale, low-return, investment projects involving the state as a source of funds or of sovereign guarantees can aggravate situations of financial stringency.

Third, generic issues of economic governance tend to dominate market perceptions well after the crisis has hit. And these issues are far more difficult to tackle in the short run than issues of macroeconomic policy or institutional reform.

Development assistance is no longer simply a matter of funding sound projects. It is not even good enough to ensure sound macro fundamentals through adjustment lending. Institutional development has come center stage. Conditionality has moved from the IFIs to the markets. A sound public sector is increasingly recognized as a vital factor of national economic competitiveness. Markets reward countries able to manage public expenditures with care at all stages of the business cycle and countries able to establish processes which screen out dubious schemes before they are undertaken.

Thus, evaluation capacity development is an integral part of development and this is why the principles which underlie the 1994 ECD strategy are still largely valid. It is clear that ECD strategies and tactics must take account of country conditions (of demand and supply factors as highlighted in the Lessons and Practices document). But in addition, ECD design must take account of how truth can best challenge power—given political economy and governance factors. Evaluation may be connected to the executive or the legislature or both. It may be centralized or decentralized. It may or may not rely on the intellectual resources of the civil society. Increasingly, it is intergovernmental and participatory.

ECD and Development Assistance

Beyond the World Bank, it is increasingly clear that the fortunes of development assistance hinge to a very great extent on the quality of evaluation capacity in developing countries. Here are three of the major reasons for this.

First, the entire development assistance community has turned to results-based management at project, country and global levels. And for such an approach to work effectively, capacity to collect, verify, assess and use performance information is essential. Experience among all development assistance agencies converges on one lesson: without institutional capacity on the ground, monitoring and evaluation components of projects are mere paper exercises, of no value whatsoever.

Second, at the country level, the demand for policy and program evaluation capacities and skills is rising rapidly as country assistance strategies become participatory and involve an increasingly wide range of partners. The time is past when development assistance agencies could neatly divide their interventions so that they did not interfere with one another.

As the development consensus on priorities solidifies, and as efforts to move development assistance to the higher plane of country programs and institutions become successful, development coalitions are emerging. In this context, it is increasingly difficult to ensure coherence in development interventions through headquarters-based coordination mechanisms. The country is now at the center of the development partnership and evaluation needs to be located as close to the ground as possible.

Third, the need to make development assistance more understandable to the electorates of the industrial democracies hinges on getting credible evidence in support of the effectiveness of development interventions. In this connection, the Shaping the 21st Century Initiative of the Development Assistance Committee needs the active support of all bilateral donors and multilateral agencies. And evaluation needs to be an explicit part of this initiative to deliver credible information about results achieved. Here again, the defining constraint lies in domestic evaluation capacity.

So evaluation capacity development means more than imparting evaluation skills. It means fitting evaluation structures, systems and processes to new public sector reform strategies. In turn, this requires a clear vision of where country policy priorities lie and of the role of development assistance in promoting progress towards jointly agreed development objectives.

ECD and Public Sector Management

But the main rationale for ECD has to do with public sector reform. In the new global economic order, the state is increasingly operating in partnership with the private and voluntary sectors and it operates on the basis of subsidiarity vis-a-vis lower levels of government. Whether, in an era of globalized markets and increasingly active civil societies, a sound public sector means similar things in developing countries as in developed countries is a fundamental issue which merits thoughtful examination.

Assuming that the new public management is indeed relevant to our developing member countries, what capacity building in the public sector is about should not be very hard to describe. It aims at a public sector staffed with a lean, well trained and adequately compensated civil service subject to adequate oversight. It involves helping developing country governments keep out of productive activities best handled in the private sector. It means assisting developing country agencies to evaluate their policies and programs regularly and objectively. And it implies helping them to subject their public expenditures to strict accountability, to contestability and to the scrutiny of citizens.

Finally, it involves connecting developing countries to the rest of the world, not only by opening their borders to trade, capital and ideas but by connecting them to best practice in the economic, social and environmental realms. And this is where evaluation capacity development comes in. Increasingly, good government means government which judges its performance by results and which, wherever feasible, devolves its operations to the private and voluntary sector while protecting the public interest.

The papers in this volume represent a spectrum of experience with evaluation which ranges from mature, national systems to pioneer work in countries where evaluation has yet to earn acceptance.

Even where evaluation is firmly established, important changes are affecting what it does and how it works.

The emphasis now is less on the control and punitive aspects and more on the developmental benefits of the evaluative function. In many countries governments now choose a less interventionist stance, instead, providing managers with more autonomy—and of course correspondingly higher accountability. This means an increased need for objective and early reporting. It also means that our responses to evolving needs must be

flexible and pragmatic. Sharing the experience synthesized in these pages should open a wider horizon for our thinking about ECD.

The Role of Evaluation

Mark Baird

I would like to emphasize four issues which are central to how we evaluate our performance:

- well-defined objectives—so we know what we’re trying to achieve.
- a clear strategy—explaining how we propose to get there.
- monitorable indicators—to know if we’re on track.
- evaluation of results—for both accountability and learning.

Objectives

In the New Zealand (NZ) public service, performance contracts are usually related to outputs (goods and services produced) rather than outcomes (the effect of those outputs on the community). But this does not mean that outcomes do not matter. As Bale and Dale state “Chief Executives are responsible for specified outputs from their departments, while the Minister chooses which outputs will be purchased to achieve certain outcomes.”¹ The focus is on both the quality and quantity of outputs. And in deciding budget allocations, departmental performance is assessed against the contribution of outputs to the outcomes desired by the government. As Scott elaborates, “If outcomes could be established with equal clarity and controllability by managers, then outcomes would be superior to outputs for specifying performance.”² Indeed, when this is the case—such as the role of the Reserve Bank in controlling inflation—performance and accountability have been tied to outcomes.

In the World Bank, we are one step further removed from outcomes—in the sense that we work through intermediaries (normally governments) to achieve our objective of reducing poverty. But this does not mean that we can avoid defining our objectives. Shareholders—and the public—have a right to know what we think we are achieving with the use of their funds. And we cannot make rational budget decisions without assessing the contribution of different programs to our ultimate objective of poverty reduction. Hence the importance attached to results-based management.

As a first step, we have started to define more clearly (at the corporate level) the development objectives supported by the World Bank. We begin with the development goals established by the OECD’s Development Assistance Committee (DAC): halving the proportion of people in poverty, achieving universal primary education, and reducing infant mortality by two-thirds by 2015. These goals have political significance because

they have been endorsed by the governments of OECD countries. But for our purposes, these goals have their limitations:

- first, the feasibility of these global goals needs to be assessed against the reality of country-specific situations. And we need to understand what must be done—and where—to achieve the goals.
- second, goals for the next 10-20 years, are not particularly useful for making decisions—and evaluating performance—now. We need well-defined intermediate goals against which progress can be measured.
- third, these goals apply to low-income recipients of concessional aid, but IBRD is active in many middle-income countries—where progress has to be measured against a second generation of development indicators.

We are now working with the networks to address these issues within sector assistance strategies. The outcome should be a set of well-defined objectives in each sector and a set of target countries (where the performance gap is large and progress is critical to achieving the global goals). These objectives can then be used to assess the relevance and ambition of the objectives set in individual country assistance strategies.

Strategies

The World Bank's Executive Development Program exposes participants to many case studies from the private and public sectors. One case study is the New York Police Department's successful program to reduce major crimes in the City. The first step was to acknowledge that crime reduction—not administrative effort, such as the number of patrols dispatched—was the objective of the police. But having defined the objective, the focus of performance measurement and accountability was the agreed strategy to reduce crime. As Deputy Commissioner Maple said, "You don't get into trouble for increased crime, but for not having a strategy to deal with it." Performance was then monitored daily. And if crime was not reduced, tough questions were asked about the relevance and effective implementation of the strategy.

Our approach to accountability in the World Bank should be similar. We have a hierarchy of strategies: from the Strategic Compact at the institutional level, to sector assistance strategies which cut across regional boundaries, to regional business plans which cut across sectors and countries, and finally to Country Assistance Strategies (CASs)—working at the country level, the primary unit of account. These strategies should all influence each other—it is not purely a top-down or bottom-up model. But clearly the

CAS is the most important and most developed strategy document we have.

The CAS objectives are set in close consultation with our client countries. The Bank naturally focuses on those that are most relevant to our primary objective—poverty reduction—but these should also be high on the government’s own list of objectives. The CAS then assesses the potential contribution of possible Bank Group activities on country objectives, and prioritizes among these accordingly, within a given country budget and resource envelope. Three criteria are important:

- *Potential magnitude of impact.* How important is the activity? If the activity is fully implemented, what and how large will be its impact on overall country economic performance and sustainable poverty reduction?
- *Likelihood of country action.* What is the likelihood that the activity will be successfully completed and its full impact realized, taking into account country capacity and commitment to act?
- *Additionality of Bank contribution.* What is the Bank’s specific role and contribution in this activity, taking account of its track record and comparative advantage vis-a-vis other partners?

Indicators

This brings us to the critical question of measurement. Clearly performance is usually easier to measure—and certainly easier to attribute—from closer to the performer. That is why we have traditionally focused on inputs. And why the NZ public service has focused on outputs rather than outcomes. However, ultimately, we do want to assess the impact of performance on outcomes, and to measure progress toward our objectives. It is useful to have indicators at each of these levels: inputs, outputs and outcomes. However, there is also a dilemma here. With a strong theory which links inputs to outputs and outcomes, the level at which one measures is almost immaterial. But without that theory any one indicator (or all indicators) can be misleading. Hence the importance of not just measuring—but understanding why a particular indicator matters and how it should be interpreted.

In the World Bank, we have started tracking progress toward development objectives and on sector and country strategies. In addition, we have started to use indicators of internal performance, along the lines of the balanced scorecard used by many private sector companies. For private companies, the objective of the balanced scorecard is to focus on the main drivers of performance rather than the bottom line of return to

shareholders. We do not have a good measure of return to shareholders (development effectiveness). Hence the value of tracking progress on a number of indicators of internal performance, which we believe will improve the Bank's development effectiveness. These include measures of the quality of our services: from quality at entry of our projects and our economic and sector work (ESW), to measures of how well we manage the portfolio (realism and activity). We have also set service standards for timeliness—a primary concern of our clients.

The contribution of the Quality Assurance Group (QAG) has been critical to the development of meaningful, real indicators—and setting targets and tracing progress—in each of these areas. As a result, we now have real-time feedback on how well we are doing. Obviously, we have to be careful how these indicators are used. Some staff are concerned, that we will discourage risk-taking (by focusing on projects at risk) as a performance measure or emphasize speed over quality (through the rigid application of service standards). And the usefulness of some indicators may be questionable at the unit level. However, given the objectives of the Strategic Compact, these are clearly areas where we need to do better. And where improved performance will lead to better service delivery and higher development impact.

Evaluation

To be sure, we must continue to evaluate. Evaluation contributes to three basic functions:

- **Accountability:** making sure that public institutions, and their staff, are held accountable for their performance.
- **Allocation:** making sure that resources are allocated to those activities which contribute most effectively to achieving the basic objectives of the institution.
- **Learning:** making sure we learn from our successes and failures, to do things better in future.

Within the World Bank, we have one of the best evaluation systems in the development business. OED provides the center of this system. Its strength does not come from any particular advantage in the quality of its staff or its evaluation techniques but from its independence, which allows it to find things—and say things—others may prefer to leave alone. This is not an easy task. But it is made easier if the institution recognizes its value. When I joined OED in 1991, we were debating a controversial OED review of the experience with industrial policy. I was uncomfortable with some of this report's findings. But I was impressed when Mr. Stern said that it should go to the Board without

management interference. And when it met some resistance at the Board, President Preston insisted that it be published without change. He said, “We may suffer from the publicity this report generates; but we would suffer more by trying to suppress it.” Evaluation needs this kind of support.

OED recognizes that it needs to do more to make its evaluations useful for both accountability and allocation purposes. The key will be to move from the current project-by-project approach to a more holistic evaluation of the impact of the Bank’s programs in a country. This should cover both lending and advisory services, and assess the current impact of all completed and ongoing activities (not just the historical impact of completed loans). This is already happening through the Country Assistance Reviews. But we need more of these, timed to contribute to the CAS process. And we need to discipline this process with country program ratings, which can be compared over time and across countries, and aggregated to assess the overall performance of the institution. This should be supported by a system of self-evaluations—generated by the country team and subject to review by management.

But OED—and other development agencies—cannot substitute for an effective system of evaluation within our client countries. No public sector can afford to overlook the importance of clearly defining its objectives and priorities, assessing performance and results against well-defined benchmarks, and changing a bureaucratic culture into one that stresses client service and achievement of results. This shift is inherent in the process of democratization—accompanied by greater involvement of civil society in making public policy decisions and holding public servants accountable. Rather than an imposed requirement of donor agencies, evaluation now becomes a key instrument of good governance and institutional development within our client countries. We all have a responsibility to make sure that this function is nurtured and supported.

Let us ask for more attention to monitoring and evaluation within individual projects and programs. This has often been a requirement of donor agencies. But as we know from OED evaluations, this is not always done, or done well. One excellent example is a recent effort by DEC to evaluate the impact of decentralization and privatization reforms under five World Bank financed education projects. Building this research directly into the projects provided a wonderful opportunity to learn and to adjust project design and implementation along the way. In Colombia, the assessment of the voucher system for secondary education led the government to consider alternative subsidy schemes for private education. This type of research also provides an excellent data set—and counterfactual analysis—for evaluating project impact.

Conclusion

In discussing how we measure performance, one CEO said that he saw himself as a speed skater who had only one well-specified measure of success (time) and knew whether he had won or lost as soon as the race was over. However, he noted that the World Bank and its partners were more like figure skaters: we complete our performance and then depend on the judgment of a panel to determine how well we did. They can assess the technical quality of our jumps and will certainly note when we fall. But much of their rating will depend on the difficulty of the routine and the passion displayed in its execution. And that is where we are today with evaluation. While our techniques are important, we should never compromise our integrity: our work remains an art. But we know that our basic purpose—fighting poverty—is essential to the future survival and welfare of our planet. That is why our job is so difficult and so important. And that generates the passion needed to do our best.

-
1. M. Bale and T. Dale, "Public Sector Reform in New Zealand and its Relevance to Developing Countries," World Bank Research Observer, February 1998. See also critique by A. Schick in the same volume.
 2. G. Scott, "Government Reform in New Zealand," IMF, October 1996.

PART 2:
The Missing Link in
Good Governance



Evaluation Capacity and the Public Sector

David Shand

The major issues of evaluation are management rather than methodological ones; they are about how and where evaluation should be organized, located, planned and managed to best affect decision-making. They are about managing for performance, rather than simply measuring performance.

A major part of this aim is creating the right incentives for evaluation—on both the supply and demand side. Much past evaluation activity was supply driven, and as a result was not effectively used. The New Zealand public management system takes the view that there should be no need for special initiatives or rules on evaluation; if the management system is correctly structured, all participants will ensure that they have the information required for decisionmaking and accountability purposes.

However the issue of incentives—demand and supply—calls into question the purposes of evaluation. Depending on how evaluation information is to be used, incentives may operate differently. Increasingly, the evaluation literature stresses evaluation as part of continuous learning for performance improvement - improving management's knowledge base. Evaluation is thus seen as a normal and valued part of the management cycle.

This can be contrasted with the emphasis placed in countries, such as New Zealand and the United Kingdom, on evaluation for accountability or judgmental purposes. For example, in a typical contractual environment, did the ministry supply the required volume, quantity and quality of outputs and at the correct cost as specified in the purchase agreement between the minister and the ministry? Did the chief executive of the agency meet the terms of the performance agreement signed with the minister? Evaluation may operate differently in such an environment—and be seen as a special and potentially threatening activity. It may elicit defensive responses which do not improve program performance. This has led some commentators to suggest that evaluation activity should be functionally separated from questions of audit or accountability.

Those regimes which adopted a minimalist view of the role of the public sector, showed a distinct lack of interest in evaluation. However, if programs remain publicly funded but their delivery is contracted out, it is hard to see why evaluation would not be important

in program design and implementation. (Note the recent comment of a senior official of the UK Treasury, “we are not against evaluation per se, but we do worry that it might lead to demands for additional spending.”)

Aside from the uses to which evaluation information is to be put, there seem to be two other major issues.

1. To what extent should evaluation focus on strategic as opposed to operational issues, or on outcomes, as opposed to outputs or processes? On the latter point, the use of process evaluation seems to be increasing, particularly where there is an interest in benchmarking organizational performance through techniques such as total quality management—which involves comparing actual processes and systems with a set of normatively specified ones, or where difficulties in measuring outputs or outcomes lead to the use of process evaluation as a surrogate. The New Zealand and Australian work on evaluating policy advice is an example of this approach.
2. How does evaluation relate to the burgeoning activity of performance measures or indicators? In some countries the development of such indicators is being adopted with a great deal of enthusiasm—but perhaps with rather less skill, or realization that such measures or indicators, while they may be suggestive indications of the level of program performance, do not explain the reason for the particular level of performance. Such an explanation requires an in-depth evaluation to develop an understanding of the linkages between inputs through to activities or processes, through to outputs, through to outcomes.

Some very useful work has been done by the Public Management Service of the OECD, in developing best practice guidelines for evaluation. This reflects its view that there is a need to demystify program evaluation and to emphasize its performance management, rather than methodological aspects. The guidelines discuss issues such as:

- ***the need to generate effective demand for evaluation information.*** This includes the need for top level support—for both the political and managerial level—to request and use evaluation information in decisionmaking. The need has been suggested for “sticks, carrots and sermons”—sticks, such as refusing funding if evaluation is not undertaken; carrots, such as providing funds for evaluation—or requiring that a certain portion of program funds be used for evaluation; and sermons, through proselytizing through the written and spoken word the importance of evaluation. Flexibility in reallocating resources is also an important incentive issue; for example, if organizations are to be

encouraged to be self-evaluating they should be able to reallocate resources away from poor-performing programs—rather than lose those resources.

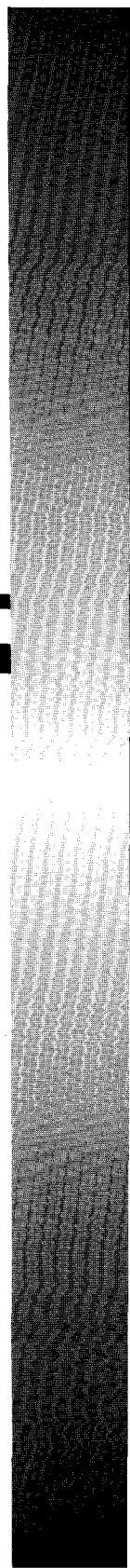
- **generating realistic expectations:** evaluations do not provide final answers; their findings are generally more persuasive than conclusive. Unrealistic expectations about “final answers” may lead some people to question the value of evaluation.
- **the need to institutionalize evaluation,** so that it becomes a regular part of the decisionmaking process. But it must not degenerate into a routine or “check list” exercise. Institutionalization raises the important issue of the linkage with decision-making processes; where should this occur, and to what extent should it be linked with the budgetary process? Some suggest that the budget process, because of its short-time horizon and political influences, is inherently hostile to the use of evaluation; others suggest that as the budget is the key decisionmaking process for resource allocation, evaluation can have little effect unless it is so linked. Whatever the judgment on this point, there are countries (USA, Canada) which illustrate considerable evaluation capacity and activity, which have had limited impact because of not being related to the decisionmaking processes. Related to this point, the timing of evaluations is important—they should fit into the decisionmaking cycle.
- **where to locate evaluation?** Should evaluation be centered in central agencies such as the Ministry of Finance? Should it be located in central offices of spending ministries? Should evaluation be undertaken primarily by operating units? What is the role of audit institutions, both external and internal? These questions can be phrased in a different way. To what extent should evaluation be internally as opposed to externally located? To what extent should it be undertaken by specialist evaluators as opposed to program staff? The OECD conclusion is that evaluation can and probably should be located in a number of different places.
- **planning evaluations is key,** and should probably occupy as much of the total evaluation effort as the actual carrying out of the evaluation. If the evaluation does not address the correct questions, it may be wasted.
- **ensuring relevance** is important. There may be some programs for which governments will not make significant use of evaluations in making decisions. Scarce evaluation resources should be targeted to areas where they are more likely to have an impact. To what extent should evaluation be ex ante or ex post? There may be

less interest in evaluating completed programs, as opposed to ongoing or new programs.

- ***communicating findings in an appropriate way*** is also important. Issues which arise here are the extent to which the report should contain recommendations as well as findings, and whether the evaluation should be publicly available. While transparency may be important for accountability, it may elicit defensive behavior rather than performance improvement.
- ***ensuring appropriate stakeholder involvement***. There seems general agreement that stakeholders, such as staff and users (clients) of the program, may provide useful information for the evaluation, but through consultation rather than direct participation. Measures of client satisfaction appear to be an increasingly important aspect of evaluations, reflecting an increasing emphasis on consumer-oriented government; but client-satisfaction cannot be the standard of judgment on a program; the perspectives of taxpayers are arguably just as important.
- ***ensuring high technical quality of evaluations***. This requires that the evaluation be methodologically sound, with due allowance made for limitations of information and other conditions. Evaluations should be as objective as possible—not unduly influenced by the vested interests of stakeholders nor by the feeling of consultant evaluators that the person paying their fee expects them to come up with a particular finding. Internal quality reviews of evaluations before their finalization may be a useful mechanism in maintaining quality; external reviews of the quality of evaluations are also found valuable, for example, the French Conseil Scientifique De L’Evaluation and the work of Auditors-General in Australia and Canada in reviewing the quality of evaluations by ministries. Professional evaluation societies have also developed codes of ethics to be used in carrying out evaluations.

The US budget critic, Aaron Wildavsky, said that the trouble with program evaluation was that “while a few butterflies were caught, no elephants stopped.” Our challenge is to ensure that evaluation “makes a difference.”

PART 3:
Experience of
Developed Countries



The Development of Australia's Evaluation System

Keith Mackay

Introduction

The Australian federal government has given high priority to ensuring that evaluations of its programs are conducted, and that the findings are used. The approach, which has been adopted to develop evaluation capacity, has entailed a combination of formal requirements for evaluation plus their strong advocacy by a powerful, central department (the Department of Finance). This has enabled evaluation to be linked both to budget decisionmaking and to the on-going management of government programs.

