Towards the Institutionalization of Monitoring and Evaluation Systems in Latin America and the Caribbean:

Proceedings of a World Bank / Inter–American Development Bank Conference

ERNESTO MAY   DAVID SHAND   KEITH MACKAY
FERNANDO ROJAS   JAIME SAAVEDRA
TOWARDS THE INSTITUTIONALIZATION OF MONITORING AND EVALUATION SYSTEMS IN LATIN AMERICA AND THE CARIBBEAN:

PROCEEDINGS OF A WORLD BANK/INTER–AMERICAN DEVELOPMENT BANK CONFERENCE
TOWARDS THE INSTITUTIONALIZATION OF MONITORING AND EVALUATION SYSTEMS IN LATIN AMERICA AND THE CARIBBEAN:

PROCEEDINGS OF A WORLD BANK/INTER–AMERICAN DEVELOPMENT BANK CONFERENCE

*Ernesto May*
*David Shand*
*Keith Mackay*
*Fernando Rojas*
*Jaime Saavedra*

Editors
This report was prepared in joint collaboration with the Inter-American Development Bank. The publication was principally financed by the Latin America and Caribbean Region’s Knowledge Management Group of the World Bank. The findings, interpretations, judgments and conclusions expressed in this publication are those of the authors and should not be attributed to the World Bank, to its affiliated organizations, or to members of the Board of Executive Directors or the governments they represent.

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

Rights and Permissions
The material in this publication is copyrighted. Copying and/or transmitting portions or all of this work without permission may be a violation of applicable law. The International Bank for Reconstruction and Development/The World Bank encourages dissemination of its work and will normally grant permission to reproduce portions of the work promptly.

For permission to photocopy or reprint any part of this work, please send a request with complete information to the Copyright Clearance Center Inc., 222 Rosewood Drive, Danvers, MA 01923, USA; Telephone: 978-750-8400; Fax: 978-750-4470; Internet: www.copyright.com.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Acronyms</td>
<td>vii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>ix</td>
</tr>
<tr>
<td>Foreword</td>
<td>xi</td>
</tr>
<tr>
<td>Welcome Addresses: Conference Overview</td>
<td>xiii</td>
</tr>
<tr>
<td><strong>PART ONE: SCENE-SETTING</strong></td>
<td></td>
</tr>
<tr>
<td>M&amp;E Systems and M&amp;E Tools</td>
<td>1</td>
</tr>
<tr>
<td>Impact Evaluation</td>
<td>7</td>
</tr>
<tr>
<td><strong>PART TWO: EXPERIENCE OF FIVE LATIN AMERICAN COUNTRIES - CHALLENGES AND OPPORTUNITIES</strong></td>
<td>9</td>
</tr>
<tr>
<td>The Chilean Experience</td>
<td>11</td>
</tr>
<tr>
<td>Country Presenter: Marcela Guzmán</td>
<td>11</td>
</tr>
<tr>
<td>Discussant: Verónica Silva</td>
<td>16</td>
</tr>
<tr>
<td>Discussant: Ignacio Irrázaval</td>
<td>17</td>
</tr>
<tr>
<td>Discussant: Keith Mackay</td>
<td>18</td>
</tr>
<tr>
<td>The Plano Plurianual and M&amp;E in Brazil</td>
<td>21</td>
</tr>
<tr>
<td>Country Presenter: Ariel Pares</td>
<td>21</td>
</tr>
<tr>
<td>Discussant: Romulo Paes de Sousa</td>
<td>25</td>
</tr>
<tr>
<td>Discussant: Marcos Holanda</td>
<td>26</td>
</tr>
<tr>
<td>Discussant: Yasuhiko Matsuda</td>
<td>26</td>
</tr>
<tr>
<td>Colombia’s National System for Evaluation of Management and Results</td>
<td>29</td>
</tr>
<tr>
<td>Country Presenter: Manuel Fernando Castro</td>
<td>29</td>
</tr>
<tr>
<td>Discussant: Carlos Pinzón</td>
<td>35</td>
</tr>
<tr>
<td>Discussant: Gladys Lopez-Acevedo</td>
<td>36</td>
</tr>
<tr>
<td>M&amp;E of Public Expenditures in Peru</td>
<td>37</td>
</tr>
<tr>
<td>Country Presenter: Fernando Zavala</td>
<td>37</td>
</tr>
<tr>
<td>Discussant: Carlos Rice</td>
<td>40</td>
</tr>
<tr>
<td>Discussant: Livia Benavides</td>
<td>41</td>
</tr>
<tr>
<td>Key Issues from First Day of Conference</td>
<td>43</td>
</tr>
<tr>
<td>Conference Rapporteur: David Shand</td>
<td>43</td>
</tr>
<tr>
<td>Conference Facilitator: Sonia Ospina</td>
<td>44</td>
</tr>
<tr>
<td>M&amp;E of Social Programs in Mexico</td>
<td>47</td>
</tr>
<tr>
<td>Country Presenter: Gonzalo Hernandez</td>
<td>47</td>
</tr>
<tr>
<td>Discussant: Guillermo Bernal Miranda</td>
<td>52</td>
</tr>
<tr>
<td>Discussant: Mark Hagerstrom</td>
<td>53</td>
</tr>
</tbody>
</table>
PART THREE: SYNTHESIS SESSION................................................................. 55

Institutionalizing Monitoring and Evaluation – Issues and Experience in OECD
Countries and in Latin America ................................................................. 57
   Conference Rapporteur: David Shand ..................................................... 57
   Panelist: Harry Hatry ............................................................................... 66
   Panelist: Philip Joyce ............................................................................ 68
   Panelist: Inder Ruprah .......................................................................... 70

Country Reactions: Reflections of the Five Country Presenters, and Views of Other Participants .......... 73
Closing Remarks ............................................................................. 77

PART FOUR: LAUNCHING AN M&E NETWORK FOR THE REGION ................................................................. 79
The Case for an M&E Network ................................................................. 81

PART FIVE: CONFERENCE CONCLUSIONS ................................................................. 87
Some Key Conclusions ......................................................................... 89
Closing Session .................................................................................. 93

Boxes

Box 1: Key Dimensions of Country M&E Systems...................................................... 4
Box 2: Brazil-M&E during the 2000-2003 Multi-year Plan: Critical Retrospective ...................... 22
Box 3: Brazil-System of M&E for the 2004-2007 Multi-Year Plan ...................................... 23
Box 4: Colombia-Achievements and Goals 2005 – 2010: Monitoring and Budgeting for Results ............................................................. 32
Box 5: Colombia-Achievements and Goals 2005 – 2010: Strategic Evaluations ..................... 33
Box 6: Colombia-Achievements and Goals 2005 – 2010: Accountability ................................. 34
Box 7: Peru-Evaluation Component ..................................................................... 39
Box 8: Peru-Government’s Evaluation System: Strategies & implementation ......................... 40
Box 9: Mexico-External Evaluations for Congress ....................................................... 49
Box 10: Mexico-External Evaluation: Strengths and Weaknesses ................................. 50
Box 11: Mexico-Goals .................................................................................... 52

Tables

Table 1: Chile-Performance Indicators —2005 ...................................................... 12
Table 2: Chile-Achievement of Performance Targets 2003 .................................................. 14
Table 3: Chile-Budget Effects of Evaluations 2000 to 2004 — Includes Evaluation of Government Programs and Impact Evaluations .................................................. 15
List of Acronyms