As a result, there is a high level of evaluative activity, and evaluations are actually used to assist Cabinet's decisionmaking and prioritization in the budget, and to support internal program management within line departments.

Each country is unique, and the Australian success has been supported by the existence and ongoing development of a sound public service infrastructure—including a high level of institutional and human capacity. Nevertheless, many lessons from Australia are highly relevant to the development of evaluation capacity in developing countries.

Many success factors are identified in this paper. These have included:

- the creation of an explicit, whole-of-government evaluation strategy;
- the existence of a powerful central department which has been a committed champion of evaluation, and which has continually identified and created new opportunities for influence and development;
- sustained commitment and support of the evaluation strategy over a decade; and
- implementation of related public sector management reforms which have given considerable autonomy to line managers and which emphasize bottom-line results—these reforms have provided incentives to conduct and use evaluation findings.

The Australian evaluation system has evolved from one of tight, formal controls and requirements to a more voluntary, principles-based approach. In this new environment it is hoped that the strong pressures for line departments to achieve and to demonstrate high levels of performance will be increased, and that the existing evaluative culture and

infrastructure will be strengthened. The latest reforms include, for example, sharper accountability and performance standards for managers, the widespread application of competitive tendering and contracting of public service activities, and the setting and reporting of explicit customer service standards. The latest wave of public sector reforms promise a closer integration of performance measurement—including evaluation—into performance management and into governance more broadly.

Genesis and Stages of Development

An understanding of the development of evaluation capacity in the Australian government is important because of the insights it provides for other countries. Evaluation capacity development (ECD) did not progress in a linear, logical sequence in Australia; it grew in response to prevailing imperatives. These various developments mean that much experience has been amassed concerning what works, what does not, and why. The objective of this paper is to share these insights.

Evaluation is not a new phenomenon in Australia. Cost-benefit analysis—the appraisal of investment projects—has been part of the scene for many decades. It has been conducted by several government economic research bureaus, and by other advisory bodies and line departments concerned with infrastructure and agriculture investments.

Formal program evaluation has a more recent origin. It too has been conducted by specialist research bureaus attached to line departments, and by special policy review task forces, focusing on areas such as labor market programs and social welfare policy. But as a discrete and on-going focus of government activity its heyday started in the 1980s.

The 1983 Genesis of Public Sector Reforms

The election of a reformist Labor government in 1983 provided an environment favorable to evaluation. The new government was determined to improve the performance of the public sector and at the same time restrain public expenditure through the annual budgetary process.

A series of public sector management reforms was implemented in the first several years of the new government. One aspect of these reforms was the desire to “let the managers manage” by devolution of powers and responsibilities—reflecting the philosophy that public sector managers would be strongly encouraged to improve their

performance if they were given greater autonomy and the potential to manage their departments with fewer central agency controls and less interference. Another aspect of the reforms was “making the managers manage”, and this thread was more directly linked to powerful budgetary pressures.

The tangible changes included:

- substantial autonomy for departments in their spending of administrative expenses (including salaries), but with these administrative expenses being strictly cash-limited;
- greater surety about future resource availability to departmental managers via a system of three-year forward estimates of administrative and all other program expenses; and
- a major reduction in the number of departments through amalgamation, to achieve less balkanized policy advice and to encourage the internal reallocation of resources through portfolio budgeting.

A related set of principles was embodied in the Financial Management Improvement Program, which included program management and budgeting. These principles emphasized the importance for departmental managers of ensuring that program objectives were realistic, to help guide managers and staff. The principles also encompassed a focus on the efficiency and effectiveness of programs—on program performance—through sound management practices, the collection of performance information and the regular undertaking of program evaluation. Guidance material which extolled the virtues of these principles and clarified the concepts was issued by the Department of Finance (DoF) and the then Public Service Board, another central agency.

The public sector reform initiatives were introduced in a stringent budgetary climate. Macroeconomic concerns provided sufficient motivation to induce the government to reduce the share of federal government outlays in GDP from 30% in 1984-85 to just over 23% in 1989-90. The impact of these cuts on government programs was even greater than these raw statistics suggest, because the government was also determined to increase the real level of welfare payments to the most disadvantaged in society. It achieved this by means of tight targeting of benefits via means-testing, which entailed substantial reductions in “middle-class welfare.”

This package of financial management and budgetary reforms was substantive and wide-ranging. For several years it placed the framework of public sector management in Australia at the forefront of developed nations.

These initiatives helped to create an enabling environment which encouraged performance management. However, while the reform initiatives were necessary they were not sufficient to ensure that evaluation and other types of performance measurement became accepted as desirable routine activities.

DoF was a major architect of many of these reforms, reflecting its role as budget coordinator and overseer of the spending of other departments. DoF was keen to distance itself from the detail of spending issues, where it was often bogged down in minor spending bids and disputes with departments. Its concern with budget spending encompassed both a priority on cutting government outlays, and finding ways to make spending more efficient and effective.

Growing Focus on Evaluation

The concern with “value-for-money”—that is, the efficiency and effectiveness of public expenditure—helped to lay the groundwork for DoF’s provision of advice to departments on the evaluation of their programs. DoF began this formally with the publication of an evaluation handbook in 1986.

At around the same time there was growing disquiet in DoF and other central agencies about the lack of any real progress made by line departments in managing their performance. In early 1987 the Minister for Finance secured Cabinet’s agreement to a formal requirement that all new policy proposals for Cabinet consideration should include a statement of objectives and performance measures, as well as proposed arrangements for future evaluation. Line Ministers and their departments were required to develop plans for the systematic and comprehensive monitoring and evaluation of the performance of their programs, and for the reporting of these reviews to government. DoF was to be kept informed of departments’ evaluation plans.

DoF augmented its earlier advisory panel through providing additional guidance material in 1987, and by presenting a basic evaluation training course. By early 1988, however, the evaluation plans prepared by departments could best be described as patchy (many were poor), and it had become apparent that a more fundamental examination of evaluation practices in departments was warranted. Thus DoF undertook a diagnostic study reviewing departments’ evaluation progress and the overall health of evaluation activities in the public service. The study found:

- a lack of integration of evaluation into corporate and financial decisionmaking;

-
- that evaluations tended to focus on efficiency and process issues rather than on the more fundamental question of overall program effectiveness—whether programs were actually meeting their objectives;
 - a poor level of evaluation skills and analytical capacity; and
 - that the role of central departments, especially DoF, was unclear.

The Formal Evaluation Strategy

This diagnostic study laid the groundwork for a major submission from the Minister for Finance to Cabinet in late 1988 seeking—and securing—its agreement to a formal, ongoing evaluation strategy for all departments. A key and continuing principle underlying this strategy was that “the primary responsibility for determining evaluation priorities, preparation of evaluation plans and conduct of evaluations rests” (with line departments).

Cabinet’s agreement to the evaluation strategy was expressed in a formal cabinet decision. For the federal government and its public servants, such decisions virtually have the force of law. Indeed, efficiency audits conducted by the Australian National Audit Office (ANAO) often focus on the extent to which cabinet decisions have been complied with. So expressing public sector reform initiatives as cabinet decisions has encouraged the oversight involvement of the ANAO.

The evaluation strategy had three main objectives. It encouraged program managers within departments to use evaluation to improve their programs’ performance. It provided fundamental information about program performance to aid Cabinet’s decisionmaking and prioritization, particularly in the annual budget process when a large number of competing proposals are advocated by individual Ministers. Lastly, the strategy aimed to strengthen accountability in a devolved environment by providing formal evidence of program managers’ oversight and management of program resources. The emphasis was on transparency, which is of interest to Parliament, particularly in the Senate’s processes of budget scrutiny and approval.

The evaluation strategy has provided the framework and driving force underlying the progress with ECD in the Australian government since that time. Its key components comprised four formal requirements. These were:

- that every program be evaluated every 3-5 years;
- that each portfolio (comprising a line department plus outrider agencies) prepare

an annual portfolio evaluation plan (PEP), with a 3-year forward coverage, and submit it to DoF. PEPs comprise major program evaluations with substantial resource or policy implications;

- that Ministers' new policy proposals include a statement of proposed arrangements for future evaluation; and
- that completed evaluation reports should normally be published, unless important policy sensitivity, national security or commercial-in-confidence considerations apply, and that the budget documentation which departments table in Parliament each year should also report major evaluation findings.

A crucial aspect of this evaluation strategy was the role which Cabinet agreed to assign to DoF. Cabinet expressed its expectation that DoF would have the opportunity to make an input to PEPs and to the terms of reference of individual evaluations to ensure that these were consistent with government-wide policies and priorities, and that DoF would be available to participate directly in selected evaluations, subject to negotiation between DoF and the line department (or between their Ministers if a dispute arose). DoF was also to provide detailed advice and handbooks on evaluation methodology, and on management information systems, and take the lead in identifying and sharing best practice.

A key to the public sector reforms which DoF had advocated was the need for central departments to get away from detail and out of a control mentality. While the leadership of DoF were true believers in the utility of evaluation, they had advocated a more hands-on involvement for DoF with some hesitation, because it represented an activist and closely monitoring approach. DoF advocated a powerful role for itself to Cabinet only after there was clear and strong evidence of the failure of line departments to live up to the rhetoric of program management and budgeting which departments had also espoused. The immediate focus of the evaluation strategy was to ensure that good evaluations of the right programs were available, rather than the creation of a performance-oriented culture. The latter focus has emerged as the main area of emphasis only relatively recently.

While the evaluation strategy was still at the proposal stage, most line departments stated their acceptance of the utility of evaluation but expressed real concern about 'intrusive' DoF involvement. This stated acceptance, at odds with the actual reality of management in most line departments, was a feature of other performance measurement and performance management initiatives introduced in later years.

Later Developments

The next milestones for ECD in Australia were two reports, from a parliamentary committee in 1990 and from the ANAO in 1991. These reports acknowledged the substantial effort devoted to the planning and conduct of evaluation, but argued for renewed efforts. In particular, they noted the variation in the extent of evaluation activity in different departments. They criticized some departments for their poor choice of programs to evaluate, and for the focus of, or issues covered by, their evaluations—particularly an insufficient focus on effectiveness issues. They argued that DoF should be more active in encouraging departments to plan and undertake evaluations.

These reports were soon followed by the creation in 1991 of a separate branch within DoF, responsible for providing evaluation advice, support, training and encouragement to other departments (and also within DoF itself). This branch, which had nine evaluators, acted as a focal point and catalyst for evaluation throughout the Australian public service. The branch was co-located with two other branches, responsible for overall coordination and management of the government's annual budget process, and for public sector management reforms more generally.

Evaluation in the Australian government—as measured by the extent of evaluation planning, conduct and use—had achieved a healthy and vigorous state by the mid-1990s.

However, by that time DoF was concerned about departments' poor progress in articulating clear and achievable objectives for their programs, and in collecting and reporting meaningful performance information. These concerns were confirmed by two reviews of departments' annual reports and of their budget documentation which DoF commissioned. This might appear paradoxical, because evaluation is at the more difficult end of the performance measurement spectrum, and was generally being done well, yet the setting of program objectives and the collection of frequent performance information, at the easier end of the spectrum, were being done poorly. But in fact the situation reflects the emphasis and success of the evaluation strategy in encouraging and mandating evaluation, with much less emphasis being placed on ongoing performance monitoring.

Thus in 1995 DoF secured Cabinet's agreement to a rolling series of comprehensive reviews (staggered over 3 years) of the program objectives and performance information of all programs in all departments. These reviews are being conducted jointly by DoF and each line department, with the results being reported to their respective

Ministers and to Cabinet. The reviews focus in part on analyzing the existing situation, but attach more importance to identifying ways in which objectives and performance information could be improved, and in mapping out and committing to a plan of action to achieve these improvements.

This illustrates that public sector management initiatives often require a push from the center to make them happen in many departments.

The election of a Liberal/National Party government in 1996 led to an emphasis on cutting bureaucracy, red tape and formal reporting requirements. The new government required that existing programs be comprehensively reviewed by Ministers and their departments to establish whether they should continue, or be abolished or be devolved to another level of government. For those programs which were to continue to be delivered at the federal level, Cabinet expected that a competitive tendering and contracting process be undertaken wherever possible.

One issue which had emerged with the formal evaluation requirements in recent years was that their concern with bureaucratic process was no longer appropriate—the length of some portfolio evaluation plans, for example, had grown from a recommended 20 or 30 pages to over 120 pages. A consensus had emerged within the bureaucracy that while it was important to have evaluation findings available to assist decisionmaking by program managers and by Cabinet, detailed and elegantly-worded plans were not necessary to achieve that end.

A related and powerful strand of thinking holds that departments should not be encumbered by excessive controls on their internal activities, as long as departmental heads and senior executives are responsible for performance, and that responsibility is reflected in their employment contracts. This is essentially a “let the managers manage” philosophy, and is analogous to the one adopted in the early 1980s. A difference, however, is the greater scrutiny by Ministers of the performance of their department heads—as exemplified in the more widespread use of employment contracts—plus the apparent readiness of Ministers to remove department heads where their performance is judged to be wanting. And perhaps the most important difference is the progress in the intervening years in establishing an evaluation culture, and in establishing departmental infrastructures to support performance measurement and management.

This led in late 1997 to a further development in the evaluation strategy, into a principles-based, performance management framework. This approach was accepted by Cabinet, and is now government policy. Such a Cabinet-endorsed, principles-based

approach provides guidance to the heads of line departments by emphasizing the good practice features of performance management and measurement (the latter includes evaluation and ongoing performance monitoring). It reflects the strong expectation that CEOs and senior executives will continue to plan, conduct and use evaluation, and so it implicitly takes the progress achieved to date as a given. These issues go to the heart of the question whether central controls should be tight or loose.

The Development of Evaluation Capacity: Evidence

It is important to have a realistic understanding of the extent of ECD in Australia. National reviews of progress with evaluation in developed countries often present too rosy a picture, particularly when external reviewers confuse rhetoric with reality. Checklists of evaluation activities undertaken do not necessarily translate into the widespread conduct of evaluation, nor into quality and rigor in evaluation, nor into its actual use in program management and government decisionmaking.

The true extent of ECD in Australia can be assessed by considering the planning, conduct, quality and use of evaluation.

Evaluation Planning

All government departments have prepared portfolio evaluation plans since 1987-88. These were intended to comprise the major evaluations in each department and its outrider agencies—in recent years about 160 of these evaluations had been underway at any given time. Most of these evaluations were major, in that the programs had significant policy or spending implications, although a significant minority, particularly for the smaller departments, were of less important programs or of the efficiency aspects of large programs. (The plan guidelines issued by DoF recommended that the main focus of these evaluations be on issues of program effectiveness. Departments were separately encouraged to plan and to undertake minor evaluations for their own internal management purposes).

Line departments themselves decided which programs should be included in the plans for evaluation, and which issues the evaluation terms of reference would cover. However, DoF would usually endeavor to influence departments' choice of evaluation priorities by making direct suggestions. In doing so DoF would attempt both to anticipate and to create the information needs of Cabinet. Where DoF has had difficulty in persuading departments, it has

sometimes approached Cabinet directly to seek its endorsement of particular evaluation suggestions and of detailed terms of reference; Cabinet almost invariably accepts DoF's suggestions.

The Cabinet-endorsed, formal requirement under the evaluation strategy that portfolio evaluation plans be prepared and submitted to DoF certainly provided a powerful incentive to line departments to prepare plans and to take them seriously. Another influential factor was the issuing by DoF of formal guidelines to departments on the desirable content of these plans, together with follow-up monitoring and reminders to departments about the need for the plans. The evaluation branch of DoF conducted internal reviews of the content and coverage of these evaluation plans, and provided feedback and prompting to departments as well as by identifying good practice examples. In seven efficiency audits and two 'better practice' guides on program evaluation and performance information, the ANAO has also repeatedly reminded departments about the importance of systematically planning their evaluation activity.

The formal requirement that all programs be evaluated every 3-5 years was also influential in creating a climate in which evaluation is the norm rather than the exception. The concept of regular, comprehensive coverage of programs also encourages a planned, staged approach to evaluation. This formal requirement should not be accepted at face value, however. It is very seldom that all aspects of a program¹ are included in any single evaluation. Instead, it is usual that an evaluation will focus only on certain key problems or aspects of a program. The challenge is to ensure that these difficult issues are actually evaluated, and this is where DoF has been active via persuasion and direct involvement in individual evaluations.²

Conduct of Evaluation

Most departments have chosen to set up evaluation units to coordinate their formal evaluation planning. At their smallest, these units comprise two or three individuals. In some departments, such as employment, a separate branch of 20-25 staff members is responsible for evaluation planning, provision of advice on evaluation methodology, participation in steering committees, and the conduct of major evaluations, particularly in labor market programs (but typically not of education programs, which comprise a substantial proportion of the department).

There is no standard approach in departments as to how evaluations will be conducted—this is viewed as a line management decision. Some evaluations involve a wide array of external and internal stakeholders, either by participation in an evaluation steering committee, or less

commonly by participation in the actual evaluation team. Some evaluations are conducted by a central evaluation unit, but it is more common for line program areas to take this responsibility. These line areas would be responsible to the top management of the department for the quality and rigor of the evaluation.

For the more important evaluations—those listed in portfolio evaluation plans—some external involvement would be typical, via provision of suggestions and comments on the terms of reference and proposed evaluation methodology, participation in the steering committee, and provision of comments on drafts of the evaluation report. But there is no standard approach to this external involvement—it would be determined by the willingness of line department to involve outsiders, and also by the interest and availability of outsiders such as central agencies. For programs with major resource or policy implications, DoF would usually be keen to be involved, and would apply pressure to ensure its participation.

A recent ANAO survey found that, for evaluations conducted over 1995-1997: about half examined the delivery of products or services to external clients, and a further 30% were associated with matters internal to the department. One-third of the evaluations examined the appropriateness of new or established programs, and 15% were directed at the development of policy advice for the government.³

The large number of evaluations in progress, and the fact that over 530 evaluation reports have been published over the last four years or so, attest to the existence of extensive evaluation activity in the Australian government. This has provided a growing library of evaluation findings.

DoF publishes a register of published evaluation reports, and this helps to monitor the progress of individual departments' activities. More importantly, it helps to share evaluation practices and methods among departments and this provides some quality assurance because the public availability of these reports exposes them to peer scrutiny. A recent survey of all departments by the ANAO found that 75% of evaluations conducted in 1995 and 1996 were released to the public or were available on request.⁴

Evaluation Quality

Quality of evaluation reports is a more difficult dimension to measure. The rigor of program evaluations depends on the expertise and objectivity of the evaluators. A recent assessment of the quality of a small sample of evaluation reports was commissioned by the ANAO. It found that over a third of a sample of evaluation reports suffered from

methodological weaknesses. It is certainly true that some published evaluations are of low quality: some of these may be produced for self-serving purposes, such as to provide a justification for the retention or expansion of the program.

DoF's own experience of evaluations is that their quality can vary enormously. The extent to which this should be a matter of concern is another matter: the issue to consider here is the intended uses of evaluations. If the intended audience of an evaluation is Cabinet (to aid its decisionmaking) or Parliament (for accountability purposes) then a poor quality or misleading evaluation gives cause for serious concern. DoF has certainly been willing to provide Cabinet with a dissenting view on the quality of an evaluation when it is used by a line department to attempt to influence Cabinet debate. (Line departments would typically try hard to avoid such disagreements, which would be virtually guaranteed to attract the condemnation of Cabinet).

Where line departments allow poor-quality evaluations to be conducted (and these evaluations are intended for internal program management purposes) however, then there is an element of *caveat emptor*. Thus the extent of evaluation quality assurance or quality control could be regarded as an issue for departments' internal management to address.

A commonly-asked question is how evaluation quality can be assured, and what the role of DoF is or should be in guaranteeing quality. In past years, parliamentary committees and the ANAO have argued that DoF should take a strong role as an independent check on departments. DoF has preferred to seek to participate directly in certain major evaluations, usually via steering committee membership, thus ensuring that evaluations address the difficult questions and do so in a rigorous manner. But it would be a very resource-intensive activity to undertake detailed reviews of all evaluations, and it would also be inconsistent with the devolutionary reform philosophy for DoF to do so.

The ANAO has consistently argued that departments should set up central oversight procedures to achieve quality assurance of evaluations conducted by line areas within the department. There is certainly evidence from those few departments which have followed this approach, that it is an effective means of bringing to bear needed evaluation skills and expertise, and of ensuring evaluation quality.

Use of Evaluation

A bottom-line issue is the extent to which evaluation results are actually used. If their use is patchy or poor then there really is little point in conducting evaluations. The large

volume of evaluation activity in itself provides some reassurance that evaluation findings are being used—in an era of very tightly limited administrative expenses, departments would not bother to conduct evaluations unless they were going to be used.⁵ (And DoF would not bother to advocate and work to influence the evaluation agenda unless it perceived high potential value in their findings).

There are differences in the perspectives of a line department and a central agency, of course, and these influence the types of evaluation which are conducted and their probable uses. Line departments have traditionally been focused on evaluations concerned with program management and improvement, whereas the primary focus of central agencies is on overall assessments of the worth—the

cost-effectiveness—of the program. But this distinction has become considerably blurred in recent years with the initiative of portfolio budgeting—this provides line Ministers and their departments much greater say in making decisions about portfolio spending and program priorities. Portfolio budgeting has encouraged line departments to focus more on cost-effectiveness.

A recent survey of departments by the ANAO found that the impact or use of evaluations was most significant with respect to improvements in operational efficiency, and to a lesser extent to resource allocation decisions and the design of service quality improvements for the benefit of clients.⁶

A 1994 review of evaluation activities in one of the largest departments, the Department of Employment, Education and Training (DEET) found that considerable use was being made of evaluation findings. The review, conducted jointly by DEET and DoF, surveyed a large sample of completed evaluations and found that:

- 55% had led to changes in program management;
- 8% had resulted in an improvement in the quality of program outputs;
- 10% had influenced new policy proposals sent by the line Minister to Cabinet for its consideration; but
- there was considerable unevenness within DEET in the quality and use made of evaluation findings—some line areas of DEET had maintained a high standard, while other areas were either consistently poor or were uneven.

Source: Crossfield and Byrne (1994).

There is clear evidence that evaluations have been used intensively in the budget process. DoF has conducted several surveys of the extent of the influence of evaluation findings on the

budget proposals submitted to Cabinet. These have been surveys of DoF officers to seek their judgment concerning the extent of the influence of evaluation. While the survey results provide no more than a broad indication of the extent of influence, they are revealing.