BGI  Comprehensive Management Reviews (Balance de Gestión Integral)
CI   Institutional Counterparts (Contrapartes Institucionales)
CLAD Latin-American Center for Development Management (Centro Latinoamericano de Administración para el Desarrollo)
CMA  M&E Commission (Comisión de M&E)
CONPES National Economic and Social Policy Council (Consejo Nacional de Política Económica y Social)-Colombia
DCG  Management Control Division (División de Control de Gestión)-Chile
DE   Strategic Definitions (Definiciones Estratégicas)
DF   Finance Department (Departamento de Finanzas)
DIPLAP Directorate of Planning and Budget (Dirección de Planificación y Presupuesto)
DIPRES Budget Directorate (Dirección de Presupuestos)-Chile
DNP  National Planning Department (Departamento Nacional de Planeación)-Colombia
EPG  Evaluation of Governmental Programs (Evaluación de Programas Gubernamentales)
FOSIS Social Investment and Solidarity Fund
GAO  Government Accountability Office
GDP  Gross Domestic Product
GPRA Government Performance and Results Act
IDB  Inter-American Development Bank
IMF  International Monetary Fund
INPAE Inter-American Network for Public Administration Education
LAC  Latin America and the Caribbean Region of the World Bank
M&E  Monitoring and Evaluation
MIDEPLAN Ministry of Planning and Cooperation (Ministerio de Planificación y Cooperación)-Chile
MOF  Ministry of Finance
NASPAA National Association of Public Management
NGO  Non-Governmental Organization
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NGP</td>
<td>New Public Management (Nueva Gestión Pública)</td>
</tr>
<tr>
<td>OECD</td>
<td>Organization for Economic Co-operation and Development</td>
</tr>
<tr>
<td>OED</td>
<td>Operations Evaluation Department of the World Bank</td>
</tr>
<tr>
<td>PART</td>
<td>Program Assessment Ratings Tool</td>
</tr>
<tr>
<td>PISA</td>
<td>Program for International Student Assessment (Programa para la Evaluación de Estudiantes Internacionales)</td>
</tr>
<tr>
<td>PMG</td>
<td>Management Improvement Program (Programa de Mejoramiento de Gestión)</td>
</tr>
<tr>
<td>PPA</td>
<td>Multi-Year Action Plan (Plano Plurianual de Acao)-Brazil</td>
</tr>
<tr>
<td>PPBS</td>
<td>Planning, Programming, Budgeting System</td>
</tr>
<tr>
<td>SEDESOL</td>
<td>Social Development Secretariat (Secretaría de Desarrollo Social)-Mexico</td>
</tr>
<tr>
<td>SIAF</td>
<td>Integrated Financial Management Information System (Sistema Integrado de Administración Financiera)</td>
</tr>
<tr>
<td>SIGOB</td>
<td>M&amp;E information through internet (Acceso en tiempo real a la información de M&amp;E via Internet)</td>
</tr>
<tr>
<td>SIGPLAN</td>
<td>Planning and Management Information System (Sistema de Informações Gerenciais e de Planejamento)-Brazil</td>
</tr>
<tr>
<td>SINERGIA</td>
<td>Evaluation System, Colombia</td>
</tr>
<tr>
<td>SNIP</td>
<td>National System for Public Investment (Sistema Nacional para la Inversión Publica)</td>
</tr>
<tr>
<td>SPPP</td>
<td>Sistema de Planificación, Programación y Presupuestación (PPBS, in English)</td>
</tr>
<tr>
<td>SSEGP</td>
<td>System for Monitoring and Evaluating Public Expenditures (Sistema de Seguimiento y Evaluación del Gasto Publico)</td>
</tr>
<tr>
<td>UMA</td>
<td>Unidades de M&amp;E (M&amp;E units)</td>
</tr>
<tr>
<td>WB</td>
<td>World Bank</td>
</tr>
<tr>
<td>WBI</td>
<td>World Bank Institute</td>
</tr>
</tbody>
</table>
THIS BOOK IS BASED ON the presentations and discussions at a conference jointly sponsored by the World Bank and the Inter-American Development Bank (IADB). The conference topic was “Towards the Institutionalization of Monitoring and Evaluation Systems in Latin America and the Caribbean”, and it was held on June 6-7, 2005, in Washington, D.C. The conference was organized by the Latin America and Caribbean (LAC) Region of the World Bank, under the leadership of Ernesto May, Sector Director for Poverty Reduction and Economic Management. The partnership support of the IADB, in particular Roberto Garcia Lopez, is gratefully acknowledged. Within the World Bank, the units which provided advice and financial support included, in addition to the LAC Region, the World Bank Institute, the Operations Evaluation Department, and the anchor unit of the Poverty Reduction and Economic Management Vice-Presidency.

The team which organized the conference comprised Jaime Saavedra, Fernando Rojas, Keith Mackay, Blanca Moreno-Dodson, Ruxandra Burdescu, Patricia Mendez, Anne Pillay, Azul del Villar and Lydia Ndebele. Special thanks are due to the logistics, information technology, and administrative support teams at the World Bank HQ and country offices.

David Shand was the conference rapporteur, and the conference facilitator was Sonia Ospina. These conference proceedings are based on transcripts as edited and interpreted by the editors Ernesto May, David Shand, Keith Mackay, Fernando Rojas, and Jaime Saavedra.

Finally, debt is owed to the speakers, discussants, and other participants of the conference, whose focus and interest in the topic of institutionalizing monitoring and evaluation systems ensured the conference was a success.
Many governments in the Latin America and Caribbean (LAC) region have gained an increased understanding of the value of monitoring and evaluation (M&E) to help both governments and donors alike better understand what public investments and interventions work well, which do not, and the reasons why. Monitoring and evaluating the performance of public programs and institutions can help increase their effectiveness, providing more accountability and transparency in how public monies are used, informing the budgetary process and the allocation of public resources, and assessing their effectiveness in attaining their desired objective such as improving welfare, reducing poverty or enhancing the equality of opportunities.

To further promote the institutionalization of M&E systems, the World Bank, in partnership with the Inter-American Development Bank (IADB), organized a regional conference on June 6-7, 2005 to take stock of the lessons learned. The participants represented finance and sector ministries from eleven countries, as well as experts from academia, bilateral donor agencies, consulting firms, and various sectors of the host institutions. Experiences from five countries, namely Chile, Colombia, Mexico, Brazil and Peru, were shared during the conference, and served as input to determine lessons learned in institutionalizing systems of monitoring and evaluation. The conference also served as a springboard to launch a regional network of policymakers and M&E practitioners to allow such national experiences to be shared much more widely and systematically throughout the LAC region.

It was evident from the country experiences presented that there is no single “destination” for countries in terms of what a well-performing M&E system looks like. Some countries stress a system of performance indicators, while others focus on carrying out evaluations (program reviews or rigorous impact evaluations). And while some countries have created a whole-of-government approach driven by finance or planning ministries, others are more focused on sector M&E systems. Yet the shared experience of these countries has led to some collective wisdom about the development of solid M&E systems, and these lessons are outlined in these proceedings. One key characteristic of most of the systems that are now at different stages of implementation in LAC, is that they reflect country-led—rather than donor-driven—efforts to institutionalize M&E.

It was evident from the June conference that a growing number of LAC countries are initiating efforts to strengthen and systematize their M&E functions. Senior officials from the eleven countries represented at the conference found highly valuable the sharing of their country experiences in institutionalizing M&E. They were impressed by the substantive progress achieved in several LAC countries in this field.

The proposal was, therefore, made at the conference to create a regional network of key stakeholders, to focus on the institutionalization of M&E and to facilitate South-South learning throughout the region. It was proposed that the network, which will be the first of its kind in the developing world, will be open and flexible, attracting decision-makers, practitioners, experts in and outside the government; high-level officials from sector ministries, finance ministries, and planning departments; parliamentarians and their advisors; academics; consultants and experts; multilateral organizations and interested bilateral donors. Network activities are likely to include a virtual discussion space, staff secondments, joint training on M&E, study tours to neighboring countries, and follow-up regional and subregional conferences.
Perhaps the principal value-added of the proposed network will be to create a forum for dialogue among M&E system managers, and between them and M&E practitioners — just as the conference was designed to allow them to share their own country experiences with each other, and to provide them with additional regional and some global experience. The World Bank and IADB are envisioned only as facilitators, so that country ownership is preserved and assured.

We all look forward to working together in further enhancing the M&E systems in the LAC region.

Ernesto May
Sector Director
Poverty Reduction and Economic Management
Latin America and the Caribbean Region
The World Bank
Monitoring and evaluation is a very important input to the process of reaching the millennium development goals and in reducing poverty. It is important for a number of reasons. First, because it allows us to track where we are. In doing so, it allows us to use that knowledge in mid-course corrections, whether in the re-design of programs or, more importantly, in the change of practices. I am going to highlight this issue of learning from practice, because the practice of monitoring and evaluating has been evolving.

Evaluation has been found to be necessary not only in our day to day work but also in the longer term assessment of how organizations are performing. One thing I have learned is what has actually been happening in Latin America, which is not very well known outside of the region or the individual countries in which these achievements have been institutionalized.

So we need to learn from practitioners so that we can go from understanding the issues of monitoring and evaluation to a process where we can exchange those ideas within a particular country project or program.

As you may know, our new World Bank President has called for us to become an ideas bank, to go beyond finance and beyond knowledge, to be a place where ideas get exchanged and ideas get developed.

At the Shanghai Global Learning Conference, which was held exactly a year ago, we looked at over a hundred case studies. One of the key features in achieving significant poverty reduction was the approach that program or project managers or the countries took in monitoring and evaluation.

One of the case studies presented was the evaluation of Mexico’s Oportunidades, which we will have a chance to hear more about at this conference. Its methodology was highlighted by many of the people attending the Shanghai conference as one of the ways in which we can use evaluation to achieve significant poverty reduction.

At the World Bank Institute, we like to look at these specific examples not as episodic or country-specific issues, but to see how one can extract things that can be replicated or changed and used in a different way in other settings.

Also, we would like to look at the implications for the budgetary process in countries, if they do this right; and what are the implications for the Bank in the way it thinks about monitoring and evaluation, and what are the implications for other international financial institutions.

In achieving the millennium development goals and in moving towards significant poverty reduction, we need to make sure that the tracking of progress moves hand in hand with thinking about where we go next.

So we would like to enrich the knowledge that has already accumulated from the Shanghai conference in terms of what works, what doesn’t, and why. By adding this evaluative dimension we look for the evidence of how one gets final impact on poverty reduction.

We would like to take the knowledge sharing from this conference one step further and not just leave it as a two-day talking event, but to end up with a live community of practice. Etienne Wanger defines communities of practice as groups that are informally bound together because they have shared expertise and passion for a joint enterprise. I know it is hard to talk about passion when it comes to monitoring and evaluation, because it is a rather tough, hard-nosed sector, but there is passion in measures and people who are passionate about doing them well. So I am hoping that that passion translates into a solid community of practice that can take this forward.
John Seely Brown has a different definition. He calls it peers in the execution of real work. What holds them together is a common sense of purpose and a need to know what each other knows. So that is another expectation I have for these two days — that we push beyond talking to each other about what we know, to finding a way to learn and engage beyond the presentations.

A number of key features are important in success of communities of practice. One is focusing on topics that are important to the business of the community members. A second is to find a well-respected community member who can coordinate the effort. A third is making sure that people have the time and are encouraged by their home organizations to engage, and to get key resource persons involved, so that the knowledge moves at the same time as the sharing of practice.

But there are three other aspects to a well-functioning community of practice — building personal relationships so that an active and a passionate group is developed; creating fora for thinking together and systems for sharing information, so that it is easy to contribute toward this community’s knowledge and practice; and the most important one is an opportunity for real dialogue about cutting-edge issues.

So I just wanted to give you some of the expectations of what a well-functioning community of practice would be, and to conclude by saying that the World Bank Institute will be supporting this initiative. On behalf of the World Bank and the World Bank Institute, I would like to wish you a successful conference.

**Ajay Chhibber**
*Acting Director–General, Operations Evaluation Department, The World Bank*

It is really excellent to see a conference such as this with such senior representation from so many countries in Latin America, from the donors and from the academic community.

It is also very good to see such strong support from various parts of the World Bank group, including from the LAC region, the World Bank Institute, the Poverty Reduction and Economic Management Network, and the Operations Policy and Country Services unit. This reflects the priority that we individually and collectively give to this very important topic.

Let me share with you the perspective from the Bank’s Operations Evaluation Department — the Bank’s independent evaluation arm, which has about 70 evaluation professionals. Our mandate is to evaluate the Bank’s operational work at various levels — at the project, sector, country and global levels. OED reports directly to the Bank’s board.

The Bank’s operational areas conduct self-evaluations of their work, and OED evaluates these self-evaluations. This provides a consistency check and a second opinion. Our role within the Bank, and as the largest evaluation unit in the donor community, allows us to see regional and global trends in the emphasis given to the strengthening of government monitoring and evaluation systems.

We also have a mandate to support country efforts to strengthen their monitoring and evaluation systems. We typically do this in very close partnership with the Bank’s country teams. In fact, the Bank is currently working to support over 30 governments to help in strengthening their M&E systems through loans and grants, and OED has provided support to most of these countries, as well.

The interest in this topic in Latin America reflects country-driven demand, not prompted by donor pressures, and I understand that there is at least one Latin American country which represents world class best practice for government M&E systems.

So what are the reasons for this strong interest and the priority that is given to monitoring and evaluation? One is pressure from countries themselves, which are seeing three main uses of the findings and information which M&E provides. First, to aid budget resource allocation decisions, particularly in a climate of continuing budget constraints. Second, in response to accountability pressures, especially from presidents who have committed their administrations to achieve real results. Thirdly, to support learning about what works well, what doesn’t, and why, and this helps ongoing management and service delivery.

We have also observed that middle-income governments in the Latin American Region are becoming increasingly demanding. They can borrow from the Bank, from other donors, or the private capital market. So they require fairly solid evidence about the effectiveness of development interventions. This has resulted in a growing number of rigorous impact evaluations now being done jointly between the Bank and countries, especially in the human development area. We can expect this pressure from coun-
tries to have a flow-on effect to the Bank, to help the Bank's own operational areas improve the quality of their self-evaluations. OED views this as a priority for the Bank itself, because of our finding that weak monitoring and evaluation is a contributing factor to poor project performance in 38 percent of Bank projects which are rated unsatisfactory or highly unsatisfactory.

We are also seeing growing commitment and support for government M&E systems from the donor community. This has several sources, including a growing realization that M&E is an important dimension of sound governance. Also the Bank and other donors are being pressured by our stakeholders to demonstrate greater results and to be accountable for larger aid volumes in the context of debt relief for the poorest countries.

And there is an understanding that strengthened government M&E systems will add to the growing body of evaluative evidence about development work. Thus, there is a growing library of evaluation findings, and this was also echoed at the Shanghai conference.

This should also lead to a more receptive environment for the Bank's own evaluations.

Finally, there is also an element of self interest: better government M&E systems and country capacities help donors do their own M&E work.

Since its creation over 30 years ago, OED has been a consistent advocate and supporter of countries' efforts to strengthen their M&E systems. This experience has yielded many lessons about M&E capacity building, and these are reflected in the valuable collection of resource materials which are publicly available. We also do self-evaluations of our own capacity building work, which are also available.

All of us — the World Bank, IADB, other donors, Latin American governments, and governments in other regions, can learn a lot from the five countries that are being showcased at this conference.

A question which I would pose is how best the donor community can work with and support countries in the region, both those which have already institutionalized their M&E systems and those which would like to do so.

Let me end with a saying which is in some way associated with Einstein, which goes something like, 'What often counts is not counted, and what is counted often doesn't count.' I think that is an appropriate thought for this conference going forward.

---

**Ernesto May**

*Director of Poverty Reduction and Economic Management in the Latin America and the Caribbean Region of the World Bank*

Back in 1997, the Bank produced the World Development Report: The State in a Changing World. That report, based on a survey of 3,600 local entrepreneurs in 69 countries, produced an index of credibility of institutions and then based on that index showed how both investment and growth in our countries is very much linked to the quality of institutions. That work shows that a seven to eight percentage point difference in investment to GDP is associated with the difference between high-quality governance institutions and low-quality governance institutions in our countries, and in terms of per capita GDP growth a difference of three percentage points.

So quality of institutions clearly has enormous impact on the results that we are seeing in our countries. Thus it is very fitting to now look at the experience that has been accumulating in Latin America.

We will have the case studies of five different countries. In the case of Chile, probably the country in our region that marks the frontier in terms of monitoring and evaluation systems, there is a 30-year accumulated experience in evaluation of government programs. As some of you know, the Bank was asked by the Chilean Government to evaluate their national evaluation system and we will be hearing about that work.

The national evaluation system in Colombia goes back more than a decade when the Bank was asked by the Department of National Planning to review the design of their evaluation system, SINERGIA, and to provide recommendations to help Colombia to move to a results-based management system. I had the fortune of leading the Bank team, composed of people like Graham Scott from New Zealand, and Harry Hatry, from the U.S. I am looking forward to seeing what has happened in these intervening years in terms of the development of SINERGIA.

In the case of Mexico, last year I was there for a conference on institutionalizing social policy for poverty reduction. Mexico has the experience of the evaluation of Progresa (now called Oportunidades). This evaluation is internationally known, and it has really set the basis for Mexico to move towards a more systematic framework for monitoring and evaluation, an effort that is being led by SEDESOL.
We will also be hearing where we stand at this stage in the efforts in this country.

We will also hear from the new experience being gathered in the case of Peru, and Brazil’s work in moving to a results-based public sector management, through its well-known multi-year planning system.

So an enormous amount of experience is being accumulated in the region. We are here to share those experiences, and to reflect on them in terms of how to move forward.

A key issue for discussion is the idea of forming a regional network of monitoring and evaluation people who are directly engaged in the day-to-day work of advancing the M&E systems in our region — with national experts, international experts, government officials, the donor community, and international financial institutions involved in this effort.

We all look forward to the outcome of this conference and to support the continued efforts to enhance the M&E systems in the LAC region.
PART ONE:

Scene-Setting
M&E Systems and M&E Tools

Keith Mackay  
Senior Evaluation Officer, Operations Evaluation Department of the World Bank

In this session I would like to help set the scene for the country presentations we will hear next. We have asked country presenters to be our guides for each country, to lead us through the often complex structures of government, and to distinguish between what exists only on paper, and what is real. We have asked them to tell us which parts of their M&E systems are working well, which are not, to explain the reasons for the progress achieved in each county so far, and to outline future plans. To help focus these presentations a list of key dimensions of country M&E systems has been prepared (Box 1).

The five countries on which this conference is focused — Brazil, Chile, Colombia, Mexico and Peru — can be considered the leading countries in Latin America, in terms of the progress they have made in institutionalizing their M&E systems. The main presenter for each country is a senior official who is highly experienced in this area, and is able to present an insider’s perspective. For each country presentation, we have asked a second country representative, from either a central ministry or a sectoral ministry to be a discussant; we have also arranged for an experienced World Bank staffer to be another discussant. This approach, one presenter and two (or three) discussants, provides a diversity of perspectives on each country; in effect, a ‘triangulated’ approach.

Most of the five countries emphasize whole-of-government M&E systems. Of course, it is also possible to have a sound M&E system at the sectoral level, such as the M&E system of Mexico’s Ministry of Social Development (SEDESOL). However, irrespective of whether the M&E system is anchored at the whole-of-government or sectoral levels, the relationships between central and sectoral ministries are important in determining the way in which the system functions, and its effectiveness.

This is not a conference about M&E techniques; instead, the focus is on the institutionalization of M&E systems in Latin America. However, during the conference we will hear about many M&E concepts, tools and methods and this can be confusing. Thus we may hear about ‘results-orientation’, results-based management (RBM), results-oriented management, performance management, performance-based budgeting, performance-informed budgeting, and so on. All of these are related, and we hope that presenters and discussants will clarify exactly which concepts they are talking about.