In the 1990-91 budget, some A\$230 million (then about US\$175 million) of new policy proposals submitted by line Ministers were judged to have been directly or indirectly influenced by the findings of an evaluation. By 1994-95—the latest year for which estimates are available—this had risen to A\$2300 million. (Measured in dollar terms, the proportion of new policy proposals influenced by evaluation rose from 23% to 77% over that period). In most cases the influence of evaluation was judged by DoF to be direct.

Some unique features of the 1994-95 budget resulted in these figures being particularly high.⁷ Nevertheless, the results indicate the importance which public servants—in their preparation of the details of new policy proposals—and Ministers have attached to having evaluation findings available. Overall, it has been very important to have had the support of key Cabinet and other Ministers in encouraging portfolios to plan and conduct high-quality evaluation.

Evaluation can also have a significant influence on the “savings options” put forward by DoF or by portfolios for Cabinet consideration in the budget process. (Savings options are areas of government expenditure which could be trimmed or abolished entirely). In 1994-95 about A\$500 million of savings options—or 65% of the total—were influenced by evaluation findings.⁸

It seems likely that this emphasis on evaluation findings has been encouraged by the nature of the budgetary system in the Australian government. The country has a well-functioning policymaking mechanism which makes transparent the costs of competing policies and encourages debate and consultation among stakeholders within government.⁹ In this “marketplace of ideas” evaluation findings can provide a competitive advantage to those who use them.

One issue which it is important to appreciate is the realistic limits to the influence of evaluation on Ministers’ or Cabinets’ decisionmaking. The evaluation paradigm in an investment project is typically that of cost-benefit analysis: a project is warranted if, but only if, its benefit-cost ratio is greater than one. But program evaluation is a more qualitative science: it can help identify the efficiency or effectiveness of existing, ongoing programs but it can rarely provide an overall conclusion that the activity is worthwhile.

An Example of the Use of Evaluation to Help Government Cut and Reprioritize Its Programs

In the 1996-97 Budget the new government was determined both to reduce and to reprioritize government spending. Particular focus was given to labor market and related programs, which accounted for spending of A\$3,800 million annually (about US \$2,900m).

The Minister for Employment articulated the government's overall policy goal as being to provide assistance to the long-term unemployed and to those at risk of entering long-term unemployment. This focus was adopted both for equity and efficiency objectives—the latter pursued by achieving a better matching of labor supply and demand. At the same time, the Minister wanted to achieve better value for money from labor market programs in the tight budgetary environment.

Australian and international evaluation findings were drawn on heavily to help guide the policy choices made. The Minister highlighted the relative cost-effectiveness of different labor market programs. A key measure of this was estimated by calculating the net cost to government for each *additional* job placement from different programs—as measured by the *increased* probability of an assisted person being in a job 6 months after they had participated in a labor market program. (The baseline was a matched comparison group of individuals who did not participate in a program).

Evaluation findings showed that the JobStart program, which provides wage subsidies, had a net cost of A\$4,900 per additional job placement, whereas the JobSkills program, which was a direct job creation program, had a net cost of A\$76,600. The Minister noted that “the Government will be ... concentrating its efforts on those programs which have proven most cost-effective in securing real job outcomes.” As a result, the JobStart program was retained while the JobSkills program was substantially scaled back and more tightly targeted to jobseekers who were particularly disadvantaged.

Total savings to the government from its reduction and reprioritization of labor market programs were about A\$1,500 million over two years.

Cabinet also commissioned a series of major evaluations of its new labor market programs and of the new arrangements for full competition between public and private employment service providers.

Source: Senator Vanstone (1996); DEETYA (1996, 1997); Commonwealth of Australia (1996).

To give an example, if a government decides to allocate a large amount of spending to the unemployed, program evaluation findings can help to map out the probable consequences of alternative types of labor market intervention such as wage subsidies, public sector job creation, or labor market regulation. Program evaluation can be used to identify the cost-effectiveness of each type of intervention. But program evaluation can usually not identify the overall amount of resources which should be allocated. That is a political decision. The most that program evaluators can realistically and legitimately hope is that their findings are an influential input into government's decisionmaking and prioritization among competing proposals and programs.¹⁰

Success Factors and Impediments—What Has/Has Not Worked, and Why

The preceding discussion identified a number of success factors and impediments to success. These are now considered, together with some thoughts on their relative importance.

The Department of Finance (DoF)

DoF has been central to the development of ECD in the Australian government: there have been advantages and some disadvantages in this. Overall, however, DoF's considerable authority with departments and with Cabinet has given it the strength to influence the acceptance of evaluation. This has been probably the single most important factor in the substantial degree of progress with ECD in Australia.

DoF has been an influential, devil's advocate in advising Cabinet about the level of funding which should be allocated to departments for different government programs. As part of this function it provides advice on the new policies proposed by line Ministers, and on possible savings options.

Being the devil's advocate does not endear DoF to other departments, and in fact is an impediment to close cooperation and trust between other departments and DoF. On the other hand, DoF works day-to-day with departments advising on funding and policy issues, participating in reviews and evaluations of their programs, and providing advice on evaluation and other public sector management tools. DoF's evaluation branch with its nine evaluators provided desk officer assistance to departments with advice on methodology, best practice, provision of training courses, publication of evaluation handbooks and guidance material, and support for evaluation networks of practitioners.

The nature of these relationships can vary considerably, with other departments viewing DoF at best as a useful source of advice and treating it with wary respect, and at worst with downright hostility. The former relationship is much more common.

As the evaluation champion, DoF has succeeded in getting evaluation 'on the agenda' in its work with departments. This stands in contrast to the situation which faced the Office of the Comptroller-General (OCG) in Canada in the early 1990s. OCG was a stand-alone, specialist body responsible for attempting to influence line ministries to adopt evaluation as a management tool. But OCG was seen as tangential to mainstream government activities, and this undercut its influence. It is interesting to note that the OCG's functions have now been relocated as part of the Treasury Board Secretariat, the Canadian equivalent of DoF, to increase its leverage in dealing with line ministries. This relocation was undertaken after the 1993 review by the Auditor General of Canada of overseas practices, including Australia's.¹¹

Another advantage of having DoF responsible for evaluation oversight is that it ensures a direct influence on the line areas of DoF which oversee the line departments. Before the devolutionary reforms of the past fifteen years DoF was heavily involved—some would say “bogged down”—in the detailed scrutiny of departments' spending activities. The more recent focus on evaluation and other public sector reforms has helped foster a greater focus in these line areas on bottom-line outcomes and value for money; DoF is simply too important a bureaucratic player to allow it to remain with outmoded attitudes and activities. However, achieving this needed cultural change in DoF has taken a number of years, and has involved substantial staff turnover.

The greater focus on value-for-money has also flowed through to the nature and quality of policy advice which DoF provides to Cabinet. That advice has increasingly drawn on available evaluation findings, thus also helping to raise the profile of evaluation with line departments. DoF's involvement in selected evaluations also provides some quality assurance to Cabinet about the evaluation findings on which proposals for new policy might be based.

One example of the evaluative culture which has grown in DoF, and arguably in the Cabinet itself, was Cabinet's agreement to commission some 60 major reviews of government programs. These reviews had been suggested by DoF in the 1993-94 to 1995-96 budgets, with most focusing on issues of effectiveness and appropriateness; many of these reviews surveyed and summarized existing evaluative information, rather than conducted in-depth evaluations themselves.

The reviews related to aspects of programs which collectively involved about A\$60 billion in annual expenditure. These reviews were designed as an urgent response to emerging budget pressures, and might best be viewed as complementary to the regular cycles of evaluation as reflected in portfolio evaluation plans. Such Cabinet-endorsed reviews can be a useful vehicle for a DoF to use if line departments strongly resist DoF suggestions about evaluation priorities.

One benefit to line departments from DoF involvement in individual evaluations is that DoF can draw on evaluation skills and experience, spanning the whole breadth of government activities. Most DoF officers are usually not specialists in technical evaluation issues—they are expenditure and financial policy specialists. Perhaps their greatest potential value-added is the objective approach which they can bring to bear to an evaluation—DoF officers are comfortable in asking difficult questions about a program's performance, and in steering an evaluation towards these issues. This independent and questioning approach has provided a useful counterpoint to the in-depth knowledge (but often partisan approach) of program areas within line departments.

Line Departments

Most of the day-to-day work of line departments relates to ongoing program management. One of the three objectives of the evaluation strategy is to encourage program managers to use evaluation to improve their programs' performance. This has proved surprisingly difficult at times.

To those who understand the potential contribution of evaluation its utility seems self-evident. Evaluators often plead along the lines "How can you program managers improve your programs and better meet the needs of your clients unless you carefully evaluate program performance?" Unfortunately, appeals to the professionalism of senior executives are not notably effective. Most managers understand the potential benefits of evaluation, but do not place it at the top of their day-to-day priorities, particularly in a climate of tightly constrained and diminishing resources.¹²

Experience within line departments in Australia shows that a highly supportive culture is necessary if major evaluations are to be planned, resources allocated to properly manage and undertake them, and the findings implemented. The commitment of departmental secretaries (the CEO) to achieving improvement in program performance is paramount in fostering such a results-based management culture.

Over at least the past decade, the tenure of secretaries has often been brief. This turnover has meant that some departments have had a series of secretaries who have placed varying priority on evaluation and the departmental effort devoted to it. While an evaluative culture can be slow to build, it can be reduced very quickly.

Reprioritization of labor market programs provides one, high-profile example of the potential benefits of evaluation to Ministers and their departments. More generally, there has been an emphasis in recent years on “portfolio budgeting.” This includes the setting by Cabinet of portfolio spending targets at the start of each annual budget round. In the tight budgetary environments that have predominated, the nature of these targets usually imply that if Ministers wish to propose any new policies then these must be funded from within the portfolio’s spending envelope. Evaluation has been one tool to assist Ministers and their secretaries in the design of new policies and in the prioritization among existing policies (subject, of course, to Cabinet’s endorsement).

A portfolio budgeting approach helps to ensure that the focus of Ministers and their departmental secretaries is on value-for-money issues, as well as management-oriented efficiency issues. Thus portfolio budgeting is a key part of the devolutionary public sector reforms in Australia.

The requirement for portfolio evaluation plans has necessitated that departments set up a bureaucratic infrastructure to prepare them (which may be a flaw when the government is determined to cut red tape and the bureaucracy, as at present). In about three-quarters of departments this has involved the creation of a committee, usually chaired by a deputy secretary of the department, to meet regularly, canvass candidate programs for future evaluation, and monitor the progress of evaluations already underway.¹³ This work itself generates bureaucratic momentum. Most departments involve their Minister and their Minister’s office by seeking their comments on (and clearance of) the draft evaluation plans.

It is difficult to speculate with any confidence how the evaluation “scene” in Australia would have looked in the absence of a powerful champion such as DoF. Some larger departments, such as the Departments of Employment and Health, would no doubt have had a substantive evaluation effort in any event—the evaluation emphasis in the Department of Employment pre-dated the government’s formal evaluation strategy. However, informal discussions with senior executives in those departments have emphasized the catalytic influence of DoF even in their departments. Executives responsible for the central evaluation areas in line departments have generally found DoF a natural ally in helping to persuade more traditional administrators in their departments to adopt evaluation as a valued management tool.

Most departments have chosen to rely on program managers and their staff for the actual conduct of evaluations. This devolutionary approach has helped to “mainstream” evaluation as a core activity of each line area, and has ensured that evaluations draw heavily on the program knowledge and experience of those who actually manage the program. It has also led to a greater appreciation of the complementarity—and sometimes the substitutability—between in-depth program evaluation and the more frequent monitoring of performance via the collection of ongoing performance information.

The devolutionary approach has also secured “ownership” by program managers of the evaluation findings. These are important advantages, and provide a strong contrast with externally conducted evaluations, reviews or performance audits where lack of program knowledge and commitment to implement the findings has often significantly undermined the impact of findings.

But there have also been disadvantages to this devolved approach. One has been a lack of evaluation skills in many program areas and inexperience in conducting evaluations, (as suggested by the recent survey by the ANAO of a sample of evaluation reports).

Basic training in evaluation skills is widely available in the Australian government—provided by DoF in particular¹⁴—as is DoF and departments’ own guidance material such as evaluation handbooks. There is a substantial community of evaluation consultants in Canberra, including numerous academics with either subject area knowledge (such as health issues) or with specialist research and analysis skills. Despite this, the recent ANAO study showed that 20% of departments are concerned about the lack of available training in advanced evaluation techniques.

Some departments have addressed the need for more advanced skills and experience by setting up a central evaluation unit to provide advice on methodology and to participate in evaluation steering committees. The Department of Health has pursued evaluation quality assurance in a devolved environment by ensuring that adequate skills and resources are available to program managers together with structural arrangements in place, such as technical panels and steering committees.¹⁵ That department, like some others, puts a lot of effort into training its staff to enhance their analytical and research skills.

Another disadvantage to the devolved approach is that program staff are often too close to their program to view it objectively and to ask the hard, fundamental questions concerning its performance and the need for the program to continue. External participation by a central evaluation unit or by peers from other programs in working groups or steering committees

has been one way to address this. External participation has often included DoF for major evaluations, and this has been another means of fostering objectivity and rigor.

In only one agency, the Aboriginal and Torres Strait Islanders Commission (ATSIC), is there a separate evaluation body—the Office of Evaluation and Audit (OEA)—which has statutory independence. The independence of the OEA helps to answer claims of ethical difficulties and corruption in the administration of some ATSIC programs. A body such as OEA can be effective in ensuring that accountability objectives are met. The impact of its evaluations and reviews may have been reduced, however, by perceptions that OEA lacks fluency in program understanding and has not secured ownership by ATSIC program managers.

Australian National Audit Office (ANAO)

The ANAO has been a central presence since the inception of the formal evaluation strategy. In endorsing the strategy, Cabinet agreed with the proposition that “it is expected that (the ANAO) would contribute to the proposed evaluation strategy through audits of evaluation processes within departments and agencies, including their follow-up on evaluation findings.”

Since 1990 the ANAO has pursued this “sheepdog” task vigorously, both with respect to line departments and to DoF. (It is notable that the Canadian Auditor General has worked in the same way.)¹⁶ The ANAO has conducted seven performance audits into the evaluation and performance information practices of a number of departments during that period, as well as on the overall, government-wide progress with the evaluation strategy. It has also published two “good practice” guides, one of which was prepared jointly with DoF.

In addition to these reviews of evaluation and performance information, the ANAO has placed increasing emphasis on the conduct of performance audits into the economy, efficiency and effectiveness of programs. The ANAO completes about 40 performance audits annually, and these now account for about one half of the Office’s overall activity. The ANAO takes care to ensure that its performance audit activities—which can be regarded as a form of evaluation¹⁷—do not overlap or duplicate those of departments, and departments and DoF avoid duplicating ANAO’s audit activities when planning their own evaluation priorities.¹⁸

The impact of the ANAO’s activities has been felt in several ways. First, it has focused attention on evaluation as a legitimate and important area for senior management attention in departments.

A different impact was felt in the earlier years, when the ANAO pursued its performance audits into evaluation (and into program administration more generally) with a “gotcha”, fault-finding zeal. The value-added of such an approach is highly doubtful as it strongly discouraged the ‘victim’ departments from ownership of the audit findings. The resistance of these departments to accepting the ANAO findings was often evident in their formal, published responses to ANAO reports.

A “gotcha” approach may have satisfied a narrow interpretation of the accountability function of an audit office, particularly in its reporting to Parliament, but it undermined the potential value-added contribution which a considered performance audit could provide to a line department’s future management of a program.

In more recent years, with a new Auditor-General and a different audit philosophy in the Office, there has been a much stronger emphasis on finding ways to help departments improve their performance. A high priority has also been attached to the identification and sharing of good practices, and the ANAO has been active in disseminating these among departments.

Other Factors

Having a government-wide evaluation effort, involving all departments, has proved helpful in developing a general climate of expectation that evaluation will be conducted and used, and in developing an evaluation community, especially in Canberra. It has also helped to develop a labor market for evaluation skills, including advanced data analysis skills. The labor market includes the growing number of staff with experience in evaluation units, in the various economic research bureaus, and in the national statistical agency.

One expression of this community has been the monthly meetings of the Canberra Evaluation Forum. The meetings have been organized by a steering group of departments, with DoF support, and each meeting involves several speakers and discussants of topical evaluation issues. About 100 participants attend each month.

The sharing of insights and good practices through these meetings has strongly encouraged networking. There have been several special-interest conferences and seminars on particular evaluation issues organized by DoF and others, on issues such as the evaluation of policy advice and evaluation/audit links.

A feature of the Australian scene has been the frequent availability of commercially-organized, for-profit conferences on evaluation and other performance measurement issues, and on public sector reform issues more broadly. Various departments including DoF work collaboratively with conference organizers to identify topical issues and provide speakers. The conferences allow an opportunity for federal public servants to be exposed to evaluation issues in state governments, local government and the private sector, and academia. However, the contribution of academia to ECD in the federal government has been more limited than might have been expected.

The role of Parliament has not lived up to the ambitious aims of the designers of the evaluation strategy, who viewed accountability to Parliament as one of the foundations of the public sector reforms, including the evaluation strategy. In practice, Parliament has generally possessed neither the infrastructure resources nor the perspective to focus on the insights into program performance which evaluation findings can offer. While Parliament exercises general oversight, including oversight of annual appropriations, it has provided little scrutiny of strategic issues of performance, preferring instead to focus on administrative errors which might embarrass the government.

But there have been some notable exceptions to the narrow focus and interests of Parliament, and these have involved parliamentary committees inquiring into issues such as the Financial Management Improvement Program and service quality.¹⁹ These have been useful in emphasizing performance issues, such as the impact of government programs on their ultimate clients, to departments and to public servants more generally.

A number of success factors are often taken for granted in Australia, which become evident when making cross-national comparisons, particularly with developing countries which do not possess a well-developed public service infrastructure:

- strong institutional and human capacity in the public sector;
- well-developed management capacity;
- public service managers with a reputation for integrity, honesty and impartial advice;
- a well-developed budget management system, and accounting standards and systems;
- a tradition of transparency and accountability in the conduct of government business; and
- a credible and legitimate political executive.

Current Developments and Prospects

The government, elected in early 1996, expressed considerable unhappiness with the federal public service, and considers it to be rule-bound and caught up in red tape. The government has a strong ideological preference for the private sector, and appears to regard it as being inherently more efficient than the public sector. It has noted with dismay a major evaluation which has shown public service administrative efficiency to be lagging considerably that of the private sector, particularly in personnel practices, and this comparison has strengthened its resolve to develop a smaller, more efficient public sector.²⁰ The government has, therefore, embarked on a wave of major public sector management reform.

The new Cabinet has directed that Ministers and their departments review the nature and extent of continuing need for existing government programs; and it has expressed the expectation that competitive tendering and contracting processes be applied by departments to their programs wherever possible—this could result in considerable outsourcing of the delivery of government programs in coming years. These types of review require a close scrutiny of performance to be successful. In particular, the application of competitive tendering and contracting necessitates a clear understanding of program objectives, followed by *ex ante* assessments of the performance of alternative tenders (in-house and external). Once contracts have been let, there is a need for ongoing performance scrutiny, and at the completion of contracts a review of past performance; this information then feeds back into decisions about the next round of contracts to be put out to tender. Evaluation of performance is central to all these activities.

The government's initiatives are part of a strong push towards commercialization and the private sector delivery of public services, and have already resulted in a significant reduction in the number of public servants, with further larger reductions in prospect. Under some authoritative scenarios, the size of the public service in a decade could be only a fraction of its recent levels.

The government is taking steps to achieve sharper accountability for public service managers, together with fewer centralized controls and fewer formal requirements, partly via a strongly devolutionary approach. Departmental CEOs, who are on employment contracts, will be required to perform to high standards. These expectations have recently been made explicit in new legislation on financial management and accountability, and will increase pressure on CEOs to ensure high standards of corporate governance.²² This may also create both the scope and a requirement for the ANAO to take a greater quality assurance role in performance measurement and reporting than in the past.

Service delivery agencies will be required to set explicit customer service standards, with actual performance being reported publicly, including to Parliament. The focus on performance will be further enhanced by the government's decision to adopt accrual accounting—this will facilitate scrutiny and benchmark comparisons of departmental costs and performance. Changes to output/outcomes reporting are also in prospect, and these will seek to marry the output specification and focus of governments such as those of New Zealand and several of the Australian states and territories, with the outcomes and performance focus of the federal Australian government. This development would invite closer scrutiny of departments' planned and actual performance.

Collectively, the latest wave of reforms are likely to result in considerable changes in performance management, accountability and reporting. For these reforms to be successful, however, there will need to be a high level of scrutiny of departmental and CEO performance to further strengthen management incentives.

This environment helps explain and put in context Cabinet's recent agreement to the replacement of the formal requirements of the evaluation strategy by a principles-based approach. This emphasizes the uses of evaluation and other performance information for performance management purposes, including links with corporate and business planning and the other reform initiatives now underway. Thus the approach in one sense reaffirms the merits of the Financial Management Improvement Program, and of program budgeting. The new approach continues to emphasize the advantages in planning, conducting, reporting and using evaluation findings, the main difference with the previous evaluation strategy now being the absence of formal requirements.²²

How should the new approach be viewed? If at least part of the success of the requirement-based approach was because it mandated evaluation activity, then there will be some risk with a new approach which intentionally excludes such formal requirements.

But a counter-argument is that the focus of concern should be with outcomes, not with processes to achieve them. If the public sector management framework provides sufficient incentives to achieve a strong focus on performance and outcomes, then this provides support to devolutionary approaches which provide management with the autonomy to achieve this performance in any manner it chooses. In a government where performance measurement has been strengthened, and there is greater accountability for results, there is scope to provide departments with greater autonomy and flexibility.²³

The new approach to evaluation accepts performance measurement as an integral part

of performance management²⁴—reflecting a philosophy that if the environment of public sector governance is strongly conducive to evaluation being done and used, then that will happen. Thus the emerging public sector environment would be expected to be even more encouraging of evaluation than in previous years. But this expectation might mirror a similar but erroneous one in the early 1980s, when it was assumed that if the structural framework of public service management was “correct,” then an evaluative culture would almost automatically follow.