Despite this range of concepts, the main proposition of an M&E system is to use monitoring information and evaluation findings for one or more purposes:

1. To inform budget decision-making, by providing information on the actual or likely performance of government programs;
2. To support government planning, such as in the development of national plans;
3. To help the ongoing management of government programs and activities, by providing information on their efficiency and effectiveness; and
4. To underpin accountability relationships — accountability to the President’s office, to the finance ministry (the hacienda), to the planning ministry, accountability within ministries, to Congress, and to the people.
A key objective of the conference is to take stock of the lessons from the more successful monitoring and evaluation (M&E) systems currently implemented in some LAC countries: Brazil, Chile, Colombia, Mexico and Peru. These countries have undertaken significant steps but are at different stages of the process. The conference is being designed as a first step in the creation of a regional network involving the main actors/specialists from these and other countries in LAC, as well as Bank experts and other interested stakeholders. The conference provides the opportunity to engage in frank, open and honest dialogue among peers — sharing with others the lessons about the strengths and weaknesses of each country’s M&E system, i.e. which aspects work well, which do not, and the reasons why.

The country sessions comprise a main presentation, followed by a country discussant and a Bank discussant. The country presentations focus on the key issues only, with little emphasis on descriptive issues or historical development of the M&E system, or on inventories of public sector reforms. The following list comprises the key dimensions of each country’s M&E system:

1. **Genesis of the M&E system**: role of M&E advocates or champions; key events which created the priority for M&E information (e.g., constitutional amendment, election of reform-oriented government, fiscal crisis).

2. **The ministry or agency responsible for managing the M&E system**: Roles and responsibilities of the main parties to the M&E system — e.g., hacienda, planning ministry, president’s office, sector ministries, Congress. Incentives for these stakeholders to take M&E seriously — strength of demand for M&E information. Possible existence of several, uncoordinated M&E systems, at the national and sectoral levels. Importance of federal/state/local issues to the M&E system.

3. **The main aspects of public sector management which the M&E system supports strongly**: (i) budget decision-making; (ii) national or sectoral planning; (iii) program management; (iv) accountability relationships (to the hacienda, to the President’s office, to Congress, to sector ministries, to civil society).

4. **Actual role of M&E information at the various stages of the budget process**: such as policy advising and planning; budget decision-making; performance review and reporting. Possible disconnect between the M&E work of sector ministries and the use of such information in the budget process. Existence of any disconnect between the budget process and national planning. Opportunities to strengthen the role of M&E in the budget.

5. **Extent to which the M&E information commissioned by key stakeholders (e.g. hacienda) is used by others, such as sector ministries. If not, what are the barriers to utilization? Any solid evidence concerning the extent of utilization by different stakeholders (e.g., a diagnostic review or a survey). Examples of major evaluations which have been highly influential with the government.**

6. **Types of M&E tool which are emphasized in the M&E system**: regular performance indicators; rapid reviews or evaluations; performance audits; rigorous, in-depth impact evaluations; other. Cost of each of these types of M&E. Manner in which evaluation priorities are set — are they focused on ‘problem programs’, pilot programs, high-expenditure or high-visibility programs, or are they based on a systematic research agenda to answer questions about program effectiveness?

7. **Who is responsible for collecting performance information, and for conducting evaluations (e.g., academia or consulting firms)? Any problems with data reliability, or with the quality of evaluations which have been conducted. Strengths and weaknesses of local supply of M&E. Key capacity constraints and the government’s capacity-building priorities.**

8. **Conclusions: overall strengths and weaknesses of the M&E system. Its sustainability, in terms for example of vulnerability to a change in government. How dependent is it on donor funding or other support? Plans for future strengthening of the M&E system.**
During the conference we will see that the five countries differ considerably in the emphasis they place on these potential uses of M&E information. Another difference between countries is the emphasis placed on different types and methods of M&E. For performance monitoring through reliance on performance indicators, an M&E system can focus on part or all of the results chain — from inputs to processes to outputs to outcomes to impacts. And for evaluation, there exists a range of available types and methods. These include, among others, relatively straightforward program reviews (which rely, for example, on desk review of documents, on interviews of key stakeholders, focus groups, and use of whatever program data are available). Evaluation also includes other types such as performance audits, and rigorous impact evaluations (which use control groups, and sophisticated data analysis). The next speaker, Ariel Fiszbein, will talk in greater depth about this latter technique. Evaluations have many uses, and can focus on operational or management efficiency, on the targeting effectiveness of programs, and on impacts such as learning gain or health improvement.

So, conference participants should be careful in their listening. The term ‘M&E’ is often used in a generic sense, almost as a mantra. Thus when speakers talk about their country’s experience, it is important to be clear which types of M&E are being discussed. Each type differs in its applicability and uses; each has strengths and limitations, so we should use the various M&E types and methods appropriately, according to the circumstances.

The final message from this scene-setting session is likely to be repeated many times at this conference: that to institutionalize an M&E system does not involve a straightforward ‘technical fix’. Creating a working M&E system requires more than relatively straightforward measures such as issuing a Presidential Decree and arranging M&E training courses for officials. Such measures are helpful, but they are not sufficient. It is essential to have high-level commitment and ongoing support from powerful champions at the ministerial and senior official levels. And strong incentives — both to conduct M&E, and to use M&E information, — are needed if ministries are to take it seriously. It is also necessary that staff or consultants are able to conduct good quality M&E. And staff and managers must be able to use the M&E information which is produced, as part of their day-to-day work: for budget preparation; for policy-advising; for planning; for ongoing management; and for operational activities.

In understanding these demand and supply prerequisites for a workable, sustainable M&E system, it is important to bear in mind that conservative ministries and staff may resist efforts to conduct M&E and to use it. In summary, to make an M&E system work well requires strong demand for M&E and strong supply — skills, good-quality data, and reliable evaluations, and appropriate M&E coordination and infrastructure arrangements.
Impact Evaluation

Ariel Fiszbein
Lead Economist, Human Development, Latin America and the Caribbean Region of the World Bank, and Coordinator of the Development Impact Initiative

There ought to be a very close link between monitoring and evaluation. But I will be focusing on evaluation, specifically on impact evaluation, because it has a critical role in helping us learn about effective development interventions.

In looking for the right description of impact evaluation, I found it not in a technical book, but rather in a book by the Italian author Primo Levi. Even though he wasn’t talking about impact evaluation, his reference to being suspicious of prophets and to reveal and seek the truth through step by step processes with no shortcuts, focusing on things that can be verified and demonstrated, provides exactly the argumentation as to why we need to invest much more in impact evaluation.

The argument is quite simple. The reason why impact evaluations are a primary way of verifying if a particular approach to solving a development challenge works or does not work or under what conditions, is because it is comparing outcomes between people who have been treated by that intervention and people who have not. The challenge can be reducing infant mortality or getting kids to learn when they go to school or things of that sort. Rather than having isolated impact evaluations, we have a set of impact evaluations that evaluate similar approaches under different conditions.

And there is an additional benefit. In monitoring, we can feel much more comfortable using performance benchmarks that are derived from a set of good impact evaluations. Furthermore, it provides the basis for doing what we ought to do but seldom do, which is carrying out cost-effectiveness assessment, in terms of determining what is the cheapest and most effective way of, say, reducing infant mortality rates.

The term evidence-based policy design is quite popular in the United States and it comes from evidence-based medicine. It is essentially the practice of regularly putting the issues of design and adoption of particular programs and policies through the filter of systematic, robust impact evaluation studies.

In this presentation, I want firstly to outline briefly what is happening with impact evaluation in what we call the development world (which includes both developing countries and the development community more broadly), and then put forward a few ideas on the challenges in developing a much more systematic approach to development impact evaluation.

So where do we stand today? There is much more impact evaluation work than many people think. In the last year or so, there have even been articles in the New York Times complaining about the lack of impact evaluation. When we take stock we see that Latin America is ahead of many other regions. Some people argue that this reflects the higher level of development in the region, and that this has created a big push for evaluation.

To put some numbers on things that we have done recently, we took just one department in the World Bank, the one that I happen to work in, and we looked at all the
investment programs that have been approved during the last six years. We asked whether we really know whether these programs are working and whether there had been an impact evaluation. We found that more than half of those programs had an impact evaluation and these covered 80 percent of the total spending. Then we started building a database of evaluations for the Bank. Currently, the database has 63 individual evaluations, 42 of them from Latin America.

This indicates that there is something going on, not only in terms of what governments are doing, but also independent evaluation. But independent impact evaluation is still very much driven by northern researchers who have gone South, although local impact evaluation is growing.

We also observe a huge diversity of methods. Among experts, there is a heated discussion about different experimental methods but now we have an arsenal of methods being used in practice.

Another interesting aspect is that we are now evaluating things that we didn't think could be evaluated. Last week I was reading a couple of papers evaluating the impact that the changes in election rules in India that reserved positions for women have had on social outcomes. Who would have thought that one could apply rigorous impact evaluation methods to issues such as this?

But we are still far from using the full potential of this tool. We are now looking at the 100-plus case studies of poverty-reduction initiatives that were represented at the 2004 Shanghai conference on poverty reduction, many of which are still work in progress. We found that a little less than 20 percent of those cases either had a completed impact evaluation or one in progress with results expected very soon. But a quick overview reveals at least twice as many cases could easily have been evaluated, in the sense that there are similar programs that have been evaluated in other countries.

One of my colleagues described the situation as follows. He said, ‘Well, we know a lot about a few things, and little about a lot of things. We somehow get fixated on a few interventions and go deeply into those, but there are other big and important areas that we haven’t focused on.

Even when we have results of impact evaluations, we don’t use them the way we should. One example is the failure to use results from impact evaluation as benchmarks for monitoring the performance of similar programs.

What would it take to have a global learning program on effective development interventions? What would we need? I suggest working backwards from key development challenges — one example in Latin America is educational learning in schools. We all know the countries in Latin America that are now performing well. We therefore need to work backwards in evaluating the approaches that may yield those results. We need experimentation and evaluation, leading to analysis of what are the cost-effective ways to achieve those outcomes.

You would think that this is what an institution such as the World Bank ought to be doing. This is a big challenge but the bottom line is that this type of evaluation is an international public good. As a result, there is under-supply. Some of the reasons relate to who bears the costs and who receives the benefits of these evaluations. This implies subsidizing or lowering the cost of evaluation to governments, building evaluation capacity, finding ways of coordinating evaluation programs, asking the right questions and using promising approaches; we also need to ensure that money is not wasted on evaluating less important issues.

Over the last year or so, we have initiated a program in the Bank to better coordinate our impact evaluations, not only increasing the number of evaluations but having a more strategic approach to the demands from the countries we work with, using promising approaches to key development challenges, and then setting the basis for eventual evaluations and for meta analysis of the collection of individual country studies. There has been a very positive reaction to this program not only within the Bank, but also externally.

Over the coming years we can expect to see the beginning of a global learning partnership. But who will be driving this effort? Will it be driven by donors, in terms of deciding what gets evaluated, how it is evaluated, who does it, and how it is used? Or is it going to be a truly global partnership, where the voice of the South is adequately represented in the evaluation agenda? At this point in time, the level of participation by the South is not there, and there aren’t adequate mechanisms to ensure its strong participation.
PART TWO:

Experience of Five Latin American Countries—Challenges and Opportunities
The Chilean Experience

Country Presenter: Marcela Guzmán
Chief of the Management Control Division (DCG) of the Budget Office of the Chilean Ministry of Finance. Head of the Chilean Management Control System and Results-based Budgeting

Our management control system and results-based budgeting originated as a need in the budgetary process. Hence, this system was developed at a central ministry level — the Ministry of Finance. The system provides a link between how allocated resources were used, and subsequent decisions on the allocation of funds. Our system has a very simple objective, which is to obtain performance information and introduce better practices to improve the quality of expenditure. We focus on resource allocation decisions and on improving the management of the institutions handling those resources — to improve the cost-effectiveness of the resources allocated to different government initiatives. A third objective is transparency, to deliver better public information to the political and social actors involved in the discussion of public policies in Chile.

Over the years the system has been strengthened by a set of instruments that are linked by common conceptual elements and procedures. That is why we call it a system. All these different instruments contribute to the working of the model and there are synergies among them. The model has been built gradually. The first initiatives emerged at the beginning of the 1990s, and over time we have been adding to the development and sophistication of the different instruments.

The basis of our model is the traditional budget cycle — budget formulation; discussion and approval; execution; and evaluation. Different performance-related instruments, including performance indicators and evaluation, are relevant at different stages of this cycle.

Performance indicators and evaluation are instruments for collecting information on performance but are different in their methodological aspects and uses. However, both have common elements and complement each other. Performance indicators are measurements containing mathematical logarithms to measure the results. Evaluation, on the other hand, is a study and is more comprehensive than a simple measurement, although it includes performance indicators. While distinguishing the methodology and the scope of each of these instruments, we have used them in a complementary manner for better decision-making.

For evaluation, we have developed over time different methodological approaches or instruments to constitute an evaluation agenda. Each of these instruments is linked with the budgetary cycle, but in different forms and at different times — but all are linked with analysis, review, and the decision-making process. When we speak of performance indicators and evaluation in Chile, we are referring to evaluations of existing programs, which already have budget funding. Another instrument — presentation of initiatives to the budget (Presentación de Iniciativas al Presupuesto) — which has its own methodological and conceptual elements, focuses on the review of new initiatives or sometimes on the expansion of existing initiatives for budgetary funding. It also contains conceptual elements of the other instruments, and covers another part of the decision-making process but is also part of the model. Finally, the comprehensive management reviews (Balance de Gestión Integral, BGI) are annual
reports of the institutions, including financial statements, that cover both financial and non-financial performance issues. These are carried out by the institutions themselves but coordinated by the Ministry of Finance. These instruments are also used in the work of the Ministry of Finance. The management improvement program (Programa de Mejoramiento de Gestión, PMG) has a more indirect link with the budgetary decision-making process, and consists of monetary incentives for officials, based on achievement of certain management system development objectives.

Performance information — developed through the performance indicators instrument, evaluation of institutional programs, and the comprehensive annual reviews — is used in our budgetary cycle. We can illustrate the working cycle of the Ministry of Finance, through three events: the formulation of the budget; the presentation of the draft legislation, its discussion in the Congress, and its approval; and subsequently the execution of that budget.

In Chile, a preliminary stage in the budget preparation is known as the internal evaluation of budgetary execution. It is internal because it is part of the standard work of the Ministry of Finance. It has a timetable, a work program, work procedures, as well as a methodology. At this stage of analysis and review, all the information from each ministry is comprehensively reviewed, including financial and all available performance information — performance indicators, results of evaluated programs and institutions, and the institutional reports on comprehensive management. This allows a more complete and integrated vision of the ministry’s financial and performance results. Our goal is to achieve a sound allocation of resources and to bring about necessary management improvements, to enable each program to produce the required results.

However, one of the difficulties here is relating actual expenditures to specific objectives. Our model, and thus the performance information it collects, is not directly or mechanically used in the budget. The Ministry of Finance does not have a mathematical model where, on the one hand, it inputs the performance indicators and the evaluation results, and on the other gets the decisions on resource allocations. This is because the budget is a political as well as a technical process, which uses information beyond performance information.

On many occasions programs do not show good results, based on the evaluations or performance indicators. Here the solution is not necessarily to cut back the resources or to eliminate a program, but to discuss and analyze the actions needed so that the program can effectively achieve the expected results. Hence, the use of this information does not necessarily imply that more or less resources will be allocated to a certain program. Naturally, the decisions on resource allocation include other factors, for example, the resources available for public expenditure each year.

**Performance indicators**

Performance indicators collect information over time and thus comparisons are mainly with past performance. Comparing performance results among public programs is a problem, because programs have different objectives, making it very difficult to find uniform outputs to use for comparisons and benchmarking. In our case, performance indicators focus on the different performance dimensions of efficiency, effectiveness, economy, and quality (Table 1). We have separated quality because it can also be understood as a concept of effectiveness, and to reflect its importance as one of the policy priorities of the government.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Processes</th>
<th>Products</th>
<th>Results</th>
<th>Total Number</th>
<th>Participation %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effectiveness</td>
<td>46%</td>
<td>46%</td>
<td>91%</td>
<td>861</td>
<td>54%</td>
</tr>
<tr>
<td>Efficiency</td>
<td>13%</td>
<td>13%</td>
<td>2%</td>
<td>177</td>
<td>11%</td>
</tr>
<tr>
<td>Economy</td>
<td>34%</td>
<td>4%</td>
<td>7%</td>
<td>191</td>
<td>12%</td>
</tr>
<tr>
<td>Quality</td>
<td>7%</td>
<td>37%</td>
<td>0%</td>
<td>359</td>
<td>23%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>1,588</td>
<td>100%</td>
</tr>
<tr>
<td>Total Number</td>
<td>396</td>
<td>901</td>
<td>291</td>
<td>1,588</td>
<td>100%</td>
</tr>
<tr>
<td>Participation%</td>
<td>25%</td>
<td>57%</td>
<td>18%</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

Our indicators are also used in different areas of control, such as processes, products and results, distinguishing the intermediate results from the final results. This is because, as has been pointed out, impact indicators or the outcomes are not easy to obtain, they require more resources and more complex methodologies, and they cannot be measured regularly, say annually. The programs need to have a certain maturity before we can begin to measure them. As a result,
in the annual budgetary work, we need to collect information from other areas to know the true impact of a program. Hence on occasions we use indicators that are not strictly outcome indicators.

Our performance indicators meet traditional requirements such as relevance, verifiability and measurability at reasonable cost. We have a work matrix covering the indicators which are used in discussions with the institutions during the budgetary process. This matrix is widely known because we have been working with ministries for more than 6 years. It includes the core information on what is being measured, the historical values of these indicators, and the goals that each ministry commits to for the following budgetary year.

There is also a work process within the institutions that is part of the budgetary preparation. It starts with the discussion of the indicators that will be formulated each year and finishes with the definition of the goals associated with the indicators, linked to the resources that are finally agreed upon in the budget to be delivered to Congress along with the Draft Budget Law each year. In the following year, the performance results reflected in the indicators are compared with the commitments made. This information is also sent to Congress and to the institutions, and is analyzed in the Ministry of Finance.

The main actors in this work are the institutions themselves, since the indicators originate with them. The Budget Office of the Ministry of Finance is responsible for coordinating this work and giving technical assistance, presenting these indicators to the Congress and making use of this information in its internal work. Congress receives this information for use in its budgetary discussions.

By 2005 we had developed 1,588 indicators. This gives the impression of being an enormous number, and it is indeed a big system. This is the maximum number of indicators that we are capable of managing. But this number represents approximately three indicators per institution. When one looks at three indicators per institution for an institution that may have 15 or more programs, the number no longer appears to be so big. In general, when the information is being used and the management of a ministry is being analyzed, one has the feeling that these three indicators are not enough to provide a comprehensive vision of the institution.

The challenge we face is to limit the size of the system for administrative purposes, while moving forward on the relevant indicators and recognizing the difficulties of impact evaluation, and advancing towards needed outcome indicators. This may involve focusing more strongly on those intermediate outcome indicators that are easier to develop and that can be evaluated in shorter periods during the implementation of the programs.