The impact of the new reforms on the culture and management of the public service will partly depend on progress already achieved since the early 1980s. To the extent that an evaluation culture—including management commitment to review and learn from past performance, and an evaluation infrastructure—has already been achieved, this will enhance the speed and widen the extent of impact of the new reforms.

Conclusions from the Australian Experience

The requirements-based, formal evaluation strategy in Australia constituted a model of central force-feeding to ensure the planning, conduct, quality and use of evaluation. It reflected the belief that managers would not do this if left to their own devices. Formal requirements and a whole-of-government approach helped to kick-start the process and achieve significant momentum. However, it is not formal rules and requirements that determine the extent of the conduct and use of evaluation findings—it is the commitment of individuals and their organizations, and the nature of their understanding and motivation.

The recent move to a principles-based approach reflects the evolution of governance arrangements and the particular circumstances of the Australian scene. This evolution represents a migration from tight to loose controls over departments.

The Australian experience provides a wealth of lessons. However, although it is possible to identify features of that system which have contributed to the substantial success achieved so far, this does not necessarily mean that these success factors are *preconditions* for success.

Some key success factors have been:

- macroeconomic pressures which have led to tight budgets and a priority on finding ways of achieving better value-for-money;
- a powerful department (DoF) willing to champion evaluation, react to changing circumstances and identify new opportunities for influence and development;
- the sustained commitment over a decade of the government, and especially of its

-
- main champion (DoF), to the evaluation strategy;
 - having a second central agency (the ANAO) willing to prompt and prod departments to focus on evaluation and performance management more broadly;
 - the creation of an explicit evaluation strategy with formal evaluation requirements;
 - a whole-of-government strategy to help achieve and maintain momentum in evaluation capacity development;
 - a budget agency (DoF) able to link evaluation into both the budget process and into public sector management reforms;
 - a budget system which makes transparent the costs, and the pros and cons of competing policies;
 - the implementation of related public sector management reforms, particularly portfolio budgeting, which provide substantial autonomy to line managers and which emphasize bottom-line results and outcomes—these have provided powerful incentives to managers;
 - the support of Cabinet and a number of key Ministers, and the emphasis they have placed on having evaluation findings available to assist their decisionmaking;
 - the priority given to evaluation in several large and important line departments, which has helped to highlight and legitimize it; and
 - the devolutionary approach to evaluation within line departments, which has helped to mainstream evaluation as a core activity, together with internal quality assurance processes.

Some Implications for the World Bank

A challenge for the Bank is to foster ECD in developing countries. So what lessons can be drawn from the Bank's own experience in fostering ECD, and to what extent are these similar to the Australian experience?

The report of the Bank's 1994 Task Force on ECD diagnosed a number of problems in developing and embedding evaluation capacity in developing countries.²⁵ These included:

- a lack of genuine demand and ownership by politicians and officials—demand was judged by the Bank to be the key precondition;
- lack of a culture of accountability—often reflecting problems of ethics and corruption;
- lack of necessary evaluation, accounting and auditing skills. Overcoming this often requires broader institutional development and capacity-building;
- poor quality of financial and other performance information, and of the accounting and auditing standards and systems required to provide and make use of such information;

-
- lack of evaluation feedback mechanisms into decisionmaking processes; and
 - the need for ECD efforts to have a minimum critical mass to succeed.

Lack of real demand has often been identified as the crucial deficiency. But what does this mean? It would be unrealistic to expect wholehearted support across an entire public service in any country—developing or developed—or from all government ministers.

One strategy could be to foster evaluation in only one or two departments, and to hope that the demonstration effect would cause other departments to progressively adopt performance measurement approaches too. This “enclave” approach could work—good practice examples are invaluable in demonstrating the potential of evaluation. But efforts could be set back considerably whenever there are changes in senior departmental management, or when policy or funding crises cause a diversion of focus. ECD efforts in some developing countries have stalled for exactly these reasons.²⁶ Experience in Australian departments shows that performance measurement initiatives, and other public sector reform initiatives, can be the first to be postponed—sometimes indefinitely—when external pressures occur.

In contrast, a government-wide approach offers the potential to generate sufficient momentum to sustain progress in all departments. Even departments which suffer some external setbacks can be induced to keep up with their peers if there is sufficient government-wide pressure and momentum.

A government-wide approach requires at least one strong, lead department or agency—perhaps ideally two. Central agencies, such as Ministries of Finance or Planning, or a National Audit Office, are prime candidates. Issues to consider here when examining possible champion agencies include the depth and sustainability of their commitment, and their ability to prod and support other agencies effectively. An ideal champion could influence both public expenditure management and line management within all other agencies.

Cabinet or government endorsement is a powerful lever, but this should be viewed in context. A government would be unlikely to view evaluation as anything more than a useful tool; the Australian experience is that government ministers often regard evaluation as bureaucratic business—something for officials to focus on. However, if they are advised by their senior officials that evaluations should be commissioned and that this will assist policy formulation and decisionmaking, then Ministers have been happy to endorse them.

While formal Cabinet endorsement of an evaluation strategy has been an important lever, it is necessary to have a strong lead department to champion the strategy with other departments. Unless this happens evaluation will only be paid lip service.

Another important environmental factor is a sense of urgency such as with a budgetary crisis. This can help to persuade officials and their Ministers of the need for a systematic approach to performance measurement.

In the absence of such an environment it should be possible to pursue ECD as part of broader public sector management/governance efforts. The Bank has argued that ECD should be viewed as an integral part of these broader efforts, but evaluation is often seen as a stand-alone activity. There is often insufficient understanding even among PSM reformers that evaluation is an invaluable and necessary support to policy analysis, budgetary resource allocation, and to program and organizational management discrete projects and of ongoing programs. A stronger version of this proposition is that high levels of economy, efficiency and effectiveness are unattainable unless there are sound and integrated systems of performance measurement and management.

There are potential synergies and commonalities between Bank work on ECD and its support for broad governance efforts, including civil service reform, financial reporting and auditing, and anti-corruption efforts. And if developing country governments place increasing emphasis on outsourcing the delivery of government activities, this will provide an additional opportunity for them to move to a more evaluative culture—however, this will also require that they possess sound assessment and contract management skills.

Given that it takes at least a decade to develop a national evaluation system and embed it in a government in a sustainable manner, there are implications for Bank and other development agency support for ECD. Clearly, it is necessary for the Bank to move from a short-term project focus in the technical assistance it provides, towards one which gives strong and enduring support over the long-term.

The position of the World Bank is analogous in some ways to that of the Australian DoF—the Bank as a catalyst and advocate with developing country governments. The lesson from Australia is that a focused approach, with substantive and sustained momentum, is necessary to overcome internal lack of support and to exploit external opportunities.

Bibliography

- ANAO (Australian National Audit Office). 1991. *Implementation of Program Evaluation—Stage 1*. Efficiency Audit Report No. 23, 1990-91. Canberra: Australian Government Publishing Service (AGPS).
- _____. 1991. *Evaluation in Preparation of the Budget*. Efficiency Audit Report No. 13, 1991-92. Canberra: AGPS.
- _____. 1992a. *Program Evaluation in the Departments of Social Security and Primary Industries and Energy*. Efficiency Audit Report No. 26, 1991-92. Canberra: AGPS.
- _____. 1992b. *Auditing Program Evaluation—ANAO Performance Auditing Guide*. Canberra: AGPS.
- _____. 1992c. *Department of the Treasury—Procedures for Managing the Economic Policy Program*. Efficiency Audit Report No. 36, 1991-92. Canberra: AGPS.
- _____. 1993. *Program Evaluation—Strategies, Practices and Impacts—Industry, Technology and Regional Development Portfolio*. Efficiency Audit Report No. 35, 1992-93. Canberra: AGPS.
- _____. 1996. *Performance Information—Department of Employment, Education, Training and Youth Affairs*. Performance Audit Report No. 25, 1995-96. Canberra: AGPS.
- _____. 1997a. *Applying Principles and Practice of Corporate Governance in Budget Funded Agencies*. Canberra: AGPS.
- _____. 1997b. *Program Evaluation in the Australian Public Service*. Performance Audit Report No. 3, 1997-98. Canberra: AGPS.
- ANAO/DoF (Department of Finance). 1996. *Performance Information Principles*. Canberra: ANAO.
- Auditor General of Canada. 1993. *Report of the Auditor General of Canada to the House of Commons. Chapter 8: Program Evaluation in the Federal Government; Chapter 9: Operation of Program Evaluation Units; Chapter 10: The Program Evaluation System—Making it Work*. Ottawa.
- _____. 1996. *Report of the Auditor General of Canada to the House of Commons. Chapter 3: Evaluation in the Federal Government*. Ottawa.
- _____. 1997. *Report of the Auditor General of Canada to the House of Commons. Chapter 5: Reporting Performance in the Expenditure Management System; Chapter 11: Moving Toward Managing for Results*.
- Barrett, Pat, Auditor-General of Australia. 1996a. "Some Thoughts About the Roles, Responsibilities and Future Scope of Auditors-General." Address at the Australian Society of Practising Accountants Annual Research Lecture, Canberra, November 14.
- _____. 1996b. "Performance Standards and Evaluation." Address at the Institute of Public Administration of Australia National Conference, Reshaping the Old: Charting
-

-
- the New—Public Management in the 1990s, Melbourne, November 20-22.
- Commonwealth of Australia. 1996. *1996-97 Budget Statement 3: Overview*. Canberra: AGPS.
- Crossfield, Len and Anne Byrne. 1994. *Review of the Evaluation Function in DEET*. Canberra: Department of Employment, Education and Training, and Department of Finance.
- DEETYA (Department of Employment, Education, Training and Youth Affairs). 1996. *Budget Initiatives*. Canberra: DEETYA.
- _____. 1997. *Enhancing the Effectiveness of Active Labour Market Policies*. Australian presentation to the meeting of OECD Labour Ministers. Canberra: DEETYA.
- Dixon, Geoff. 1993. "Managing Budget Outlays 1983-84 to 1992-93." In Brian Galligan, ed., *Federalism and the Economy: International, National and State Issues*. Canberra: Federalism Research Centre, Australian National University.
- DoF (Department of Finance). 1988. *FMIP Report*. Canberra: AGPS.
- _____. 1991. *Handbook of Cost-Benefit Analysis*. Canberra: DoF.
- _____. 1993. "The Cost of Evaluations: the Findings of a Pilot Study." DoF Working Paper, Canberra.
- _____. 1994a. *Doing Evaluations—a Practical Guide*. Canberra: DoF.
- _____. 1994b. *The Use of Evaluation in the 1994-95 Budget*. Finance Discussion Paper, DoF, Canberra.
- _____. 1995. "Reasons Why Evaluation Should Be Done and Why Finance Should Be Involved." DoF Working Paper, Canberra.
- _____. 1996. *Performance Information Review: Overview of the First Year of the Review*. Canberra: DoF.
- DoF and Public Service Board. 1986. *Evaluating Government Programs—A Handbook*. Canberra: AGPS.
- Duckett, Mary. 1995. *Performance Reporting in Commonwealth Annual Reports*. Report for the Department of Finance. Canberra: DoF.
- Funnell, Sue. 1993. *Effective Reporting in Program Performance Statements*. Study for the Department of Finance. Canberra: DoF.
- Keating, Mike, and Malcolm Holmes. 1990. "Australia's Budgeting and Financial Management Reforms." *Governance: an International Journal of Policy and Administration* 3(2): 168-185.
- Mackay, Keith. 1994. "The Australian Government's Evaluation Strategy: a Perspective From the Center." *Canadian Journal of Program Evaluation* 9(2): 15-30.
- MAB/MIAC (Management Advisory Board/Management Improvement Advisory Committee). 1995. *Achieving Cost Effective Personnel Services*. MAB/MIAC Report No. 18. Canberra: AGPS.
- Parliament of Australia, House of Representatives Standing Committee on Finance and

-
- Public Administration. 1990. *Not Dollars Alone—Review of the Financial Management Improvement Program*. Parliament of Australia, Canberra.
- Parliament of Australia, Senate Finance and Public Administration References Committee. 1995. *Service Delivery: Report from the Committee on Service Delivery by the Australian Public Service*. Parliament of Australia, Canberra.
- President of the Treasury Board. 1995. *Strengthening Government Review—annual report to the Parliament*. Ottawa: Treasury Board Secretariat.
- Reith, Peter, Minister for Industrial Relations and Minister Assisting the Prime Minister for the Public Service. 1996. *Towards a Best Practice Australian Public Service*. Discussion Paper. Canberra: AGPS.
- Schick, Allen. 1996. *The Spirit of Reform: managing the New Zealand state sector in a time of change*. Report prepared for the New Zealand State Services Commission and the Treasury. Wellington: State Services Commission.
- _____. 1998. "Why Most Developing Countries Should Not Try New Zealand Reforms." *The World Bank Research Observer*. 13(1): 123-131.
- Sedgwick, Steve. 1996. "Setting Ourselves for the Future." Address to DoF staff. Canberra: DoF.
- Senator Amanda Vanstone, Minister for Employment, Education, Training and Youth Affairs. 1996. *Reforming Employment Assistance—helping Australians into real jobs*. Canberra: AGPS.
- Task Force on Management Improvement. 1992. *The Australian Public Service Reformed: an Evaluation of a Decade of Management Reform*. MAB/MIAC, Canberra.
- Uhr, John and Keith Mackay, eds. 1996. *Evaluating Policy Advice: Learning From Commonwealth Experience*. Canberra: Federalism Research Centre (Australian National University) and DoF.
- World Bank. 1994. *Evaluation Capacity Development—Report of the Task Force*. World Bank, Washington, D.C.
- _____. 1997. *The State in a Changing World: World Development Report 1997*. New York, N.Y.: Oxford University Press.

Endnotes

1. The three dimensions on which program evaluation can focus are (i) the efficiency of a program's operations (minimizing costs for a given level of output), (ii) its effectiveness in achieving its objectives, and (iii) whether the program's objectives remain consistent with the government's policy priorities, (the appropriateness of the program).
2. There was only modest success with the requirement that Ministers' new policy proposals include an evaluation plan of action that would be undertaken if the proposal was accepted. Feedback from portfolios indicated that this requirement was onerous for portfolio managers during the busy budget period. Only about 30% of proposals broadly met this requirement in the 1993-94 budget, for example, although an additional 50% of proposals included a clear undertaking to evaluate the proposal if accepted (DoF 1994b). These percentages were only achieved after considerable prodding by line areas within DoF. In recent years the extent of such prodding (and of departments' willingness to provide such plans in their budget documentation) has fallen off considerably.
3. ANAO 1997b.
4. Ibid.
5. The ANAO (ibid) recently surveyed 20 departments and agencies and found that 11 had reviewed their evaluation activities within the past four years.
6. Ibid.
7. A major review of labor market policies resulted in a particularly heavy emphasis on evaluation findings in that budget (see DoF 1994b).
8. This compares with A\$1060 million in the preceding year (reflecting a composition effect because of small numbers of large savings options).
9. The Australian budgetary system is discussed in World Bank 1997, p. 82.
10. In the absence of evaluation findings, decisions will probably be influenced more by ex ante analysis or anecdotal information and case study examples.
11. Auditor General of Canada 1993.
12. A 1992 evaluation of public sector management reform in Australia concluded that "there is widespread acceptance of the importance of evaluation." But it went on to note that "the bulk of (senior executive managers) state that it is using evaluation information only sometimes or infrequently during the conduct of their job. This suggests that information generated by evaluations is not yet a key element in program management." (Task Force on Management Improvement 1992, pp. 378 and 379.)
13. See ANAO 1997b.
14. DoF has provided introductory evaluation training to over 3000 public servants since 1991.
15. The Department of Health encourages quality evaluations through: selection of experienced officers to manage the evaluation; involvement of internal and external stakeholders; ensuring that technical advisory panels are available to help assess the work of consultants; having steering groups available to help manage consultants; and ensuring that sufficient resources are available for the evaluation.

-
16. See, for example, Auditor General of Canada 1996 and 1997.
 17. A broad definition of evaluation is used in Australia. It includes program evaluation, project evaluation (principally cost-benefit analysis), efficiency and performance audits, and formal policy reviews.
 18. Auditor-General 1996; ANAO 1997b.
 19. Parliament of Australia 1990, 1995.
 20. Reith 1996; MAB/MIAC 1996. The government has favorably acknowledged the existence of “a system of performance management and program budgeting based upon an explicit evaluation of outcomes.” (Reith 1996, p.ix)
 21. The new legislation was passed by Parliament in October 1997. See also ANAO 1997a.
 22. And yet even some of the formal requirements will remain, but in a different guise: departments will continue to be issued guidelines for reporting to Parliament about their annual appropriations, and these will now include the need for summary statements of evaluation intentions. In addition, guidelines for preparation of departments’ annual reports will note the need to report on past performance, including results as shown by completed evaluations and other performance information.
 23. This approach is advocated by the World Bank in its recent World Development Report (World Bank 1997).
 24. In contrast, it could be argued that the earlier requirements-based approach had been added on to the then-existing suite of reform initiatives in the 1980s almost as an afterthought. If that is a fair interpretation, it was a very effective afterthought.
 25. World Bank 1994.
 26. Ibid.

Evaluation in the Federal Government of Canada

Stan Divorski¹

Evaluation has been a requirement in the Canadian federal government since about 1977. The Office of the Auditor General of Canada has examined the implementation and success of those efforts, most recently reporting the results of its examinations in 1993 and 1996.

The Auditor General has argued that program evaluation has significant potential, and that evaluation can provide information to support resource allocation decisions, to help Canadians determine the value obtained from tax dollars, and help public servants manage for results and take responsibility for results. But this requires more than a set of social indicators or statistics on the outputs of government departments. From time to time, systematic disciplined studies are needed which examine program accomplishments. These studies are evaluations.

In Canada, federal evaluation has been governed by administrative policy. Unlike the U.S., where the system of performance measurement is governed by legislation, Canada has been trying to accomplish the same results focus through administrative policy. Current federal policy requires federal departments to evaluate their key policies and programs. Initially, policy required—as with Australia—that all programs be evaluated on a cyclical basis. The length of the cycle has varied over time between three and 7 years. That policy was amended to require that all programs be considered for evaluation over a fixed cycle (current policy simply requires that key policies and programs be evaluated strategically and cost-effectively).

The accountability to Parliament for departments is through elected officials appointed by the Prime Minister to be the head of that agency, as one of the members of Cabinet—usually with the title of Minister. Current policy makes the deputy heads of departments (Deputy Ministers) responsible for the performance of programs, for appointing an evaluation manager independent of the activities being evaluated, and for using the results of evaluations. The deputy head is the first level of public servant, an appointed official. He or she is responsible in effect for program evaluation.

The earlier policy specified that the deputy head was the client for evaluation. That policy has now been amended to require that evaluation meet the needs of the deputy and of senior program management, an important distinction. The Treasury Board Secretariat (roughly equivalent in function to the US Office of Management and Budget)

is responsible for administrative policy. The Secretariat which promulgated this policy is also responsible for monitoring its implementation.

The policy defines three core evaluation issues. These issues are meant to be considered in the planning of all evaluations. The first is relevance. This speaks to the question of whether the program is consistent with existing government and departmental priorities, and whether it continues to address an actual need. The second is the success of a program in meeting objectives established for it. The third is the cost effectiveness of the program, that is, whether means used by the program to achieve its objective are appropriate and efficient compared to alternatives.

The 1993 and 1996 audits had a number of similar findings. First, both audits found examples in which program evaluation had been clearly capable of meeting its potential. Some evaluations, identified significant cost savings. An evaluation of Canadian Coast Guard search and rescue activities found that the six largest vessels (which cost more than half of the entire fleet to operate) accounted for less than 10 percent of the lives saved. This finding led to recommendations to decommission or reassign these vessels, with a potential for cost savings in the order of \$30 million.

Evaluations have also contributed to increased cost effectiveness and to improved services. Human Resources Development Canada, evaluated its job training programs and found that privately-sponsored projects were more successful than publicly-sponsored ones.

Evaluations have also led to identifying the consequences of major policy changes. A Finance Department evaluation of Canada's tobacco tax found that a ten percent increase in the price of cigarettes would lead to approximately a nine percent decrease in consumption (although it was expected that consumption would rebound). This evaluation also found that a 10% increase in the tobacco tax would lead to a ten percent increase in exported tobacco. These examples illustrate government activities where evaluations have made significant contributions to the understanding of policies and programs and to cost effectiveness.

There were few examples of major program changes or large cost savings that could be clearly demonstrated to result from evaluations. Rather, evaluations tended to meet the needs of immediate program managers and focus on smaller program components and on operational issues, such as streamlining activities, improving relationships between headquarters and regions, and increasing smoothness of delivery to clients. In general, important accountability issues were not addressed. There was little attention to the difficult overall effectiveness questions with regard to larger programs. In general, the coverage of programs was low.

The emphasis was not on issues of overall effectiveness for government decisionmaking or for broader accountability to Parliament.

Evaluations had been successful in producing changes when they had addressed operational issues close to the needs of program managers. Evaluations were more likely to be used in departments which had long-range planning for evaluations, particularly where there was senior management follow-up of the implementation of evaluation recommendations.

Perhaps it is not surprising that in Canada evaluation has tended to focus on the needs of program managers rather than larger accountability issues. John Mayne, Donald Lemaire and I explored the implications of locating program evaluation in different parts of a government. In looking at a number of countries, we found that when the responsibility for decisionmaking about evaluations is located close to program management, the evaluations tend to focus on operational effectiveness issues and tend to limit the examination of impact issues. (They would probably not look at all at continued relevance, which is seen more as a political issue).

Further up the government structure from line management, evaluation would tend to focus less on operational issues, and look more at impact questions. Only at the legislative levels or the very senior corporate government level, would people look at relevance issues.

There have been a number of changes in Canada since 1993 with significant implications for the function of evaluation. First, the government conducted a major restructuring. It reduced the number of departments, and it significantly restructured a number of the remaining ones. This created enormous problems and challenges for planning and conducting evaluations because evaluations were suddenly aiming at moving targets. It also made it difficult to identify who was accountable for what programs and what activities.

Government has started to increase the use of third-party program delivery. This complicates evaluations again because of the challenges with regard to accountability. When programs are devolved to other levels of government, or to the private sector (or to a nongovernmental sector) those actually delivering the program may not be directly accountable to Parliament (although the public servants who have delegated the responsibilities still are). Thus, accountability influences what issues evaluations can look at and how evaluations are conducted.