Our indicators are mainly focused on the products or outputs of goods and services. Many also focus on quality of the products such as the satisfaction of users, the timeliness of the delivery and other factors related to quality — either the perceptions of users or the technical quality. Having a large number of product or output indicators is not seen as something negative, given the difficulty of developing impact indicators. Rather, this provides us with information on some key aspects of performance.

To be reliable as a management tool, results have to be assessed using the indicators. Where indicators are formulated but the performance information they provide is not subsequently utilized, it is unlikely they will ever be institutionalized. Assessment of performance information is not an easy process. In our case, we assess the performance information a year after it was collected. Also in some cases performance information can not be assessed because the indicators are programmatic and their measurement period is more than one year. What is important is that everything that should be assessed, does get assessed and under the highest standards.

Our assessment of the indicators for 2003 reveals that achievement of performance targets (‘indicator compliance’) was between 95 and 100 per cent, in over three-quarters (76 per cent) of the institutions (Table 2). To us this level represents achievement of the targets because it is very difficult for an institution to meet 100 per cent of its goals. The indicators are also complex and thus it is difficult to do a detailed analysis with respect to the achievement of a goal. Hence, we believe that a level of performance of at least 95 per cent means that an indicator target has been achieved.

What are some of the reasons for non-achievement of the indicator targets? There are external factors that the institutions cannot control, and in the assessment of performance these factors are reviewed, analyzed and allowed for. In some cases goals have been set too high due to inadequate back-
ground information or insufficient past experience when we are developing a new program. Performance information in these cases promotes and generates the improvement of information systems. There may also be internal management problems causing non-achievement of the targets. When working with indicators, this type of analysis must be performed, otherwise the indicators may be used incorrectly.

In terms of what remains to be done, I have already pointed out the need to improve the availability of performance information and the need to internalize its use in public institutions. There is still work to be done with each ministry and service provider if the performance indicators are to be properly used as measurements of internal performance. Finally, we need to focus on the relevance of the indicators, which means advancing further towards outcome indicators.

**Evaluations**

Our second instrument is the evaluations. Our evaluation program was initiated in 1997 and is called EPG (*Evaluación Programas Gubernamentales*), or *evaluation of governmental programs*. It is a relatively simple, desk evaluation. The work is carried out with whatever information is available, and takes approximately 4 months of work using as the central element the logical framework matrix used by international agencies. This matrix has been adapted to the requirements of rapid evaluation.

Subsequently, in 2001, we implemented impact evaluations, and later in 2002 we developed a third type: institutional or comprehensive evaluations that evaluate a public ministry as a whole, including the goods and services it provides. The impact evaluations were initiated because the desk evaluations, with more limited information available, often left some key questions unanswered. Therefore, it was necessary to go to a second level of evaluation where more in-depth data could be collected. The impact evaluations include identification of baselines, the use of control groups and the application of statistical tools — all involving more sophisticated analysis. Thus in our case we see this ‘set’ of instruments as complementary, each to be used depending on the information that we need.

With regard to our legal framework, both the performance indicators and the evaluations were initiated in Chile as part of the budget formulation process. There was no permanent law that mandated the Ministry of Finance or the institutions to perform this work. Development of the system was mainly done on a voluntary basis. The work was always carried out within the budget formulation process, which includes the annual law and an agreement protocol that is signed each year between the Ministry of Finance and Congress, which establishes the list of programs/institutions to be evaluated each year. Subsequently in 2003, a modification to the Financial Management Law of the State included a provision for the Ministry of Finance to carry out evaluations of programs and institutions every year.

The evaluations have to meet certain standards, and both their technical development and the work flow processes have been framed to meet these requirements. Firstly, the evaluations are independent and are not performed by the institution that is responsible for the execution of the program. The evaluations are performed by panels of experts or consulting firms from outside the public sector, hired and managed by the Ministry of Finance. The institutions participate throughout the evaluation process but the responsibility for the evaluation lies with the Ministry of Finance as the contractor of consultants and external consulting firms.

The second requirement is that the evaluation program should be transparent, with all the information from the evaluations being accessible to the public. Once the evaluations are completed, they are sent to Congress along with an attached response prepared by each of the evaluated institutions, as well as an executive summary. This information is put on the web page of the Budget Office of the Ministry of Finance, and is thus available to all the political and social actors in Chile (www.dipres.cl/fr_control.html).

The information also has to meet technical requirements and for these a set of guiding methodologies and procedu-
res has been developed. There is a public selection process to ensure that the most technically qualified consultants are used in each evaluation. Timeliness is a very important requirement in using the evaluations in the analysis and decision-making process; if the evaluations are delayed, their information may become obsolete and therefore lose its relevance. This requires us to have very rigorous procedures to ensure both quality and timeliness. Cost-effectiveness is also important; the three different types of evaluation take different amounts of time, and therefore have different costs, and naturally deliver different levels of information. Each year we have to select programs and institutions to be evaluated, and we have to optimize the use of available evaluation resources and to allocate resources to each of these three types based on information needs.

The evaluation system also has a workflow process that begins with the selection of the programs to be evaluated. This selection is based on the anticipated information needs, over the coming year, of the Ministry of Finance, and those that have been identified through consultations with the Secretary of the Presidency and the Ministry of Planning. A proposal is then prepared and agreed upon with Congress. In tandem, the competitive process for public selection of consultants is initiated, as is the work with each of the institutions being evaluated, to collect available information and to train and provide technical assistance to all those program managers who will collaborate with us during the evaluation process.

The completed evaluation reports include conclusions and recommendations which are reviewed jointly with the institutions and with the relevant unit of the Budget Office of the Ministry of Finance. Decisions are made concerning follow-up for the next year, that is, possible program changes and their budgetary effects. The evaluations are sent to Congress and at the same time detailed work is begun with the relevant service provider or ministry to analyze the results of the evaluations. Commitments with the evaluated institution are entered into; these are the detailed actions to be carried out for each evaluated program and the timetable for implementation.

Let us look at decisions made using these evaluations; see also Table 3. Of the evaluations we carried out between 2000 and 2004, 25 per cent of the programs have been determined as in need of minor adjustments, for example needing changes in internal management, monitoring, or other things. On the other hand, 10 per cent of the evaluated programs have been eliminated, while 21 per cent have been identified as needing substantial redesign of components and internal management processes. The balance comprises modifications to program design or management processes (39 per cent), and institutional relocation of the program (5 per cent).

We carefully monitor the extent of compliance by the institutions with the commitments they have entered into following the evaluations. Statistics on the compliance of the institutions are used in MOF’s internal work processes,

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories of Budget Effect on Programs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minor Adjustments, Modifications, Relocation</td>
</tr>
<tr>
<td></td>
<td>Base Budget</td>
</tr>
<tr>
<td>2000</td>
<td>148</td>
</tr>
<tr>
<td>2001</td>
<td>130</td>
</tr>
<tr>
<td>2002</td>
<td>174</td>
</tr>
<tr>
<td>2003</td>
<td>141</td>
</tr>
<tr>
<td>2004</td>
<td>94</td>
</tr>
</tbody>
</table>
and are also reported to the Congress and to each ministry. This compliance monitoring information is provided by the ministries and is verified by the Ministry of Finance; in some cases more rigorous information has to be provided.

Up to 2004, we had evaluated 158 public programs and 14 institutions, representing the evaluation of 64 per cent of public expenditure potentially subject to evaluation through these methodologies. This figure takes into account public spending that cannot be evaluated by these methodologies — for example, defense expenditure, expenditures on the public pension system and education grants (the latter involve a voucher system to which this methodology cannot be applied).

Our management of the M&E system includes the ongoing review of the work processes used by the Ministry of Finance for these evaluations. We conduct surveys both on the processes and the effects that the evaluations are having on the institutions. The World Bank has recently carried out an evaluation of our program, and we are already seeing some results from this in our evaluation program.

As to ongoing issues regarding evaluation, we need to improve the use of the evaluation results within public institutions, and also by Congress. The data on the use of the evaluation results within institutions show progress, but more remains to be done to expand their use. This also applies to utilization of the performance information which the M&E system generates. Work also remains to be done to improve the information available within the institutions, to better apply the evaluation methodologies, and to improve evaluation processes.

**Conclusions**

We have learned various lessons from our system to date. In the first place, this is a cultural issue and thus it requires time.

We have to be persistent in the application of these instruments — if instruments are applied and then stopped, it has a very high cost for the institutions. Conversely, the continuous operation of the system, with ongoing review of its processes and technical elements, has allowed us to systematically enhance and continuously improve the system.

Finally, we need to identify and keep in mind the limitations of the instruments to monitor performance and to evaluate, so that we can make proper use of the information these systems provide in improving resource allocation and institutional management.

**Discussant: Verónica Silva**

*Deputy Director of the Social Investment and Solidarity Fund (FOSIS), Chilean Ministry of Planning (MIDEPLAN)*

I would like to focus my observations on two lines of thought, with the understanding that my perspective is institutional: I work in an institution that has been both a victim and a beneficiary of the national M&E system. I am going to explain what I mean by the term ‘victim’ in terms of costs and I will explain the benefits to my institution. The institution where I’ve been working for the past 5 years, MIDEPLAN, has undergone all three types of evaluations under the Chilean evaluation system: desk evaluations, impact evaluations and comprehensive expenditure evaluations.