The government conducted a number of major policy and program reviews that looked at the question of program relevance, and the extent to which programs were in accordance with government priorities and continued to meet the needs of the day. Canada

has also implemented changes to its budgeting system (the expenditure management system). Canadian government departments must now set performance expectations and report these annually to Parliament. They are expected to report annually on prior years' achievement of or progress towards these expectations. The policy also requires that departments report to Parliament annually on the results and findings of major reviews of government activities, including evaluations. This creates the potential to fit evaluation into the expenditure management process.

The policy governing evaluation and internal audit was modified in 1994 to give greater recognition to management's responsibility for examining performance, by putting greater emphasis on management-led reviews. This led to some confusion. What exactly was the nature of evaluation as compared to reviews? To what extent was evaluation required? and What standards applied to the measurement of effectiveness?

The Standing Committee on Public Accounts of the Canadian Parliament held hearings on the 1993 audit. The Committee's report called for stronger central leadership in planning, performing, and reporting the results of evaluation. The Treasury Board Secretariat implemented some changes in response to these recommendations. The President of the Treasury Board tabled a report which discussed government's progress in improving the reporting of results in the federal government and has reported annually on similar topics since then.

The 1996 audit found that the modifications to the expenditure management system, the requirements for performance reporting, the links to evaluation and the specific reference to evaluation together constituted important progress in linking evaluation into the budget process. There is some hope that this will lead evaluations to provide more information to support accountability to Parliament. The President of the Treasury Board's Report on Review looked promising in this regard. However, there was still no systematic presentation of government evaluation priorities, nor was there a clear link between overall government priorities and departmental evaluation plans.

In Canada we find a changing environment in which various factors, such as the attempt to place a greater focus on results-based management and monitoring, have changed how evaluation links into the budget management process.

1. The author was the responsible auditor for the May 1996 Report of the Auditor General of Canada on Evaluation in the Federal Government. However, the views and opinions put forward in this article are those of the author and no endorsement by the Office of the Auditor General is intended or should be inferred.

Comments

The Role of Evaluation in Public Administration and Political Philosophy—Some Observations on Developed Country Experiences

Ray C. Rist

What can we learn from these three mature national evaluation systems? Surely there are lessons for political systems and public administration; and then for political philosophy and how it bears on this issue of ECD.

First let us go to the core—the question of capacity. The first implication for public administration and building systems of governance is the need for capacity in systems, processes, and personnel. Evaluation, if it is integral to a national system, is part of the political life of a society. The data from evaluation become part of the national discourse. And consequently, the processes need to ensure the trustworthiness and transparency of the evaluation data. And then comes the matter of human capacity. Without people who can produce good, trustworthy data, who are trained to know the distinctions between anecdotal information and more reliable data, who are able to do good financial assessments, who can do sophisticated (or even rudimentary) cost-benefit work, real evaluation becomes impossible. The capacity issue at that level is important in considering how evaluation can contribute to effective public administration.

The second issue for public administration “Where do you place the capacity?” If there is only some evaluation capacity in some parts of the system for some processes, where do you put that capacity? To what degree should you centralize versus decentralize? Do you put capacity in a central agency, or do you put it in a number of decentralized parts of the national system? The US has tried to decentralize out to the departments. Australia has a strong champion of centralization in the Ministry of Finance. So the issue is when to centralize, and when to decentralize. The Canadian evaluation system began as a centralized approach and has evolved into something more decentralized. Thus we have three countries with three different strategies. For each, the location is critical.

A second element of location is also important. That is the issue of whether capacity is emphasized in the executive branch or in the legislative branch. Do you put it in Parliament or in the government? To gain access to it, do you go to the central government ministries? To Parliament? or in the United States, to the US Congress?

The third location issue is whether the evaluation function is internal or external. To what degree should an evaluation office try to handle everything? To what degree should one use third parties? Using third party external evaluators means that whoever is responsible for evaluation in an organization is monitoring, managing and leading the evaluation efforts of people from the outside. Going to a third party for the conduct of evaluations brings some problems: span of control; design questions; quality assurance questions; and whether the external evaluators are attuned to the political issues. To summarize the key questions regarding where you locate the evaluation function: Do you centralize/decentralize?; Do you use it in the legislative or executive branches?; and Do you make it an internal or an external function?

A final question for public administration concerns the degree to which capacity should be operational versus strategic. In Canada, evaluation capacity is seen to be emphasized more operationally than strategically—a not surprising outcome when the choice is made to decentralize. The further from the central policymaking and strategic parts of the government, the more likely it is that local stakeholders will ask questions important to themselves. The looming question for the US—when the GPRA is institutionalized—will be how to manage and use both the operational and strategic evaluation data as they come in.

And now for political philosophy and ECD. First, let us define capacity. Is it capacity for learning? Capacity for accountability? Capacity for budget decisionmaking? How utilitarian is it versus how strategic? How much is it geared towards using evaluation for the learning function, the accountability, or the budget function? There are no easy answers to the trade-offs posed here. But this question needs to be thought through when considering evaluation capacity-building. If pursued simultaneously, these objectives can work at cross-purposes to each other—“learning” means that you take a positive, forward-looking perspective; accountability is backward-looking and often focuses on finding fault and apportioning blame; while “budgeting” emphasizes costs and benefits relative to a defined bottom line. All three issues have got to be taken into account, because how the managerial system resolves these tensions sends strong signals. The reality of how evaluation is used as a form of governance influences the behaviors of managers. The questions then of “use for what”? as well as of “capacity for what”? are important.

The second issue has to do with capacity in the policy cycle. Do we want it for formulation, implementation, and accountability, individually or all at once? Can we put a full policy cycle evaluation system in place? What happens if we cannot put it all in place? Where do we want to begin? With enhancing people’s capacity to do policy formulation? Would we rather have

the emphasis on formulation—when we do it and where we do it—or is it more important to think about how we do it, (the implementation issues?) Or is it even more important to think about the accountability issues? With a “big bang” approach to ECD, can we have all three at once? Perhaps the Australian model of a “whole of government” approach would allow evaluation to change government, and the entire political process. I think the realistic view would see this as a high-risk proposition.

So we have an issue of timing, we have an issue of targeting, and we clearly have an issue of what is considered a minimal baseline from which to begin. If you cannot evaluate the whole of government, or you choose not to, what is enough? These are questions where we need good research. We need case studies. We need to understand what you will get if you use a “whole of government” strategy. Alternatively, if you try to begin in a targeted fashion and start with a Ministry of Finance or an Auditor General’s office, what do you get? If you try to do it operationally versus strategically, what do you get? These are empirical issues. Unfortunately we lack the data to answer these questions.

Today we are in an experimenting mode. There are no orthodoxies, and the degree to which we can share experiences and learn, will determine the effectiveness of evaluation in the future.

Comments

Frans Leeuw

In general I agree with the selection of crucial determinants for the successful establishment of evaluation activities in developed countries. I agree that one crucial variable is the existence of one or two central government ministries. A second crucial variable is the compelling force of evidence produced by independent studies (for example by the audit office) showing that indeed there is no evaluation structure. Variable number three, the issue of evaluation guidelines, is important too. In the European context, the role of the national audit office, whether it is a court or the auditor general, can also be a crucial variable.

The articulation of goals and missions, is similarly important in the European context. The level of preciseness and testability is something that we definitely could learn from. Our policy goals often are not as specific as the American cases – in Europe there are no such major laws on performance or monitoring performance assessments. In the Netherlands, for example, this is done on the lower level of rules and regulations. Of course, there are laws on the financial system, and governments have to be active in setting priorities and doing evaluations.

In the Netherlands, a small country with some 16 million inhabitants, the empirical evidence shows that until the late '80s not many evaluations took place. The level of government information and what was known about subsidies, vouchers, levies, inspectorates, loans, public information campaigns, and many other tools of government was not there. That is compelling evidence indicating a need for evaluation—the government did not know what was going on. However, on its own, the pressure of a budget deficit was not enough to get the evaluation machinery going. But it triggered a government-wide investigation which asked questions like: Which managers and which organizations are involved in evaluations? How much money is spent on it? How many studies have been reported? and What is their impact? That government-wide audit looked back over a three-year period (1988-1990) and was published in 1991. The result was tremendous. It led, within six months, to the government bringing out a seven-page, written statement on why evaluation is necessary, what are basic (methodological) criteria for it, why it should be done, and who should do it. This was the first big step forward. This government-wide study, carried out by the National Audit Office, on top of the previous findings, caused the government to create the momentum for evaluation. It multiplied the number of evaluations being conducted.

A couple of years later, the National Audit Office looked into Quangos, the quasi-autonomous non-governmental organizations, that are semi-independent but are largely

financed through the annual central government budget (these Quangos have a total employment greater than that of central government). After the review, which showed that evaluation was still an infant industry, there was a huge expansion of evaluation capacity within the Quangos.

Evaluation now, in 1998, is much more big business. Between 500 and 600 evaluation studies are being produced annually by the central government ministries. There are growing numbers of inspectorates and supervisory/regulatory offices. Almost every month, a new inspectorate or supervisory agency is created. The most recent was on child care. This has a very important task to evaluate the quality of child care. Given our demography, we soon will have an elder care inspectorate.

Here are five points which may be of interest to those wanting to implement evaluation capacity in developing countries. The first is that the evaluation capacity, now in place in a number of western European countries, has an almost natural focus on ex post evaluations. But this is at the expense of ex ante evaluations, focusing on articulating the logic behind programs that have to be established, that have to be implemented, or the logic behind tools of government that have to be implemented. That logic should answer the question: How valid are the behavioral, the social, and the economic assumptions underlying programs? But these questions are not often actually tackled. One reason for this is that it is dangerous for evaluators to question, to debunk, to demystify the program. When an evaluator starts to show that certain assumptions are not valid, or are truisms or are received wisdoms, that is probably dangerous. So while an evaluation capacity infrastructure is probably acceptable, it may be even more important to look into the ex ante type of assessment than to look into setting up an evaluation infrastructure.

The second forgotten factor could be what is called the performance paradox. Is it true that when organizations invest heavily in evaluation structures, in performance measurement techniques, and similar things, that these organizations are more effective? The answer is: not always. Indeed, there is a performance measurement industry already in the United States. And now the same evaluation industry is growing in Europe. So the performance paradox has to be taken into account. It is an unintended consequence of a full-grown evaluation and, in particular, audit infrastructure.

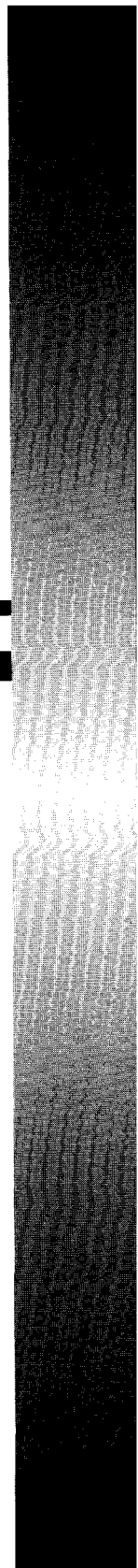
Number three—organizational culture. Apart from structures, we have culture. You have commitment, you have social capital, you have trust, but I know of only a few western European audit and evaluation studies that look into how to measure, for example, trust and effectiveness. A similar question is How to measure social capital? One of the EDI

programs, on helping develop and implement anti-corruption programs, tries to do that. The program uses social capital as one of the mechanisms.

Finally, two other forgotten variables. Number four—ossification. This means organizational paralysis or lack of innovation brought about by too many or too rigid systems of measurement. There are examples in the United Kingdom of ossification. Number five is manualization—the consequence of a tick and flick approach. The more activities that have to be checked, the more likely it is that the evaluators will use the checklist approach and manualize everything. That can have unintended side-effects, too.

A healthy skepticism seems the best approach. We should certainly be skeptical but not cynical (although we must be alert to the unintended, even perverse, consequences of fostering an unquestioning commitment to evaluation). Objectivity will lead us to develop an effective and productive evaluation capacity infrastructure.

PART 4:
Experience of
Developing Countries



Lessons from Chile

Mario Marcel

This session is supposed to tell success stories on evaluation in the public sector of developing countries. In my country it is still too early to label evaluation as a success, except in a rather general sense. But this is not a minor achievement because the experience with public sector reform in many Latin American countries exhibits many false starts. For example, Bolivia has one of the most modern pieces of legislation in public sector management (the SAFCO law). But the problem is that the law is not complied with. In order to manage information on public sector financial management, Bolivia has a large system (funded mostly from donors), but this system is so heavy and so slow that the Ministry of Finance has to run a parallel program to provide information on financial programming for macroeconomic purposes.

In Chile, the main progress that the government has made is coordinating a series of management tools and processes already standing on their feet. One key facilitator of this process has been the incremental approach taken by the Chilean government, avoiding the idea of a grand reform that would solve all the problems at the same time. The background of the Chilean experience reveals how this choice evolved.

It is usual to attribute most of the progress made in Chilean reform to the military government, but public management is an area in which not much progress was actually achieved during that time. There was a lot of rolling back of the functions of the state, a lot of privatization, reforms in social security and so on. But as far as public management is concerned not much happened. During the 1980s there was a huge fiscal adjustment effort which strongly affected the public sector. Public wages fell about 40 percent, and the operational resources of government agencies were dramatically run down. Even decentralization did little to improve public management since the transfer of responsibilities from the central government to municipalities was not accompanied by a similar effort to strengthen the managerial capabilities of local governments.

The figures on public employment in Chile show that it is relatively small compared to the averages for most Latin American countries. But it is equally true that the Chilean public sector was still rule-driven, highly bureaucratic and insulated from civil society.

At the start of the first democratic government, in 1990, the incoming authorities were hardly aware of this. They thought that because there had been a lot of reforms in public policies, public sector managerial practices had also improved. However, during its first

years the democratic government concerned itself with reforms related to decentralization and with a few initiatives in the area of administrative reform (which did not get through Parliament).

At the same time, however, there was a growing concern within the administration about what was happening with the management of public resources. In the Finance Ministry, there was an increasing view that scarce resources were being allocated to sectors without getting reasonable results. That was clearly the case of the health sector which doubled its budget in six years, but without substantially improving its results. By the end of the first democratic administration, the Budget Directorate of the Ministry of Finance launched a pilot scheme that focused on how to improve management within government agencies. That pilot scheme promoted strategic planning exercises in government agencies seeking to identify their mission, their main objectives, their main products and their clients. Agencies could then develop projects for improving the quality of management and build management information systems to provide consistent information for setting benchmarks and targets for performance.

That initiative lasted a couple of years. By 1994, however, a new government was more motivated to improve performance in the public sector. To reflect that concern, a joint ministerial committee was formed by the Ministry of the Interior, the Ministry of the Presidency, and the Ministry of Finance. This committee was to coordinate and promote new initiatives on public management.

The current administration has undertaken a series of initiatives to improve the effectiveness and efficiency of public sector management. These have ranged from the endorsement of agency-specific modernization agreements to the widespread introduction of performance-based pay. These initiatives have had varying success.

Looking at what the Chilean government has done on evaluation and other performance assessment in the public sector, we see two major initiatives. The first is the development of a system of performance indicators. By 1994, the Budget Directorate took the initiative to request information and indicators on performance from government. That was a voluntary scheme and 26 out of more than 100 agencies came up with, in total, about 100 performance indicators. Since then, the system has been evolving rapidly.

There has been an important increase in coverage, growing from 26 to nearly 70 agencies in the past four years. There has also been an important increase in the number of indicators (from 100 to nearly 300 in 1998).

There has also been a substantial diversification of indicators. At the very beginning, we had indicators of “effectiveness”, most of which were really only output indicators. But now, after four years, there has been a significant increase in the number of indicators related to the quality of service, efficiency, and economy. The same goes for the sectors covered by the system of performance indicators. At the beginning, we had very few agencies in the system, most of which concentrated on economic functions like inland revenue, and customs. Now more and more agencies from the social area and infrastructure are also providing performance indicators.

The quality of performance indicators has improved considerably. The center has provided guidance and quality control and as the system has evolved, the Budget Directorate has become more selective in the indicators included in the set of information handed over with the budget to Parliament.

The development of the system now allows the quality of the databases that generate information on performance to be tested. In some cases, failure to achieve stated goals has been due to inaccurate information or unrealistic targets: databases and target-setting have been subsequently improved.

These indicators and the performance against targets have been systematically reported to Parliament and to the public—an important feature of our experience. Being able to use indicators not just for people that work within government—people that meet each other all the time—but also with people who are looking at the public sector from the outside, amounts to a major change in organizational culture, and this usually causes changes in behavior.

But the main benefit of performance indicators is that they have enabled a change of focus within government from rules and procedures to performance. For instance, performance indicators have made performance increasingly relevant for discussing budgetary requests and salary increases in the public sector. In some cases performance agreements have been reached between the Center and executive agencies.

Some pioneering agencies those with greater capabilities have also been able to move in the direction of what has been labeled the Citizens’ Charter, establishing standards for the quality of service to the public and offering a compensation mechanism when standards are not reached.

There has been a request for more consistency on performance reporting and account-

ability. And now Annual Performance Reports are issued by each agency as a part of the overall reporting on the state of the nation. There has also been a request for in-depth evaluations following the view that performance indicators do not provide the full picture of what is happening in an agency.

The second component of the Chilean system of government performance assessment is program evaluation. Such a system evolved from the view that performance indicators enable us only to monitor whether an agency's performance is improving or worsening, but not whether its' programs are good enough or sufficiently justified by their performance. This has to do with the difficulty of measuring outcomes. Government outcomes are very difficult to monitor day-to-day, not only because of measurement problems but also because outcomes are subject to influences outside the control of the agencies in charge of government programs. In Chile, we have dealt with outcomes by carrying out in-depth evaluations.

The aim of the program evaluation system as developed in Chile has been to assess the cost-effectiveness of government programs in a meaningful and effective way. This means addressing the need to carry out evaluations that are respected by executive agencies and ensuring that evaluation conclusions and recommendations are seriously considered. This is not a minor issue. Cost, timeliness and feedback to policy decisionmaking, should ensure that evaluations are taken into account when decisions are made within the public sector. These concerns led to a political agreement to launch a system that within four years would evaluate all government programs in Chile. This clear target helped to deal with initial resistance from agencies; by setting such a comprehensive scope for the evaluation system, each agency was told: "no matter how good your programs are, every program is going to be evaluated."

The current target is to evaluate 200 programs in four years, with the selection each year taking into account the concerns raised by parliamentarians regarding programs that they would like to see evaluated. Every program is evaluated by a panel of five experts: three from outside the public sector, either from NGOs, universities, or consulting firms; and the remaining two from central departments within government. Every panel has a chairman who acts as a liaison for the effective coordination of the system. Each panel has a counterpart in the corresponding executing unit throughout the evaluation process. The panel has the authority to request information, either statistical data or previous assessments for those programs (as in the case of programs that have been funded by multilateral agencies but which are relevant to the program in question).

A panel may also commission studies in specific areas that it considers relevant for the

task in hand. The panels use the logical framework approach (and this work has received support from the IDB).

The process through which a panel issues its report has been devised to maximize effectiveness. The panel first issues a draft report that must be replied to by the executing unit. The final report should address any relevant issue or concern raised by the latter and include a statement from the executing unit indicating how it is responding to the recommendations issued by the panel. A summary of the final report is sent to the main authorities in the central government and to Parliament. This means that the evaluations are public.

Twenty programs were selected for evaluation in the first year of the system. All evaluations were finished on time and publicized. The target for this year is to evaluate another 40 programs. This year is going to be the real test for this system because it is going to be the first one in which there will be enough time to take the results of evaluations into account for decisionmaking in the budget process.

One lesson that can be drawn from our experience is the usefulness of a gradualistic and evolving approach. The Chilean reform process has involved several initiatives dovetailing with each other, trying to build something coherent and effective, but without concerning this process as a grand plan. This effort has been supported by public institutions that are stronger than their counterparts in some other developing countries.

In the Chilean experience, the development of the system of performance indicators has evolved on a voluntary basis. This has proved very useful in overcoming resistance from within the public sector. The system is not yet consolidated; there is still considerable ground to cover—for instance, regarding performance indicators.

Finally, the relationship between public sector reform and democracy. The openness of the system has been important in the Chilean experience—we have a public sector that has traditionally been closed to the public, and been averse to public scrutiny. That is not only because of the military government. The political structure of Chile is based on a deeply rooted authoritarian tradition. Opening-up information, and shedding light on the results that the public sector is achieving is a radical departure from this tradition. Such a difference does not only relate to the way in which the public sector works, but also in the way in which it relates to citizens. This has very much to do with building citizenship in the country, and this aspect of our experience is of considerable interest and appeal to other countries. Chile has been able to reduce the gaps between the interests of politicians, bureaucrats, and the citizens, and all Chileans should keep working in this direction.

Indonesia's National Evaluation System

Alain Barbarie

Introduction

Experience in Indonesia confirms that success in ECD depends on the awareness, appreciation and common understanding by decisionmakers and evaluation managers of the importance and necessity for evaluation, and on the subsequent commitment of adequate financial and human resources to support a professional, dedicated and effective cadre of evaluators.

In many developing countries evaluators are limited to producing evaluation information, usually for a restricted number of people. But the presence of evaluators must encompass the whole process of ECD, generating relevant and timely evaluation information, and seeing that this information can be readily understood and properly used.

So the challenge to evaluators is to continue development on the supply side of evaluation, while doing more aggressive work on the demand side. Working on the supply side alone, hoping that demand will come after the evaluation function shows it can deliver useful evaluation information, will not work. A focus for evaluation information requires a demanding client, and such clients must be developed concurrently with the production of such information.

ECD in Indonesia

A Steering Committee was established in 1994 by the Government of Indonesia (GOI) to oversee the development of a national strategy and framework for performance evaluation. This Committee (and a supporting Technical Committee) have been meeting for the past four years. Both committees have been

ECD in Indonesia:

Some Milestones

- Steering Committee in 1994
- IDF Grant by the World Bank,
- Diagnostic Study (30 institutions)
- Development of Evaluation Framework
- Training (120 middle managers)
- National Policy on Development Project Performance Evaluation in 1996

assisted by the Bureau for Project Implementation Monitoring in Bappenas (the National Development Planning Agency) and by the World Bank, through an Institutional Development Fund Grant.

A diagnostic study of some thirty government institutions was first carried out in 1994. The study confirmed the absence of any performance evaluation function in GOI. By tapping into the vast amount of knowledge and experience available internationally, Bappenas was then able to quickly adapt this knowledge and develop its own performance evaluation strategy and framework.

The National Policy on Development Project Performance Evaluation, promulgated in December 1996, stipulates that performance evaluation is to become an essential part of the management process of departments. The intention of GOI now is to strengthen its capacity for performance evaluation, and progress towards a more comprehensive approach when evaluating its development activities.

A number of departments proceeded, under the general guidance of Bappenas, with their own capacity-strengthening activities. For example, performance evaluation sub-directorates have now been created or re-organised in the Program Development Directorates of various departments. Other capacity-strengthening activities have taken place in the central planning bureaus of many departments. Efforts are beginning at the regional level in the provincial offices of some departments.

The focus of performance evaluation in GOI is on the results and benefits, not on the procedures and processes which lead to these. The purpose of performance evaluation in GOI will be to derive lessons from experience regarding the results and benefits of government activities, mainly development projects, and to feed these lessons back into the government management process to improve performance and better allocate resources.

Performance evaluation will not be only an ex post exercise. Performance evaluation will be done primarily during implementation, whenever appropriate. Concern about performance evaluation should be present throughout the management process.

Evaluation will focus on two separate tracks. The first will be the formulation of performance indicators for development projects. The second will be the carrying out of performance evaluation studies of development projects, programs, and sectors. Initially, efforts will concentrate on the development of performance indicators. Later a more

comprehensive approach will be developed to include more evaluation studies. Bappenas has the functional responsibility for performance evaluation in GOI, while departments actually carry out the evaluation work. Bappenas assumes that functional responsibility by overseeing and assisting the development, use and overall effectiveness of evaluation in supporting decisionmaking. Bappenas will work closely with departments to develop and use performance indicators for GOI development projects, as part of the national budgetary prioritisation and decisionmaking system.

Underlying the new national policy on performance evaluation is a belief that a better understanding of the accomplishments of government efforts, in particular those of development projects, will

improve the capability of departments and agencies, and of GOI as a whole, to allocate scarce developmental resources more appropriately and to manage them more effectively. The successful implementation of performance evaluation is seen as an institutional development needed to provide better information for decisionmaking.

ECD in Indonesia:

Basic Principles

- Focus on results and benefits
- Integrated management function
- Two tracks:
 - performance indicators
 - evaluation studies
- Bappenas has functional responsibility
- Departments do the evaluation work

Key ECD Success Factors in Indonesia

The importance of the awareness, appreciation and common understanding of evaluation by decisionmakers and managers implies that those responsible for evaluation are responsible for the supply of good evaluation information, and for the creation of a demand by decisionmakers and managers.

By being more active in developing a viable evaluation function for their country, evaluators contribute to public sector reforms. Evaluation should not wait for public sector reform but contribute to it while it is in process.

Success for ECD in Indonesia is attributable to the sustained efforts of a few key senior public servants, and the aggressive stance they took to promote and encourage the supply and the demand for performance evaluation in GOI. These efforts were greatly enhanced by the active participation of the World Bank.

ECD in Indonesia:**Success Factors**

- Clear terminology, common understanding
- Strong direction
- Participatory approach
- Functional responsibility to Bappenas
- Support and participation of the Bank
- Lobbying for and promulgation of a national policy

Key Success Factor #1

Determined efforts to establish, from the very start, a clear terminology and common understanding related to evaluation, and to eliminate many of the misconceptions and misunderstandings about evaluation.

Decisionmakers typically see evaluation in different ways. For many it is the ex ante evaluation found in planning exercises. For others it is evaluation in the context of quality control and assurance practices. Many fail to differentiate between audit and evaluation or between process compliance and results evaluation. Many have encountered monitoring and evaluation (M&E) and see evaluation as being mainly the monitoring of inputs and outputs. Yet others have participated in donor agency sectoral or impact ex post evaluations and associate evaluation with quasi-research major undertakings.

Initially in GOI there were many definitions of evaluation. Many government institutions had evaluation specified in their mandates, but often meaning different things. The Auditor General, the Department of Finance, Bappenas, the departments themselves all have evaluation in their inception decrees. A common definition for performance evaluation had to be agreed so that who would be responsible for evaluation in GOI could be decided.

In developed countries, it is usual to try and situate management systems (like evaluation) within a broader governmental framework and to then relate such systems to each other. This is done to avoid overlaps and inefficiencies, and to ensure complete coverage of the accountability mechanisms.

ECD in Indonesia:**Lessons Learned**

- Develop a common understanding
- Create a core group of supporters
- Use a participatory approach
- Let departments do the evaluation work
- Provide both financial and moral support
- Promulgate a decree on evaluation

Many of these accountability systems differ from those in developed countries; audit, for example, is still a purely financial compliance exercise—it is process-oriented and seeks to ensure compliance. It looks for faults, and the audited party is usually on the defensive.

Linking evaluation to other accountability mechanisms such as auditing may not be the best approach to developing an evaluation function in developing countries where the focus in developing evaluation should be kept narrow, understandable and doable. Essentially, evaluation should be treated as a stand-alone function.

In GOI evaluation is needed in the context of development projects to know, for example, if the irrigation structures built have increased the amount of cultivable land and the quality of such land in a given area. To know if the new farming techniques developed are being used and have increased disposable household incomes. To know if the number of trainee days were given as planned and if the training is actually being used in the field. To know how much money has been disbursed to date on an activity and how many tool kits have been distributed to the appropriate targeted farmers as a result. To know what constitutes success and whether it has been reached.

Evaluation in GOI has to do with inputs, outputs, results, and impacts, as captured in the imagery of a “results spectrum” approach, a simplified logical framework (a logframe approach) which is now well known, understood and accepted. This puts much less emphasis on process, compliance and controls and should, in time, be of considerable assistance to public reforms in Indonesia.

Evaluation, whatever its ultimate and appropriate nature may be for a given developing country, must have a clear set of concepts to which all can agree in principle. Achieving this, as has been done in Indonesia, was a key success factor, and it should be considered an essential early step for any ECD effort. Every subsequent effort by the developing country and associated donors should reinforce the existing evaluation framework.

Lesson Learned

One first step for successful ECD has to be the development of a clear terminology and common understanding of evaluation and its purpose. From there a clear set of concepts and a workable evaluation approach can be

Common Understanding

- Evaluation has many facets. Choose one
- Evaluation must be a stand-alone function
- Develop a suitable evaluation framework
- Aim for a national policy

developed. Once in place, every effort by the developing country and associated donors should reinforce the evaluation framework and support the national policy.

Key Success Factor #2

Considered and forceful actions (that often went against bureaucratic culture) by the champions of ECD in their efforts to promote and encourage performance evaluation in Indonesia, in particular, through their careful selection of participants to committees, meetings, study tours and training.

The participants were chosen with great care according to their intellectual capabilities, current position, influence in their respective organisations, formal training, work experience, and demonstrated interest in government reform.

All too often the selection process for government-wide initiatives including committees, meetings, study tours and training is left to the decisionmakers in the departments and agencies. As a result selection of the people chosen for official functions, like the development of an evaluation function, often fails to reflect the purpose of the activity.

In Indonesia selecting the right people was done from the very start. A series of informal group discussions around a number of key topics related to ECD was held. Some sixty carefully-selected people were involved in four consecutive meetings. Later, the best participants, perhaps half, were retained for a wrap-up meeting. Those same people, who had been screened, were involved in further committees, meetings, study tours and training. These people formed the basis for a core group of evaluation managers.

In developing countries the membership of committees once established tends to shift constantly, incumbents change frequently, and there is also a tendency to send delegates. So to attain continuity it is necessary to form a solid core group of participants. In GOI the core group later acted as leaders and trainers, creating a multiplier effect in the government. Many have since had the opportunity to manage their own departmental evaluation units.

On the negative side evaluation can become a dumping ground for marginal public servants. (This continues to happen even in developed countries). Any activity which is not central to delivery such as research or training, faces this danger. An alert ECD team will guard against this.

Even before the establishment of a core group of evaluation managers, a group of senior officials, the decisionmakers, were invited to be members of a Steering Committee for

evaluation in GOI. This Committee was further supported by a Technical Committee and both benefited from secretarial support from the Bureau for Project Implementation Monitoring in Bappenas, the National Development Planning Agency. The Steering Committee established in 1994 still exists, and has now been confirmed by the Ministerial Decree regarding performance evaluation in GOI. Continuity of membership and leadership has been maintained. As a result, an experienced core group of decisionmakers in GOI now has the responsibility for the further development of performance evaluation.

Strong direction, a clear plan of action and attention to the process of bringing together the right people, resulting in the formation of an experienced core group of decisionmakers and evaluation managers, were all key success factors. When the Ministerial Decree came out in December, 1996, all these “core” people could understand it, explain it to others and use it immediately in their departments. In fact, many of these people had lobbied for the promulgation of the Decree.

Strong Direction

- A plan of action, a vision
- “Champions” of evaluation
- Selection of the right people for ECD
- Creation of a dedicated core group
- Continuity

Lesson Learned

The creation of a core group of strong advocates of evaluation, a critical mass that can later create a multiplier effect in the government, is essential. This requires forceful action by the champions of evaluation to select the right people for the job and to ensure continuity of participation. It is critically important that the appropriate leader take charge of the process and give it direction, focus and continuity.

Key Success Factor #3

A participatory approach taken in adapting the general principles of evaluation to the Indonesian context, and the resulting creation of a core group of committed ECD supporters, both decisionmakers and evaluation managers.

The general principles of evaluation (the OECD set of evaluation principles) are by now well understood and accepted in developing countries. The difficulty for developing countries is in adapting these principles to their own circumstances.

Early during the ECD efforts in Indonesia, a participatory approach was chosen that would bring together the people best suited to assist in the development of performance evaluation. These people would adapt the general principles of evaluation to the Indonesian context.

Developing countries may be handicapped by a scarcity of financial and human resources, the lack of a well-developed public service infrastructure, and the absence of a tradition of transparency and accountability in the conduct of government affairs. The implementation of sound management practices, in particular performance evaluation, under such circumstances becomes a daunting task.

Considering how costly and time-consuming ECD has been in developed countries, despite their considerable pre-existing governmental infrastructure, it is surprising that developing countries are expected to achieve the same results in less time and using considerably fewer resources.

Initially the work of developing a strategy and evaluation framework in GOI was divided into a number of "ECD" themes such as examining the merits of a central vs. decentralised approach, the limitations of human resources capabilities, and the current role of M&E units in departments. Informal working group meetings were held on these topics. Steering Committee meetings involving senior officials were held, often with World Bank staff in attendance.

Full coverage of all the key departments and agencies was a determining factor in membership to these committees and meetings. The "accountability" agencies were all specifically invited to participate. In this way a comprehensive and representative group of dedicated public servants was created for spreading ECD. Many of these same people remain involved to this day. They now form the basis of a common understanding about evaluation and a resource by which the basic principles of the national policy are being quickly passed on to all departments and agencies.

In Indonesia a number of one-week training courses were given to over 100 carefully-selected middle managers in 1995. This was done well in advance of the publication of the Ministerial Decree on Performance Evaluation, at the end of 1996. The courses succeeded in contributing to the development of a practical policy statement on evaluation, and in creating an expectation for a solid legal basis for the trainees to apply what they had learned.

Many of the middle managers trained were the same people who had earlier participated in the initial development of performance evaluation. They had been part of a participatory

process which started in 1994, and by the end of 1996, were awaiting the promulgation of the Decree by Bappenas.

This highly participatory approach adopted for ECD succeeded in creating both a demand for evaluation and the basis for the resources needed to satisfy that demand.

Participatory Approach

- Adapting evaluation principles
- Participation by all interested parties
- Representation of all accountability agencies
- Training as a development tool
- Creation of an interest in evaluation

Lesson Learned

A participatory approach to the acceptance of evaluation principles is needed to secure the support of both senior decisionmakers and middle managers who are responsible for the demand and the supply of evaluation information. In developing countries the existing conditions in the government and the capabilities of producing evaluation information must be considered when developing a consensus towards an evaluation framework.

Key Success Factor #4

The decision to have Bappenas, the National Development Planning Agency, be functionally responsible for the evaluation function, and to have the departments be responsible for its application, initially through the use of performance indicators.

In GOI most departments and agencies are responsible for evaluating what they do. Departments were not prepared to relinquish their control of departmental evaluation to Bappenas. Consequently, the decision as to who would be responsible for evaluation (and how they would be responsible) was critical.

The concept of having Bappenas as functionally responsible and the departments as responsible for doing the actual evaluation work was the key to the acceptance of an overall approach for performance evaluation in GOI.

Bappenas, with its few hundred staff, would never have been able to do the evaluation work needed. Retaining functional responsibility, prescribing how evaluation work was to be carried out and reported, and having departments do the evaluation work, was the only viable solution.

Departments are closest to the development work and they have access to the information needed for evaluation. Unlike developed countries where there may be many alternative sources of information possible for evaluation, in developing countries there is only one real source of information—the departments. It follows that, evaluation work can only be performed well if the departments are involved.

However, without central guidance from Bappenas, departments would not do evaluation at all or would do evaluation work using a wide range of different approaches, making the evaluation data of little use government-wide. Bappenas would then be unable to fulfill its national task regarding performance evaluation.

Bappenas now has the functional responsibility for performance evaluation in GOI while departments actually carry out the evaluation work. Bappenas assumes functional responsibility for performance evaluation in GOI by overseeing and assisting its development, use and overall effectiveness in supporting decisionmaking. Bappenas works closely with departments to develop and use performance indicators for GOI developmental projects (donor-assisted or not), as part of the budgeting system.

This arrangement has been made possible, in part, by developing a system of performance indicators, largely based on development results, formulated and maintained by departmental (structural) managers, and by bringing this system in close parallel with evaluation efforts practised by professional (functional) staff independent of management. All this is performed under the overall functional responsibility of Bappenas staff.

In the initial stages of implementing the national policy on performance evaluation, emphasis has been placed on the development and use of performance indicators. Later, more evaluation studies will be added to complement the performance evaluation efforts. The use of these performance indicators for development projects brings Bappenas and the departments closer together through both parties using performance indicators as a common language.

In Indonesia donors now can support the national policy on performance evaluation by co-operating with GOI in the development and use of performance indicators for development project inputs, outputs, results and impacts. The use of the same set of structured performance indicators for a given development project by both donor and recipient should strengthen ECD. Parallel reporting systems for recording the achievements of donor-assisted development projects, in line with donor evaluation systems, should be discouraged. Instead, as much as possible, donor countries should adapt to the evaluation systems of the developing countries.

Lesson Learned

In developing countries, the source of information is departments. There is usually insufficient governmental infrastructure to provide alternative sources of information. Departments tend to be strong and to guard fiercely their sphere of authority. Any

attempt at evaluation will be seen as a threat and will be resisted. Consequently, any realistic evaluation policy must recognise the key role of departments. This is usually best accomplished by having a central agency, such as Bappenas, be functionally responsible for the evaluation function and having the departments be responsible for the application of actual evaluation work, initially through the use of performance indicators.

Functional Responsibility

- Departments have the information
- Departments need direction
- Functional responsibility of Bappenas
- Decentralized implementation
- Performance indicators initially
- Donors should subordinate and integrate their evaluation efforts to existing practices

Key Success Factor #5

Active, visible and continuous participation by World Bank staff in the efforts of GOI affirmed the legitimacy of the evaluation function and reassured decisionmakers about the importance and benefits of evaluation.

In developing countries there is an understandable lack of confidence with new management approaches which have not yet been tried out or that change the status quo. This is especially true in matters having to do (or appearing to have to do) with accountability.

Senior managers will first want to be reassured by senior World Bank staff of the appropriateness of new management systems, and will want to hear more about these systems as they have been implemented elsewhere, preferably in other developing countries. For this reason, the World Bank, and other donors, should be active as advisors to senior management.

The support of donors, especially in the early days of ECD, is critical. This certainly was the case in Indonesia where the Bank staff were involved in numerous visits, steering committee meetings and seminars regarding evaluation, as well as in providing feedback to GOI on the ECD proposals and plans.

A supportive and involved approach by donors is needed early in ECD. The provision of financial resources alone is not enough.

Lesson Learned

Developing countries often lack confidence when it comes to implementing new management practices for which there is little previous internal expertise or experience.

Donors, such as the World Bank, can reassure developing countries of the importance and benefit of new management practices such as evaluation. Attendance and participation by donor senior officials at key meetings, and their implicit support for the efforts of the country, are as important as financial assistance, and will affirm the legitimacy of the evaluation function.

Support by Donors

- Compensate for a lack of confidence
- Financial support
- Moral support, to guide and reassure
- Participation, feedback and encouragement

Key Success Factor #6

Lobbying for and promulgation of a National Policy for Development Project Performance Evaluation, in the form of a Ministerial Decree, around which continued support of a core group of ECD supporters could be consolidated, and a common understanding about evaluation confirmed.

Formal instruments, like the Ministerial Decree on performance evaluation in GOI anchor future management actions and provide legitimacy, authority and credibility to those performing managerial tasks. Without such formal instruments, little can be accomplished. In Indonesia, as in other developing countries, managerial actions have to be supported by some form of state authority.

The Decree on performance evaluation confirms a uniform terminology and common understanding of what is meant by performance evaluation in the Indonesian context.

The development and promulgation of the Decree required the sustained effort of a few senior bureaucrats from early 1994 to the end of 1996. Consensus is essential in Indonesia and can only be achieved after long and often difficult discussions among the various stakeholders. Initially, departments and agencies could all see their authority and independence being diminished. Considerable time was needed to arrive at consensus as to the advantage of having a national policy.

Lesson Learned

In the bureaucracies of developing countries it is essential to have a formal instrument to provide a legal basis for the development of evaluation. The mobilisation of resources and the authorisation for action requires such legal backing: this makes it critical in developing countries that ECD secure the promulgation of a decree (or its equivalent) in support of the evaluation function. Intensive lobbying will often be needed to ensure that this happens.

National Policy

- Formal instruments are important for:
 - Consensus
 - Common understanding
 - Empowerment
- ECD goal from the very start
- Strong lobbying

Evaluation Capacity Development in Zimbabwe

Issues and Opportunities

Stephen Brushett

The Backdrop

The following information provides a context for understanding the current status and probable future trend for evaluation capacity development (ECD) in Zimbabwe:

- There has been consistent and sustained interest expressed by central government planning authorities since the late 1980s in improving the quantum and quality of evaluation. This interest was not shared evenly throughout government, but evaluation had a strong champion in the National Planning Agency (NPA) which at that time was located in the Ministry of Finance. The Director of the NPA (Dr. Stan Mahlahla) collaborated with a UNDP-funded external specialist (Adil Khan) to draw up a first conceptual framework for Zimbabwe in 1991 encompassing implementation monitoring, project completion reporting, performance monitoring and impact evaluation.
- The starting point was the perceived need for better information on the effectiveness and implementation of the five-year (static) development plan which was the primary tool for managing the government's economic development program up to the early 1990s. The focus, however, began to shift after the government adopted its economic structural adjustment program, and greater emphasis is now placed on the three-year (rolling) public sector investment program (PSIP) which is in principle an integral element of the fiscal budget processes.
- Government established a National Economic Planning Commission (NEPC) separate from the Ministry of Finance and under the Office of the President and Cabinet (OPC) with specific responsibilities for monitoring and evaluation (M&E). This was in line with the Mahlahla/Khan recommendations. The role and responsibilities of the NEPC are set out in a Presidential Directive dated July 11, 1994. These include: coordination of long-term plans and development policies; preparation of annual and medium-term development programs; assistance to the development of provincial development plans; monitoring and evaluation of policy implementation; and impact assessment of projects and programs (in close collaboration with line ministries and executing

agencies). The secretariat for the NEPC was quickly established and staffed for its M&E role and a Commissioner was appointed at Cabinet level. However the Commission itself, a decisionmaking body with Cabinet level representation, has never been established. Nor has the National Development Council which was to be the tool for wider consultation on planning matters with civil society.

- Government has enjoyed consistent, if still low-level, support (initially from UNDP, latterly from the Bank) to develop this interest in M&E and to exploit the institutional reforms leading to the establishment of NEPC into an appropriate framework and program for ECD. The Mahlahla/Khan work has been followed by exchange visits to OED in the Bank in 1992 and 1994, which helped build knowledge in Zimbabwe about best M&E practices. Then came a further study in November 1994 funded by the Bank and carried out by the Netherlands Economic Institute (NEI) to develop a proposal for a program on which basis government could request assistance from the Bank and others.
- Support for the program has to date been characterized by an 18-month pilot phase, mid-1996 to the present, which has involved NEPC and three pilot ministries (Agriculture, Health and Local Government) financed primarily by an IDF grant administered by the Bank of US\$340,000. After the pilot, a second phase was proposed to start by mid-1998 which would now involve all (15) key economic ministries. The pilot phase has (with some help from the Bank, most notably from Pablo Guerrero) produced useful output, in particular the outline of a possible evaluation framework and a detailed situation analysis of M&E in Zimbabwe leading to the policy, legal and institutional measures that need to be taken to underpin the evaluation framework. What else is suggested by pilot experience? More groundwork needs to be put into consolidating and strengthening the public sector investment program (PSIP) process before M&E can really take off; a successful program will probably depend on broader-based public sector reform to strengthen key line ministries and create better incentives in the public service; a considerable amount of effort still has to go into building skills and knowledge and to retaining staff with those skills in the appropriate institutions.
- Despite all these efforts, achievements on the ground are very few and ECD still lacks credibility in Zimbabwe. Changes in management and institutional structure within the OPC have limited NEPC's effectiveness as an apex for M&E and as the driving force for setting the framework in place. Nothing systematic has been achieved yet, and no clear guidelines are yet in place to help implementing agencies plan and implement their own M&E work. As a result almost all evaluation remains donor-driven; what has been

delivered to date by NEPC has tended to come too late in the process and has not been of great value for future project design. More focus on mid-term review and ongoing M&E would help increase the relevance of the work.

The Constraints

What are the key constraints which have to be overcome, and are still to be fully addressed, for sustainable ECD in Zimbabwe?