From the standpoint of institutional learning the positive results that we have experienced with this government management and M&E system in Chile are due to its gradual implementation over time, so that institutions can gradually adjust to these new processes. I believe that it would not have been possible if the whole system had been installed at once.

Second, in practical terms it implied the creation of a specific work area for my institution — a new work unit that is called management control. It is not strictly correct to view this function as a cross-cutting activity within the institution, although it is true that all staff participate and are part of the managerial process which responds to the formal requirements of the system. This work unit was created within the institution to coordinate the process and to safeguard two factors that are crucial in the process.

First, it ensures that all the required information is available. All of us working in public institutions know that we have a large volume of information, but that it is not necessarily well organized, and that while it can be interesting for other things, it is not necessarily useful for management control purposes. Second, it is necessary to create an internal coordination mechanism within the institution to service this M&E system, and for programming control processes. Indeed, one of the advantages of the Ministry of Finance’s M&E system is precisely that it forces the institutions to pay attention to the need for better internal coordination; this is an interesting ‘bonus’ that should be acknowledged.

This national M&E system should not replace the internal follow-up, monitoring and evaluation mechanisms of the institutional programs. It would be a serious mistake for
each program or activity not to maintain its own monitoring, recording, follow-up and evaluation system.

I believe that this national M&E system has had a very positive impact in the internal management of my institution. There was initial resistance to the change process. Such resistance is totally natural, and is to be expected. Those managing the change have to take this resistance into account in their work program. As a result, the positive impacts of the system are something that only comes later in the process, and not at the beginning.

Another point, which Marcela Guzman covered very rapidly but on which I want to put a little more emphasis, has to do with the recommendations that result from the evaluations. Each evaluation, and the related processes that are part of the management control system, includes the formulation by the external evaluators of specific recommendations for each of the areas evaluated. These recommendations are discussed by the evaluated institutions, which accept, reject, or redirect them. In general, these imply some formal commitments which are agreed with the Ministry of Finance. This process seems to me to be very important from the standpoint of institutional learning, rather than if the evaluation report were just to be produced with nothing happening as a result. I believe that would be an incomplete process.

With respect to the limitations and the risks of the Chilean system, the first is that there is an institutional temptation to formulate indicators and basic goals that have already almost been reached. This makes it easy to achieve ‘success’. But this would absolve the institution from making greater efforts to improve its performance. The risk from setting such modest goals has to be handled by the manager responsible for this process, and hence it is very important that there be an institutional counterpart of this national system, located within the institution. This is particularly important when the system requires compliance by the institution, including perhaps the reorganization of the institution itself. Marcela Guzman explained briefly the improvement and management programs, which provide a special monetary bonus, but the temptation in all the institutions is to set modest goals for themselves, which can easily be achieved.

A second limitation is that there is always the risk that institutions will view and implement the evaluation recommendations as a set of independent tasks to be undertaken by different people within the institution in an uncoordinated manner. I have observed significant differences among institutions — between those which have a systemic vision of what the system is intended to achieve, vis-à-vis those where there is simply an institutional mindset to respond to a list of individual requirements.

A third limitation is that the system requires collection of performance information and the execution of the agreed commitments using the institution’s existing resources. But these activities cost money. For example, the surveys of user satisfaction which are conducted to collect information on the quality of service delivery impose a financial cost on the institutions that is not budgeted; no fresh resources are provided to the institutions to collect this information. This also applies to activities to register program beneficiaries. We need to be clear that this type of data collection implies reducing the resources available for other areas.

Finally, I believe that there is a limitation in the evaluations and in the management control framework, which currently focuses on narrowly-defined programs of each institution. From an expenditure stance, this makes sense. However, in the social sector in particular, there is a whole new generation of social policies that are comprehensive and cross-cutting in nature. These involve inter-sectoral interventions, where the policy is not confined to a single institution. As a result, this scatters the execution of the policies across different agencies and institutions. This shared responsibility is desirable, but it makes it necessary to seek other methodologies — other evaluations and management control systems, that are pertinent to the new generation of policies.

**Discussant: Ignacio Irarrázaval**

*Director of the Public Policies Program at the Catholic University of Chile*

In Chile we have now done a lot to evaluate the impact of programs. But the coverage is very low as a proportion of general public spending, and this is a challenge for us. How can politicians, decision-makers, and the public in general understand the costs and benefits of alternatives in the allocation of public resources?

We apply a logical framework matrix and that is very good, but as an evaluator, I am evaluating direct impact. For example, I am currently evaluating innovation in agriculture. I can do that and I can go to the agricultural sector and to some related ministry, but I don't have time to look at public works and technology in general. So I think we
need simpler, more generic indicators that would allow me to compare the alternative use of resources.

A second goal for an M&E system would be the improvement of public policy and the institutional setting. The culture of evaluation has already been created in the public sector in Chile, and this has been a positive thing. However, the procedure has been top-down. For those of you who do not know Chile or Latin America, ministers of finance are usually very powerful, and in Chile they are even more powerful than in the average Latin American country — powerful not only because they are in charge of resources, but also because in professional skills, the ministry of finance is above other ministries.

The establishment of the M&E system in Chile has been a top-down approach and for this reason it is very difficult for institutional learning and ownership at the institutional level to take place. Many times we have seen the Ministry of Finance build some programs which other institutions do not understand. So we try to gather some resources, transfer functions from other institutions, and artificially create a program. It will add up to a program, but it will not be owned by the institution, as such. The Ministry of Finance also has a more budget-oriented vision, and is not so much results-oriented.

We also need an M&E system to support accountability in the public sector. If we look at the public sector or public institution websites, we can see all those reports which are available, and if you take the time, you can go into them in detail. This has been done with a highly institutional vision and a very parliamentarian type of vision. But here we have a great challenge. Usually, the non-expert, the layperson or the average citizen has a hard time trying to visualize this. It is very hard for them to come in and understand exactly how this optimum resource allocation is taking place. So in this sense, I would prefer that indicators would simplify matters and make it possible not only to make information more accessible to the citizens, but make it easier for the public to understand it in general.

**Discussant: Keith Mackay**

*Senior Evaluation Officer, The World Bank*

It is clear both from Ms. Guzman’s presentation and from the recent World Bank evaluation of Chile’s M&E system that the system is highly efficient and effective. The Government makes intensive use of evaluation findings, particularly in assisting resource-allocation decisions in the annual budget process, and in helping the finance ministry (the Hacienda) to decide changes to the management and structure of sector ministries’ programs.

Chile’s system is one of the strongest government M&E systems in the world. In achieving this success, the single most important factor is probably the location of the M&E system — where the system was designed, progressively developed, and is still managed — in the finance ministry, specifically its Budget Directorate (DIPRES).

The finance ministry has succeeded in creating an ‘evaluation factory’, which includes a well-developed process for planning, commissioning, managing, reporting and using a range of types of evaluation. What is not clear, however, is the replicability of Chile’s M&E system in other countries in the region. Chile’s finance ministry is very powerful, and very capable; it is respected, even feared by other ministries and agencies. Compared to Chile’s finance ministry, those in other LAC countries are typically relatively weak.

The Chilean example is significant: in demonstrating that a whole-of-government M&E system can be built and operated at relatively low cost. The finance ministry limits the amount spent on each evaluation; it ensures they are not conducted in a wasteful or inefficient manner. Evaluations are conducted in a disciplined process, and to tight deadlines which are dictated by the finance ministry’s decision that each evaluation must be planned, commissioned, conducted and completed quickly, often within one annual budget cycle. Thus its *evaluations of government programs* (these are, essentially, program reviews) cost on average the equivalent of US$11,000, and take four to six months to complete; 158 of these program reviews were conducted between 1997 and 2004. The more in-depth and rigorous *impact evaluations* cost on average US$88,000, and take up to eighteen months to complete; 14 of these have been conducted so far, i.e., about 4 per annum.

The World Bank’s recent review of Chile’s evaluation system found some unevenness in the quality of both types of evaluation; and it suggested that this might have resulted from the tight budgets and tight deadlines for these evaluations. Thus one option for the Government to consider is to increase the amount of resources available to be spent on each evaluation, and to allow a longer time frame for evaluations to be completed.
Put another way, it is clear that Chile’s evaluations are highly cost-effective. It can be inferred that additional government spending on individual evaluations would also be cost-effective. In Mexico, for example, a series of impact evaluations have been conducted of the Progresa program, (which has been renamed Oportunidades) which have collectively cost several millions of dollars — no precise figure is available — and are widely recognized to have been highly influential in persuading the Government to retain and, indeed, to scale up the program considerably. (The program now covers 21 million beneficiaries, or about one-fifth of the Mexican population). The impact evaluations, while expensive, cost only a fraction of the Government’s annual spending on the program (which was US$780m in 1999), which in turn accounts for a large share of total social sector spending. This suggests a question: would Chile’s finance ministry ever conduct such a detailed and expensive evaluation as the ones for Progresa? The danger for Chile is that its emphasis on minimizing the cost of evaluations might lead it to under-invest in individual evaluations, even when the programs being evaluated are expensive programs which have major policy significance to the Government.

Another notable feature of Chile’s M&E system is its ‘graduated approach’ to monitoring and evaluation. It regularly collects performance information for all programs and activities. These are also used in its evaluations. The program reviews are conducted selectively, particularly for newer programs. The more in-depth impact evaluations are conducted later in the life of a program, especially where its impact is unknown, or where the program is a government priority.