- There is a need to develop a receptive *culture* in which demand for and effective use of evaluation output can grow. This may prove to be easier to achieve in a market-driven policy reform environment than under the previous conditions. Certainly a positive factor is recent evidence that demand at Cabinet level for information on the implementation of policy and programs is strong and increasing. This opportunity needs to be fully exploited by clarifying roles and responsibilities and ensuring that evaluation information is properly gathered, organized and analyzed.
- The next most important requirement is *sensitization* of key stakeholders, especially in the public sector, to the need for, and benefits of, evaluation. This goes together with building *awareness* of techniques and approaches which are workable in the Zimbabwe context. These are the critical building blocks to sustainable demand for evaluation outputs. At least a start has been made during the pilot phase given in NEPC and the pilot ministries: a significant amount of training has been carried out. To be sustainable, this training has now to be applied in specific processes and products which build up a solid base of evaluation output.
- *Linkage* has to be built between evaluation and specific public sector processes so that this is well understood and straightforward to implement. In Zimbabwe, the PSIP process, for which NEPC is formally responsible, provides potential linkage in principle. In practice though this has yet to take off. Linkage provides a context in which incentives to perform and to provide information and feedback, and sanctions for non-compliance, can be properly considered. Funding for both ongoing and future projects could be withheld (or additional funding could be made available to deserving cases) using the evaluation data base.
- Some constraints to effective implementation of the evaluation framework were highlighted by the experience of the pilot phase:

-
- (i) PSIP processes are not fully articulated and reporting requirements are not fully established. Because of this it is difficult to determine compliance or non-compliance and there is no established or agreed basis for sanctions.
 - (ii) Initial approaches have focused on quantitative rather than qualitative data, something which NEPC continues to favor (NEPC wants to put in place large MIS-style systems at an early stage). This creates some problems as most data may be difficult to retrieve and may be open to misinterpretation. A broader understanding is needed about evaluation information and the uses to which it can be put.
 - (iii) While some good work has been done on preparing project profile forms and clarifying information requirements, it is still easier for line ministries and executing agencies to see the costs but not the benefits of evaluation. The costs to line ministries and agencies of “feeding the beast” are evident and seemingly increasing: the possible benefits of more data collection and regular quarterly reporting are not.
 - (iv) A “Them” and “Us” mentality can very quickly be built up (who is evaluating whom, and for what purpose?) unless there is transparency in the process and NEPC’s responsibilities are clear. The notion of an apex should have strong positive connotations (center of excellence, quality input to review of ministry submissions, lead role in complex assignments). The incentives for the line ministries need to be stronger.
 - (v) NEPC must be able to keep and develop good people to establish and maintain its credibility as an apex institution. Government as a whole must be supportive and ensure that the quality of staff is maintained (and then improved) and that issues associated with overlapping or unclear mandates are dealt with decisively. Line ministries need more people than they currently have trained and designated for evaluation. However, some good staff who have left NEPC in the past few years now work in these ministries so the public sector as a whole has not been the loser.

Next Steps

What are the next steps needed to enable the second phase to contribute to the establishment of a sustainable evaluation framework for ECD in Zimbabwe? This is the basic list of items for the second phase of the program to begin mid-1998. High-level backing and consistent follow-through will be needed for the framework to move closer to reality. Some of the critical measures are:

-
- Modify/strengthen institutional structure (within OPC) to create a more effective focal point for evaluation, and build demand for project and program evaluation at Cabinet level. Most legal instruments are in place, but the lack of a fully constituted NEPC means that no high-level focal point exists; linkage of program evaluation to policy evaluation would increase demand and interest, but the respective roles of the Monitoring and Implementation Department (MID) and NEPC still have to be sorted out. If NEPC is to be the apex, this has to be agreed and acted upon. MID and the Auditor General seem to have a ready market for their product: so far, NEPC does not. One option may be to restructure/merge MID and NEPC evaluation functions.
 - Consolidate the linkage of evaluation to public investment programming through completing the PSIP manual, the monitoring and evaluation guidelines, quarterly reporting, and project profiling. This could start in fiscal 1998/99 with a few key ministries (the original three and three/four more), then roll out to all (15) economic ministries within two years.
 - Recognize that the development of human capacity must come before the development of systems. A major training effort is still needed together with commitment to ensure continuity of staffing at the focal point, and exchanges with line ministry and agency staff are needed to build coalitions, encourage collaboration and discourage competition. The supply of evaluation skills available to Zimbabwe needs to be grown according to a series of annual targets.
 - Ensure that adequate local resources are made available commensurate with the actual annual requirements of the evaluation program. This includes, but is not limited to, the suggestion that there be a budget line item for evaluation for each line ministry and agency. A number of ministries have planning or change management units which could accommodate in-house evaluation capacity if adequate resources were provided.
 - Prepare and follow-through annual evaluation plans with clear targets and identified outputs to help build experience and strengthen the track record. The initial target should be three major evaluations a year to be led by NEPC. The first such plan should be for fiscal 1998/99.

Further Considerations

- The impact of decentralization policy on how projects are prepared, appraised, monitored and evaluated should be examined. Some provincial-level planning staff are in

place but their task in evaluation is not clearly spelled out. These resources are not being effectively used to provide support to bottom-up planning processes which should enrich the quality and relevance of public sector investment programming. In the medium-term, decentralization will probably increase demand for evaluation from the center as more responsibility for preparation, appraisal and monitoring is pushed out.

- Attention should be given to building the local consultancy resource base for evaluation. Zimbabwe has a pool of public sector expertise now in the private sector. Local consultants could readily be used to supplement the efforts of core ministry personnel on evaluations and could take the lead on some assignments. The local consultancy input to the design and implementation of the pilot phase was of good quality. A data bank of resumes of good candidates could be prepared, using what is already available on local consultants in the Bank resident mission as a start.
- The role of the Bank, and possibly other donors, needs to be considered in building and sustaining best practice initially, perhaps, through one of the three major annual evaluations. This means going much further than the usual ICR routine. One attempt to do this was the completion evaluation of the Emergency Drought Recovery and Mitigation Project (EDRMP). The ICR was completed in mid-1995 after a series of collaborative exercises with government, managed through the NEPC. The process included an early start to the evaluation exercise, a joint workshop to discuss findings at the draft stage, and the dissemination of the evaluation outcome to a broad range of local stakeholders. Both the ICR process and the final document have been cited as best practice by OED (Slade for OED to Marshall for AFIDR June 20, 1996).

Outline of a Possible Evaluation Framework for Zimbabwe

1. NEPC Evaluation Function

- sets evaluation system directives and standards
- sets guidelines for annual evaluation plans by line Ministries and key agencies
- develops and organizes training programs and workshops to disseminate methods and lessons of a cross-cutting nature
- reviews TORs prepared by ministries for evaluations of major programs
- reviews the quality of evaluations done by ministries
- undertakes or participates in selected evaluations of complex projects, for example,

-
- where more than two ministries are involved
 - uses key performance information and evaluations from ministries to inform OPC, MOF; draws implications for policy and new program formulation, monitors performance across sectors and systemic issues, and delivers information before budget allocations
 - follows up on the implementation of evaluation recommendations
 - prepares annual report on evaluation results and use of evaluation findings.

2. Evaluation Function in Line Ministries

- define evaluation plans covering all major programs, taking into account evaluations by third parties.
- define, jointly with the sector units of NEPC, the key performance indicators to be used to assess progress and results—these should be defined at the design/appraisal phase of new programs, and retrofitted to ongoing programs and projects
- delivery semi-annually to NEPC key performance information for major programs and major evaluation studies, according to annual evaluation program
- budget resources for monitoring and evaluation tasks
- establish feedback processes to ensure use of evaluation results within the ministry
- within each ministry, line departments monitor key performance indicators and submit results quarterly to internal central evaluation unit
- set up evaluation review committee at senior level to review evaluation reports and, periodically, to review performance information.

3. Organizational Implications

- Evaluation Unit of NEPC to be independent of line/sector functions and report to Head of Planning, or to OPC
- Rationalize evaluation responsibilities and working relationships between NEPC, MID and MIU
- Establish evaluation function at apex of each ministry, to focus on carrying out evaluations of major programs and projects, and analysis of performance monitoring information; central evaluation function, could initially be attached to central planning units
- Establish performance monitoring responsibilities for projects and programs in line departments within each ministry
- Auditor General to carry out a bi-annual process audit of the evaluation system established, and receive evaluation reports for major programs and projects.

Institutionalizing Evaluation in Zimbabwe: Factors to Consider Situation Analysis

Institutional Setting—Relations Between Public Sector Institutions and Their Functions

- The relationships between NEPC, other central ministries and agencies, including MID and MIU, and line ministries, with respect to: formulating policies, strategic planning, allocating public expenditures, defining public investment programs, and allocating and monitoring budgets, auditing and financial management, and monitoring and evaluating performance of public sector resource use
- The relationship of the preceding functions within major public sector agencies
- The relationship between NEPC and other executive level agencies, such as OPC, and with independent oversight agencies, such as with the Auditor General and Comptroller's Office, or the Public Accounts Committee of Parliament
- The legal or regulatory basis underpinning these relationships, including the role of the Public Service Commission, and other relevant official bodies
- The division of responsibilities between NEPC, other central agencies, such as the Ministry of Local Government, line ministries, decentralized government functions, and local authorities.

Monitoring and Evaluation Setting

- What is being monitored or evaluated at present, and by whom?
- Is the M&E focus on policies, programs or projects?
- Is the focus on public investment, or on public expenditures?
- Are Parastatals evaluated? What is the role of the Auditor General in evaluating Parastatals, for example, through value-for-money audits?
- Is priority given to evaluation of externally funded operations, and if so, what is the level of coverage? The Health Ministry suggests that 50% of its programs are covered by donor-specified evaluations. What is the situation in other ministries—line and central?
- What is the focus of monitoring and evaluation work being done? Is it focusing on input monitoring, or on development? Is it focused on lessons or on accountability?
- Who receives evaluations and monitoring data at present? Is the information kept by the sponsors, or are the data used within ministries or across ministries?
- What are the feedback mechanisms in place? Are evaluations sponsored by donors used? And, if so, for what purpose? Is there a link from monitoring and evaluation information to decisionmaking? What is the situation in the key agencies and across agencies?

-
- Who follows up on monitoring and evaluation findings and recommendations?
 - Is evaluation or monitoring information released outside the agencies? If so, what are the reporting arrangements?

Monitoring and Evaluation Resources

- Are there budget set asides for monitoring and evaluation? Are the resources built-in in externally funded operations, or are they funded out of domestic agency resources? What level of resources is being applied to monitoring and to evaluation?
- What are the professional skills of staff involved in monitoring and in evaluation?
- Is training for monitoring or evaluation being done?
- Are external agencies, donors, and NGOs inviting national staff to participate in externally driven evaluations?
- Do agencies have their own evaluation practices, methods, rules, policies or guidelines?
- What information technology and systems resources are available?

Bibliography

Task One: Situation Analysis Report—Strengthening of the Monitoring and Evaluation Capacity for Public Sector Investment and Planning In Zimbabwe by ISO Swedish Management Group and ARA Techtop Consulting Services dated September 1996.

Supplement to Situation Analysis Report (NEPC—Strengthening of the Monitoring and Evaluation Capacity for Public Sector Investment and Planning in Zimbabwe) by ISO Swedish Management Group and Ara-Techtop (Zimbabwe) dated January 1998.

Draft Monitoring and Evaluation Guidelines (NEPC—Strengthening of the Monitoring and Evaluation Capacity for Public Sector Investment and Planning in Zimbabwe) by ISO Swedish Management Group and Ara-Techtop (Zimbabwe) dated February 1998.

Comments

Eduardo Wiesner

Introduction

Perhaps there is a common thread—an analytical framework—that may help us go beyond the particulars and peculiarities of each country experience.

The Case of Indonesia

Mr. Barbarie's paper on Indonesia focuses well on the fundamentals of evaluation capacity development as a "process." And, as neoinstitutional economics remind us, "process matters." This highlights two aspects. First, the idea of beginning with a diagnostic of the evaluation capacity of Indonesia. Most countries overestimate the amount and quality of the evaluation that takes place in their public sectors. It is a revelation to them to discover how little evaluation is actually carried out. It would be interesting to explore if the evaluations conducted in the mid-80s hinted at, or foresaw the troubles that recently affected Indonesia. Second, it is commendable that the authorities focus more on results than on process compliance. This is the correct approach.

The Case of Chile

Chile has been offering, by example, good advice to many developing countries on social and economic issues. Soon that good advice will also cover evaluation. Mr. Marcel's presentation synthesized how to organize an evaluation system from the perspective of its operational features. Its most attractive merit is the gradual approach. But here, gradual means an effective process of implementation, not a justification to do little. A second highlight is the well-balanced proximity between evaluations and budget allocations. It seems that Chile will make substantial progress in establishing that critical link.

The Case of Zimbabwe

Mr. Brushett's paper raises some key issues. Two points with regard to Zimbabwe deserve special prominence. First, the separation of the evaluation unit from the Ministry of Finance. Second, the strategy of decentralizing evaluation. On the first point, while distance from the Ministry of Finance may give evaluation independence, at the same time, there is the risk of isolation from the budget process. Decentralization

certainly is the right strategy in most cases. But it is not a free panacea. Certain conditions have to be met. And they are not easy conditions to meet.

Is There an Algorithm for the “How” of Evaluation?

The unifying analytical framework to understand evaluation capacity development is the demand side of evaluation. If there is an algorithm for evaluation it will be found in the factors determining that demand. However, this algorithm is elusive because it assumes that the right incentive structure is easily “reachable” when, in fact, it is not. In most developing countries, the norm is the wrong incentive structure. This is particularly true in the budget process where public sector rent-seekers have captured a significant portion of public policies and resources. These rent-seekers do not like evaluations—particularly independent evaluations. As Professor Mancur Olson would have said, they “free ride” on public resources and are usually able to prevail over the public interest.

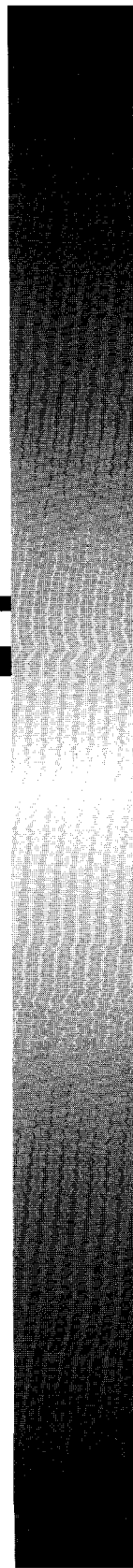
In brief, the question is not what evaluation can do for you but what you can do—through evaluation—to those who now profit from the relative absence of evaluation. They will resist the development of evaluation capacity. And they are well organized to do so.

Strategic Evaluations as a Way to Reveal the Political Economy Restrictions

An effective strategy to enhance evaluation capacity development should focus more on the restrictions than on the intrinsic merits of evaluation. This is the way to effectively move forward with evaluation development. To strengthen the demand for evaluation, the right policy incentives need to be in place.

Lastly, strategic evaluations of key projects and policies are a fruitful way to begin to deal with the real political restrictions that adversely affect evaluation. Colombia is dealing with these political economy restrictions for those who will fight the evaluations—not explicitly of course, nobody does this explicitly—by conducting strategic evaluations, where some areas, programs, or some policies are chosen and an independent evaluator, or group of evaluators show what actually happens. Chile and other countries are beginning to use this approach. Then the public, little by little, realizes and learns how inefficient the public sector is and how inefficient the budget allocations are. By divulging these results, one creates political support for more evaluations and for a more efficient public sector.

PART 5:
Two Perspectives



A View from a World Bank Network

Cheryl Gray

The Public Sector Group in the Poverty Reduction and Economic Management Network oversees the Bank's work in public sector reform, which includes both public finance reform and public sector management and governance reform. Obviously this is extremely important to almost everything we do, and one cannot discuss good public sector performance in areas like service delivery, public expenditure management, taxation, civil service reform, and judicial reform without thinking about the capacity to evaluate and feedback to government. It is a critical part of reform. The evaluation capacity in the Bank for our projects is a key issue. Another is evaluation capacity in countries. For the time being, let us focus mostly on in-country evaluation capacity issues.

On the question of evaluation capacity development, four themes come out very strongly. First, the need to focus on institutions, as institutional economists might define them, meaning the rules of the game. These could be within organizations. They could be the laws for market behavior. They could be the norms of cultural interaction. We say that we need to focus on the rules and incentives that govern behavior. But what does that mean? The idea of capacity building, narrowly defined, is the supply-side approach to building capacity. A focus on institutions is much more supply and demand taken together—how capacity is developed, how it is used, and whether it is used. Certainly, developing capacity is important, but developing capacity on its own may not achieve anything. There has to be demand for the capacity or it will not be used. It is important to keep this in mind. The Bank has been (and continues to be) supply oriented. But now the Bank should look at the rules of the game and the institutions in-country for using capacity. So the first point is to focus on institutions.

The second theme is how do you move from a dysfunctional public sector without accountability, without capacity, without clear checks and balances to a more functional one? This is the big challenge. Last year the *World Development Report* on the role of the state pointed out three different mechanisms—three different forces—that help in public sector reform. The first is reforming the hierarchy in government, that is, reforming the rules within government. For example, a medium-term expenditure framework for controlling budgeting decisions or improving tax administration, or for improving evaluation capacity in government can change the rules within government. Next is the voice of the public and the civil society as a mechanism for oversight of the public sector and feedback to government on performance. Then comes competition, contestability and exit mechanisms, bringing market forces to bear so that government itself has to compete, either with other parts of government or with the

private sector. All three modes have their counterparts in evaluation. In the first, the public sector provides the evaluation. Australia has a very good evaluation system for the public sector. In the second, the voice mechanism, perhaps the private sector or NGOs would provide the evaluation—as with the Bangalore scorecard when citizenry mobilized to evaluate government and to publish their evaluation findings. In the third mode, the market evaluates and selects. So it is not only the public sector, but also other parts of the system that provide the evaluation which is so critical for public sector functioning and reform.

The third theme is the question of instruments. Our public sector strategy is stressing the need to think about new instruments for public sector reform. The Bank has short-term adjustment lending through which it has a lot of leverage, but this happens very quickly. And it is hard to have fundamental institutional change over a short period. We have technical assistance (TA) lending, which may not have the leverage to actually accomplish much, (and there are incentive problems in our technical assistance). We have typical investment loans, and they also have their problems. We get our own projects working very well, but these too often become separated from government functioning. We have had good institutional work in some projects. But there are inherent problems with our current instruments. An idea now spreading throughout the Bank is the need for a new instrument for public sector reform which would be a longer-term instrument with a high level of resource transfer to provide leverage. This could perhaps support a time-slice of an investment program, say, make a disbursement every six months. But this would not be tied to particular purchases. Instead, the conditionality would involve fundamental reform of the public sector. One type of loan for this is a public expenditure reform loan (a PERL), where the whole public expenditure management system is under scrutiny and open to reform. Of course, the loan disburses in line with the general investment or expenditure program of the country.

Some new approaches are already being tried out. We have adaptable program lending, and we have the new learning and innovation loans to test various approaches. So the Bank is already experimenting with new lending instruments oriented towards building capacity in the public sector. For this program approach, evaluation is absolutely critical: these new instruments will increase the demand for better evaluation capacity, both in the Bank and in the country.

Now, lastly, to the Bank incentive question. And this really is an issue of in-Bank evaluation capacity. We have a complex mix of incentives within which staff, task managers and country directors design and disburse loans and help governments. OED's and QAG's work in oversight, evaluation and constant examination is vitally important in creating an internal incentive structure oriented toward greater development effectiveness.

A View from USAID

Gerald Britan

The United States has a strong, historical commitment to be transparent and results-oriented in its domestic public administration. Under the Government Performance and Results Act of 1993, the requirement for effective performance measurement and evaluation is now the law of the land. Every federal agency in the U.S. government must develop strategic and annual plans that include specific performance goals, measure evaluated results against these plans, and adjust programs and budgets accordingly. As President Clinton said when he signed the bill, “This law simply requires that we chart a course for every endeavor that we take the people’s money for, see how well we’re progressing, tell the public how we’re doing, stop things that don’t work, and never stop improving the things that we think are worth investing in.”

In the development sphere, USAID has long and strongly advocated performance measurement, evaluation and results-based management. From the invention of the logical framework in the 1960s to the development of impact evaluations in the 1980s, and the evaluation of strategic objectives in the 1990s, our commitment to evaluating performance and learning from experience has been unwavering. USAID sees the development of effective performance monitoring and evaluation, the development of good information for results-based management, as a key element for all our development assistance. Partnership, collaboration, and empowerment are among the core values that guide this development assistance. And we use performance monitoring and evaluation in all our development programs to work with, learn from, and strengthen the capabilities of our partners. Building the capacities of these partners in both the governmental and nongovernmental sectors is a key feature of our development assistance effort.

These values are explicitly recognized in USAID. Our policy on participation states, “Operating units and project teams shall involve USAID customers and partners in planning approaches to monitoring performance and planning and conducting evaluation activities as well as in collecting, reviewing, and interpreting performance information.” Planning, monitoring and evaluation are seen as closely linked.

In our policy on building performance measurement and evaluation capacity, for example, we state “The agency and its operating units shall attempt to build performance monitoring and evaluation capacity within recipient developing countries. Operating units shall integrate, wherever feasible, performance monitoring and evaluation activities with similar processes of host countries and donors.” And then the section on

information sharing says, “Whenever feasible and appropriate, the agency and its operating units shall participate in networks for exchange and sharing of development experience and development information resources, with development partners, host country development practitioners, researchers, and other donors.”

Evaluation capacity development in USAID is not only about good management, but also (and perhaps even more critically) about good government. USAID’s collaboration in evaluation capacity development is clearest in our efforts to promote more effective and accountable governance. And it is a key feature of our substantive programs in other areas. In health and population, for example, USAID pioneered the development of capacities for collecting performance information through demographic and health surveys. Recently, USAID added a module for education data to these surveys. Through the environmental programs, we focused on strengthening local capacity to monitor and evaluate areas such as biodiversity, greenhouse gas emission, and deforestation. And now USAID is starting a major effort in a more difficult field, that of democratization, civil society strengthening, and the like. This will mean establishing how to monitor and evaluate progress in those activities.

The Center for Development Information and Evaluation (CDIE), is USAID’s central evaluation office. It does not formally list evaluation capacity-building as one of its functions. Nor is there any program budget devoted explicitly to evaluation capacity development activities. But CDIE does carry out various related tasks that contribute to strengthening evaluation capacity in the development process, including participating in workshops, doing studies and conducting technical assistance activities. So by including host country evaluators in our performance monitoring and evaluation workshops and training; by including host country evaluators on CDIE’s own evaluation teams; by participating in joint and collaborative evaluations—and we hope to do more of these; by widely disseminating USAID’s evaluation findings, guidance, and lessons learned; and by organizing and leading international efforts in evaluation information exchange, we hope to contribute substantially to evaluation capacity development.

At the operational level, a review of USAID’s activities database showed that USAID often encourages evaluation capacity development. But this was rarely the exclusive focus of a project; it was usually a subcomponent found in all regions and all sectors.

In USAID’s case, it was also rarely a government-wide evaluation capacity development effort. USAID would seldom go out and work to develop the capacity of that government to monitor and evaluate programs government-wide through the Ministry of Finance or the Ministry of Planning. Typical evaluation capacity development activities were

targeted on institutions, usually project implementing agencies or partners in USAID's strategic objectives. And these were focused on developing the performance monitoring and evaluation capacities of nongovernmental organizations, local governments, NGOs, PVOs, and foundations.

We all recognize that supply without demand in the economic world leads to bankruptcy. In the development world it does not lead to effective development. Our efforts towards democratization and good governance focus on development outcomes that will increase the demand for good performance measurement, good evaluation, and good management.

Our evaluation capacity development efforts, and our performance monitoring and evaluation activities more generally, bear on operational improvement and learning (primarily in terms of how we do the development activities in question). In the last few years, USAID is shifting to a more strategic focus on evaluating capacity, performance monitoring, managing for results, in sectors and in larger program areas.

Under re-engineering, USAID is placing increased emphasis on monitoring and evaluating this broader strategic, sectoral and program performance. There is more emphasis on the performance monitoring side of evaluation, (how to link evaluation to ongoing management processes), and more emphasis on performance information use. From USAID's point of view monitoring is less for external accountability and more to provide managers with the information they need to manage. That need is answered at least in part by more formal evaluation activities. USAID is using the OECD/DAC strategy for the 21st century as a framework for collaborative effort, and the DAC goals are very clearly in our Agency-wide strategic plan required under the Government Performance and Results Act.

Evaluation capacity development is not a distinct goal for USAID or an end in itself. It is a core component of what we do. It clearly includes performance monitoring and a more strategic and results-oriented planning effort. It is clearly embodied in USAID's emphasis on managing for results. It clearly encompasses our focus on good management and good government. It is clearly related to our concern for enhancing democratization. Properly understood, evaluation capacity development is real social action for real social change. It means strengthening both the government's and the people's ability to chart a course for themselves using concrete measures of what is being accomplished and what is wanted.

PART 6:
Where Do We Go
from Here?



Overview and Conclusions

Elizabeth McAllister

Results-based management and program evaluation are central to sound public sector management. These are prerequisites for high performance, public expenditure management.

Within the institutional context, we need to understand incentives and to create incentives to sustain the interest on performance measurement. Because we have been urged to look at the links to other processes within institutional development, we now understand that ECD can influence the allocation

of resources in the public sector: and that means ECD has a strong link to the planning and the budgeting system to accountability, to creating and motivating change within organizations, and to learning how to improve the quality of public sector services.

All of this tells us, that ECD needs to be mainstreamed and made part of how the Bank and other donors evaluate, plan our country assistance strategies, and monitor and evaluate work at the project, sector and the country levels.

These seminar papers have reaffirmed the conclusions of the 1994 Task Force on Evaluation Capacity Development. One important lesson is that we need to have a long-term perspective. This cannot be done with a two-year grant, but perhaps we need a two-year grant to get started, to do the inventory and diagnosis, do the planning, and develop the social science capacity to do good evaluation. And then later a phase of looking at how we sustain and broaden the impact of evaluation capacity. But however we do it, we will have to be in for the long-run—at least eight to ten years.

Results-Based Management and Evaluation Capacity Development Are Central to Sound Public Sector Management

- Prerequisites for high-performing public expenditure management
- Powerful links to broader institutional capacity issues
- Important implications for donor country programs

1994 ECD Task Force Conclusions Are Reaffirmed

- Take a long-term perspective
- Identify and support champions
- Adjust to country-specific circumstances and culture

We have reaffirmed the importance of champion departments and champion individuals, to build and to sustain ownership in developing countries. And we have reaffirmed the need to adjust and link ECD to country-specific conditions and culture. This cannot be donor-driven; it has to be done in partnership with others.

How do we now move from words to action? One strong plea from those who are involved as professionals and those who have joined us as task managers/country directors is: Please demystify, simplify, and clarify. As a community we have to think very carefully about what terms we are using to tackle areas that remain ambiguous. There is ambiguity in a number of issues covered by the papers: about whether ECD is something that would best be developed centrally or decentrally in governments; and whether the approach should be holistic or fragmented. And we need to clarify the difference between results-based management, performance measurement, and program evaluation. And we have looked at the issue of accountability and its seeming contradiction with learning.

Part of our challenge ahead is to look at when we move on a government-wide basis and when we move to work on a decentralized basis in various departments. What do we mean? Are we concentrating first on results-based management? Do we move directly into program evaluation? And what do we mean by those terms? Interest in performance measurement can start at the managerial level, with more of a results-orientation, which is a focus in the managerial sense of defining inputs, desired outcomes and results, and the indicators and milestones for achieving these. Perhaps this can be done in countries where the social science skills are not so well developed.

A results-based approach helps us to track performance, and to have early warning signals of when our projects or programs are going off-track. This is focused largely at the program managerial level. And this feeds into reporting and demonstrating progress and learning at the end of the corporate programming and planning system. But evaluation also asks deeper questions: Would you have achieved those results anyway? Could you have done something better with the funds available? Why have you been successful?

It is important for us as a community to move forward to clarify the meaning of all these terms, such as evaluation, performance, indicators. What matters is that as a group we have to come (and as a community of evaluators we need to come) to some common definitions and use them so that the task managers, our partners in developing countries, are not confused by different uses of the same term.

Another area is accountability. We have had some confusion about whether there is a

Complementary Roles: Results-Based Management, Internal Audit and Evaluation

Defining features	RBM	Internal audit	Evaluation
Responsibility	program manager	corporate	central/corporate
Focus	both development and operational results	management processes and operational results	development results (intended and unintended); involves all stakeholders, including clients and other donors
Primary purposes	clarifies program objectives; links program and project activities and associated resources to objectives; assists managers to understand when their programs are succeeding or failing by signaling potential management problems, or fundamental design and delivery problems	assists managers to understand what results were or were not achieved, and why; and whether monies were spent as intended; proposes operational and administrative improvements	assists managers to understand why and how intended development results were or were not achieved; makes recommendations for policy, program and project improvements
Typical scope	monitors program and project operations, and levels of service; reports on progress to managers and alerts them to problems requiring action; monitors cost-effectiveness of management and delivery	examines how well management systems and practices work; assesses risk and control; seeks more cost-effective ways of managing and delivering programs and projects	questions development rationale and objectives; determines success in meeting results; seeks more cost-effective ways of achieving program results
Coverage/frequency	provides on-going measurement of programs and projects	as required, assessment of selected systems, programs and projects	as required, retrospective measurement of key policies, programs and projects
Common elements	timely, relevant and evidence-based information		

contradiction between moving into learning and using evaluation for learning, versus having it for accountability and having a price to be paid for learning. Some talk about accountability in two ways. One, that if we think about accountability in the old way of command, control and hierarchical organization, then we use it to appoint blame and to punish those who have done wrong. But when one thinks that being held accountable means being held at risk for blame, one's learning system shuts down. Another emerging approach uses the concept of partnership or shared accountability, which is based more on a learning organization environment, and focuses on trying to open up learning opportunities, and seeing learning as an insurance so that if something goes wrong you can say, "Here's what we learned from it."

We now understand that to build ownership of evaluations in developing countries we have to build the capacity to evaluate. The Canadian International Development Research Centre (IDRC) recently did an evaluation of their evaluations. The feedback from research institutes in developing countries was "Your evaluations may be useful for you in your project but they don't help us understand our institution, and they don't help us understand the greater context within which that project is operating, which is our preoccupation."

Accountability Concepts: Hierarchical Versus Partnership

Hierarchical:

- Punish wrongdoing
- Appoint blame

Partnership:

- Based on obligation to demonstrate and take responsibility for performance in light of agreed expectations
- Clear roles and responsibility
- Clear performance expectations
- Balanced expectations and capacities
- Credible reporting
- Reasonable review and adjustment

The other genesis of true partnership accountability is the recognition that we are jointly working on our projects. If we build an accountability system to feed our headquarters without making it useful at the field level, we cannot have a great deal of confidence in the information we are collecting. So for partnership accountability to

work, and to help us better manage our programs, we need to have a clear definition of roles and responsibilities. We need to have clear performance expectations from the outset that recognize that there is asymmetry between the donor and the recipient government, and between the government and civil society who may be involved in implementing a project. So our expectations need to be balanced according to our capacities. We will need credible reporting to demonstrate what has been achieved and what has been learned, and reasonable review and assessment.

Strategy for Capacity Building

- Commit to partnership with government, civil society, academics
- Create demand
- Sequential/interactive approach

We need to be more strategic and to accept, that we cannot approach evaluation capacity-building in a shotgun manner, that it is a matter of building demand, and building capacity and that each country will be different. Among ourselves we need to look at committing to partnerships with government, civil society, and academics.

We have learned that as we go forward we need to have a sequential and interactive approach. It is now time to move from telling people what we think needs to be done to knowing what, and how. How do we codify what we know already? We should look at the different players in evaluation capacity development. Each player will have a different perspective and we can learn from this rich variety of experience.

The next move is for us to work together as donors, to develop common standards of definitions and of methodologies. The OECD's DAC has warned us all that in this culture of growing accountability we cannot disempower leaders in developing countries. Making our systems more complex and more demanding dilutes the concentration needed at the country level. If we are not building evaluation systems that make sense at the country level, we are not contributing to sustainability.

Action Learning: from Knowing What to Knowing How

- Codify what we know
- Case studies—test and learn
- Rethink role of the evaluator
- Donor collaboration to support ECD, and move responsibility to developing country governments

Recommended Reading

1. The World Bank, *Colombia: Paving the Way for a Results-Oriented Public Sector*, WB Report no. 15300-CO, January 28, 1997.
2. The World Bank, *Report of the Evaluation Capacity Task Force*, June 1994.
3. Eduardo Wiesner, *From Macroeconomic Correction to Public Sector Reform: the Critical Role of Evaluation*, WB Discussion Paper no. 214, 1993.
4. Derek Poate, *Measuring & Managing Results: Lessons for Development Cooperation*, SIDA/UNDP, 1997.
5. Robert Picciotto and Eduardo Wiesner (eds.), *Evaluation and Development*, Transaction Publishers, New Jersey, 1998.
6. Pablo Guerrero O., "Evaluation capacity development in developing countries: applying the lessons from experience", in R. Boyle and D. Lemaire (eds.), *Building Evaluation Capacity*, Transaction Publishers, New Jersey, 1998.

The Operations Evaluation Department of the World Bank has initiated a series of short papers on evaluation capacity development (ECD). These present country case studies and also key ECD issues.

The series will help to present lessons learned in different contexts—what worked, what did not, and why—and thus contribute to a growing library of ECD experience. An active approach will be taken to disseminate the papers in this series to share these lessons with World Bank staff, with other development agencies, and with developing country governments.

The first four papers in the series are taken from this volume of seminar proceedings, and report the experience in Zimbabwe, Indonesia and Australia, as well as an overview of lessons from national experience. Additional papers will be prepared on other countries.

List of Authors and Discussants

Mark Baird

Mark Baird, New Zealand, obtained his MA (Hons) in economics from the University of Canterbury. He joined the World Bank in 1974 and has worked as a country economist on India, East Africa and Indonesia, including resident assignments in New Delhi and Jakarta. He returned to New Zealand for two years in 1989-91 as economic adviser to the Treasury. On his return to the World Bank, he was appointed Division Chief for Country Policy, Industry and Finance in the Operations Evaluation Department. In 1993 he was appointed to the post of Director of Development Policy in the Office of the Senior Vice President and Chief Economist. In this position, he was responsible for reviewing the World Bank's country strategies and adjustment operations and ensuring that research is linked to Bank policies and operational work. In 1997 Mr. Baird was made Vice President, Strategy and Resource Management. In this capacity, he is responsible for work in the Bank on strategic analysis and partnerships, planning and budgeting, and change management.

Alain Barbarie

Alain Barbarie completed a doctorate in experimental physics at the University of Ottawa in 1974, and the following year he obtained a diploma in public administration from Carleton University. From 1974 to 1991, he worked as a public servant with the Federal Government of Canada. During that time he held various management positions in a number of departments, including the Public Service Commission, the Ministry of State for Science and Technology, the Office of the Comptroller General, and the Research Branch of the Library of Parliament. He also worked with a number of international organisations, including the United Nations, the Organisation for Economic Co-operation and Development (OECD), and the International Institute for Applied Systems Analysis (IIASA). He concluded his career with the government as Director of Program Evaluation for the Department of Energy, Mines and Resources. In 1991 he became Deputy Project Manager for a Canadian International Development Agency (CIDA) project with the Ministry of Public Works (MPW) of Indonesia. Subsequently, in 1994, he worked with the National Development Planning Board (Bappenas) of Indonesia to help set up, with the assistance of a World Bank grant, a national performance evaluation policy, which was later promulgated in 1996. Since then Alain Barbarie has continued working, mostly in Indonesia where he lives, as a freelance consultant involved mainly in project benefit monitoring and evaluation work.

Gerald Britan

Gerald Britan is Director of USAID's Center for Development Information and Evaluation (CDIE), which leads the Agency in learning from experience through evaluation,

performance measurement, and development information activities. He received his Ph.D. in economic anthropology from Columbia University in 1974. After serving as a faculty member at Northwestern University and as Associate Dean of Graduate Studies and Director of Research at Southern Illinois University, he joined USAID in 1984, and was appointed to his present position in 1996. He has conducted fieldwork in the United States, Canada, Europe, Mexico, Africa, Asia, and the Near East. He has served on numerous US and international panels and task forces, including Vice President Gore's National Performance Review. During more than 20 years in government and academia, Gerald Britan has made important contributions to bureaucratic reform, program planning and evaluation, performance measurement, science and technology policy, and social and economic change. He has written four books, dozens of articles, and scores of professional papers and has served as a senior management and evaluation consultant to a variety of US and international agencies, organizations, and private firms.

Stephen Brushett

Stephen Brushett obtained his BA/MA in Economics from Cambridge University and his Msc in Business from London University. Mr. Brushett worked in Lesotho on industrial sector planning and project development. He has been with the Bank since 1982. He is currently the Senior Operations Officer, Transport (Eastern and Southern Africa). His present work involves managing major road sector investment program interventions in Malawi and Zambia and advising generally on road sector management and financing issues. His Bank experience includes infrastructure project development, public enterprise reform and public sector management reform with particular emphasis on planning and evaluation capacity building. Mr. Brushett was posted to Zimbabwe 1991-95 as Deputy Resident Representative. Since that time he has managed Bank support to evaluation capacity development in that country.

Stan Divorski

After receiving his Ph.D. in social psychology from Northwestern University, Stan Divorski served as a faculty member at the University of Manitoba and as a senior research officer, program evaluation manager and senior policy analyst in the Canadian government. As a member of the Office of the Auditor General of Canada, he led a government-wide audit of program evaluation and subsequent audits of performance reporting and managing for results in the Canadian government. He is currently assisting the US General Accounting Office in examining implementation of the GPRA. Stan Divorski is Associate Editor for *Evaluation: The International Journey of Theory, Research and Practice*.

Cheryl Gray

Cheryl Gray is the Interim Director of the Public Sector Group in the PREM Network. She joined the Bank in 1986 after receiving a law degree and Ph.D. from Harvard University and working for three years with the Indonesian Ministry of Finance on a major tax reform project. Her work from 1990 to 1997 focussed on legal reform and enterprise restructuring in the transition economies of Central and Eastern Europe—in particular, privatization and corporate governance, bankruptcy, and legal frameworks for private sector development and foreign investment. The group she currently heads oversees the Bank's work on public sector reform, including its work in public finance (expenditure and revenue analysis and management), civil service reform, decentralization, anticorruption, and legal/judicial reform.

Frans L. Leeuw

Frans L. Leeuw, Ph.D. (1953) is Dean of the Humanities & Social Sciences Department of the Netherlands Open University and Professor of Sociology, Utrecht University, the Netherlands. Earlier he was affiliated to the Netherlands National Audit Office as a director for Performance Audits and Program Evaluation and to Leyden University, the Netherlands. He is Vice-President of the European Evaluation Society, member of the Impact Assessment and Evaluation Group of CGIAR and member of several editorial board of journals.

Mario Marcel

A Chilean national, Mr. Marcel holds a M. Phil in Economics for the University of Cambridge, Great Britain and Bachelor's and Master's Degrees in Economics from the University of Chile. Mr. Marcel was a Research Fellow at the Corporation for Economic Research on Latin America (CIEPLAN). He has been the Budget Director at the Ministry of Finance in Chile; Executive Secretary of the Inter-Ministerial Committee for the Modernization of Public Management in Chile; and Visiting Research Fellow at the Inter-American Development Bank. He is currently the Executive Director for Chile and Ecuador at the Inter-American Development Bank in Washington, D. C. He has written several articles on the Chilean Economy, which have appeared in professional journals: the Reform of the State, Public Finance, Social Policy and Labor Economics are among the subjects.

Keith Mackay

Keith Mackay is a Senior Evaluation Officer in the Operations Evaluation Department of the World Bank. Before that he was an Assistant Secretary in the Australian Department of Finance, responsible for the government's evaluation strategy. He was also responsible for policy advice in telecommunications and science. Before that, he worked as a government economic researcher.

Elizabeth McAllister

Ms. McAllister, a Canadian national, joined the Bank from the Canadian International Development Agency in Ottawa (CIDA) where she has been Director General, Performance Review (Evaluation, Internal Audit and Results Based Management.) She joined CIDA in 1983, where she has held a variety of assignments both in Headquarters and in the field—Director, Women in Development (1983-85), Counselor, Development, Canadian Embassy in Indonesia (1985-88), Director, China Program (1989-92) and Director General, Strategic Policy and Planning for the Americas Branch (1992-94). Prior to joining CIDA, she worked for the Canada Employment and Immigration Commission where she developed the affirmative action program for federal jurisdiction employers. As Director, Operations Evaluation Department, Ms. McAllister manages and directs activities of the Department to provide an independent assessment of World Bank/IDA operations, policies and practices to satisfy the requirement of accountability to member countries; help improve Bank and borrower efficiency and effectiveness in designing and implementing development projects, programs and policies; and contribute to evaluation capacity development inside and outside the Bank. Ms. McAllister received a Master of Public Administration from Harvard University, a Certificat de la langue francaise from the Institut Catholique De Paris, a B.A. (Political Science) from the University of New Brunswick, and took courses in law, evaluation, and photography from Carleton University.

Robert Picciotto

Robert Picciotto is the Director-General, Operations Evaluation in the World Bank. He reports to the Board of Executive Directors and oversees the activities of the Bank's Operations Evaluation Department and of IFC's Operations Evaluation Group. He joined the International Finance Corporation (IFC) in 1962 and transferred to the Bank in 1964 where he worked as a development finance company analyst. He spent two years in New Delhi in 1967-69 as an agriculture economist, and on his return to headquarters, headed the Agriculture Industries Division. In 1970 he took over a Special Projects Division dealing with what was then East Pakistan (then Bangladesh). In 1972 he became Assistant Director, Agriculture and Rural Development in the Asia Region. From 1976 to 1986 he was Director of Regional Projects Departments. In 1987 Mr. Picciotto was appointed Director of Planning and Budgeting. In January 1990 he became Vice President, Corporate Planning and Budgeting. He assumed his current position in 1992.

Ray C. Rist

Ray C. Rist is the Evaluation Advisor for the Economic Development Institute within The World Bank. He is responsible for the evaluation of EDI's activities and performance. Mr. Rist has been active in evaluation for more than 20 years—as a professor, an administrator in both the Legislative and Executive Branches of the United States Government, and as a researcher. Mr. Rist has written or edited 23 books, the three most recent being *Can Governments Learn?* (1994), *Policy Evaluation: Linking Theory to Practice* (1995), and *Carrots, Sticks, and Sermons: Policy Instruments and Their Evaluation* (1998). He has written more than 125 articles and lectured in 43 countries.

David Shand

David Shand has been a consultant on public expenditure management with the Fiscal Affairs Department of the IMF for the past year. From 1993 to 1996 he worked in the Public Management Service of the OECD on issues of performance and financial management in the public sector. Prior to that he held senior budgeting and financial management positions in state and federal government in Australia.

Eduardo Wiesner

Eduardo Wiesner is Senior Partner in Wiesner and Associates in Bogotá, Colombia. His work focuses on public economics and institutional development. He has been a professor at the University de los Andes in Bogotá, where he graduated, *summa cum laude*, in Economics in 1959. He received a Master's degree in Economics from Stanford University in 1962. He has served as Minister of Finance and as Director of the Planning Department Colombia, (1978-1982). In Washington, D.C. he was Director of the Western Hemisphere Department in the International Monetary Fund (1983-1987), and a member of the Executive Board of the World Bank (1988-1990). He writes extensively on fiscal and political decentralization, evaluation, and institutional reform. He has been a consultant to the World Bank, the Inter-American Development Bank, the United Nations, the Government of Colombia, Coopers and Lybrand and other institutions.



THE WORLD BANK

1818 H Street, N.W.
Washington, D.C. 20433, U.S.A.
Telephone: 202-477-1234
Facsimile: 202-477-6391
Telex: MCI 64145 WORLDBANK
MCI 248423 WORLDBANK
World Wide Web: <http://www.worldbank.org/>

Operations Evaluation Department
Partnerships & Knowledge Programs (OEDPK)
Email: ecampbellpage@worldbank.org
Email: OED Help Desk@worldbank.org
Telephone: 202-473-4497
Facsimile: 202-522-3200

World Bank InfoShop
Email: pic@worldbank.org
Telephone: (202) 458-5454
Facsimile: (202) 522-1500