Another commendable feature of Chile’s M&E system is its ‘graduated approach’ to monitoring and evaluation. It regularly collects performance information for all programs and activities. These are also used in its evaluations. The program reviews are conducted selectively, particularly for newer programs. The more in-depth impact evaluations are conducted later in the life of a program, especially where its impact is unknown, or where the program is a government priority.

The World Bank review also found that the finance ministry uses the evaluation findings intensively; indeed, it collects regular information on the extent and manner in which the findings are used. However the review also found weak ownership of these evaluation findings by sector ministries and their agencies, which typically regarded the evaluations as having been imposed by the finance ministry. Thus there exists an opportunity for ministries and agencies to use the evaluation findings much more intensively — to assist their own strategic planning, policy development, internal resource allocation, and ongoing management and control.

One option for the finance ministry to consider is a more pragmatic approach, such as the one followed in Australia under its M&E system in the 1990s. This often involved the creation of an inter-departmental steering committee for major evaluations. The committee would typically comprise senior representatives of the finance ministry, other central ministries, the key sector ministry whose program was being evaluated, as well as other sector ministries with responsibilities for related programs. This partnership approach usually achieved high levels of ownership of the evaluation findings. The steering committees generally exercised reasonable oversight of the quality of evaluations, which as a result were usually of good quality.

The case study of Chile also leads to another question: ‘How much evaluation is enough?’ The finance ministry conducts about 13 or 14 EPGs (i.e., program reviews) each year, and about 4 impact evaluations; in other words, a total of about 18 evaluations each year. The total cost of these evaluations, including of the evaluation managers and advisers in the finance ministry, is about US$0.75 million per annum. This represents a very small fraction of Chile’s annual budget of some US$20 billion. A comparison country is Australia — which has a slightly larger population than Chile — and which is considered by many to have had in the mid-1990s one of the most effective whole-of-government evaluation systems among OECD countries. During that period in Australia, about 160 evaluations of major programs were underway at any point in time; most of these evaluations were in the nature of program reviews, but a number of rigorous impact evaluations were also conducted.

Which of the two countries has the better approach? I simply do not know. Clearly, there is no point in doing more evaluations unless their findings will be used intensively; it would not be reasonable to argue that evaluation
should be done for its own sake. Yet the high value-added of Chile’s evaluation system does indicate that there would be very positive returns from additional spending on the M&E system.

Another strong feature of Chile’s system is that the finance ministry itself commissioned an external evaluation of the system, conducted by the World Bank. The ministry also regularly monitors the system’s performance, by tracking the implementation of evaluation recommendations. This monitoring and evaluation of the system constitutes good practice. We know from experience in all countries that building an M&E system is neither easy nor quick. Some efforts encounter significant obstacles, or simply do not work well in terms of quality of information produced by the system or the extent of its utilization. Such systems typically require major modification. And none of the more successful systems has developed in a linear or predictable manner. Unforeseen obstacles and opportunities can and do emerge, so there is a need for the M&E system managers to act flexibly and opportunistically.

The message here is that it is important to monitor and evaluate an M&E system, in the same manner as any other government activity—to determine which parts of the system are working well, which are not, and the reasons why. In undertaking such a review, there are two key issues to consider. The first involves an assessment of the quality of monitoring data and evaluation reports. The second entails a review of the extent and nature of utilization of the monitoring data and evaluation findings. Answers to these questions go to the heart of the value of an M&E system. Answers to these two questions permit informed decisions about the cost-effectiveness of the current system, and the extent to which there would be additional benefits from additional spending to expand the system’s scope or from other efforts to improve the system.
The *Plano Plurianual* and M&E in Brazil

**Country Presenter: Ariel Pares**  
*Secretariat of Planning and Strategic Investments, Ministry of Planning*

**Linking M&E with Planning and Budgeting**

In Brazil there have been enormous changes in monitoring and evaluation beginning in 1999. There was a change in the strategy for formulating the national plan, and planning and budgeting have become integrated.

We have a one-year budget; this is a legislative requirement. Then there is a four-year plan, the PPA (*Plano Plurianual de Ação* — the Multi-year Plan); this means there is the budget plus three years. And then there are longer-term projects with multi-year commitments, which are always based on territorial planning, including private sector activities. Preparing a four-year plan, integrated with the budget is enormously complex because it covers four budgets, and the fiscal base changes every year. It is very difficult to prepare a political plan, and we must recall that Brazil is a presidential country with a very strong executive. And then we translate this plan into an operational budget so that it can be monitored and evaluated. So when we draft the plan we need to ensure that it can be executed, that it is feasible, that it has been funded and that it can be evaluated and monitored.

But the plan is wider than the budget — there are budgetary and non-budgetary resources, and the non-budgetary side of the four-year plan is extremely important.

**Key Principles**

In the 1999 changes we had two fundamental principles. Firstly to ensure that government activity was linked with activities and the problems they were designed to address, so that the approach was results-oriented. The second principle was that management should be based on known aspects of good management systems and good business practices. So this means a logical framework for each program, with indicators, goals and appropriate measures, together with more effective management.

In 1999 we were undergoing a major fiscal crisis and without this plan I don’t think we would have come out of it. Let me turn to the fiscal crisis and the question of legitimacy. As you know, democratization in Brazil in 1985 led to the entire reorganization of several sectors of government and the pressure on the government was enormous. This meant that we had to implement an evaluation and accountability system.

The basic principle of our M&E system is annual, results-based, self-evaluation. Whoever implements must evaluate, and the three different levels of evaluation are by program management, by the sectoral ministry and by the Ministry of Planning.

This self-evaluation system is overseen by the Ministry of Planning. Its objectives are to provide greater transparency to government actions, to support decision-making, to provide learning and to disseminate information within the respective organizations. We see it as important that line ministries acquire capacity for self-evaluation and that this improves their management systems. This will improve the
management of the plan, as this is also coordinated with the annual review of the plan, as well as the draft annual budget. Here we have similarities with the Chilean approach.

Results-based evaluation is very difficult and sometimes time is our enemy. We need evaluations to begin with the desired objectives or results. We need to think of results in terms of impact, but because that is conceptually very rigorous we need intermediate indicators. But essentially we first seek to know what is happening in society and then to see what is happening inside the program in response to society’s needs. Any organization that implements a program needs to listen to the users as well as opinion makers, and that applies to implementation as well as design. Thus we hope that the people who manage the program and do the evaluation are trying to manage their programs better, although this may not always happen.

There are three institutional levels in our evaluation system as mentioned above. There is always self-evaluation of the program by the institution and the program manager. Then the sectoral ministry evaluates their entire program, and finally the planning ministry undertakes an evaluation of the plan as a whole. All this generates a voluminous evaluation document which goes to Congress. Generally Congress is little involved in our work and this is a serious problem. We would much rather that they were more involved. As Brazil has a presidential system of administration, the executive is very strong vis-à-vis Congress and it is not always the best of relationships.

In 2003, 353 programs, out of a total of 389, were evaluated. Some are really not programs but internal services which arguably do not require evaluation, but we would prefer that they be evaluated. Their managers do not undertake evaluations and we don’t force the issue.

For the government’s 2000-2003 Plan, nearly 90 percent of programs had indicators, but only between 40 and 45 percent of these have been validated. Important issues are what has not been evaluated, and what has not improved, and both these proportions have increased in recent years. The percentage of programs which failed to meet their physical targets has risen from 33 percent in 2000 to 64 percent in 2003.

**Decentralization Challenges**

It is useful to distinguish between programs, of which we now have 374, and the 5,000 to 6,000 activities of which they are comprised. There is a challenge with decentralized actions — how to evaluate and monitor activities in 27 states and over 1,500 municipalities. Nevertheless between 2002 and 2003 we increased our knowledge about what happens with the resources that go to the states and the municipalities, with the percentage of decentralized organizations having monitoring mechanisms increasing from 49 to 62 percent. While it is important to know about the performance of programs using decentralized resources, our knowledge of the reality of these actions is very low. This is a cross-cutting problem of the federal government, and the decentralization system is not yet entirely clear — it is a decentralized system with delegation of authority, but without clearly assigned responsibilities.

**Strong Points**

The introduction of a systematic evaluation process to provide a focus on results has had some strong features (Box 2). It has changed the culture of the line ministries; there has been a change in the behavior and the attitude of public managers in that they see the need to better understand and think through their actions based on desired results. However, this is still far from being a full results-based management system.

**BOX 2:**

**Brazil-M&E During the 2000-2003 Multi-Year Plan: Critical Retrospective**

<table>
<thead>
<tr>
<th>Strong Features</th>
<th>Weak Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Introduction of a systematic evaluation process focused on results.</td>
<td>• Lack of in-depth evaluations, linking expenditure management through the budget cycle.</td>
</tr>
<tr>
<td>• Identify systemic and implementation problems.</td>
<td>• Partial incorporation of evaluation results into the revision of the Multi-year Plan and the budget.</td>
</tr>
<tr>
<td>• Promoting learning in organizations.</td>
<td>• Low quality standards for results based information.</td>
</tr>
<tr>
<td></td>
<td>• Lack of technical capacity in line ministries.</td>
</tr>
<tr>
<td></td>
<td>• Bad timing for information management.</td>
</tr>
</tbody>
</table>
Our statistical information allows us to more readily understand program implementation and, through having objective questions and answers, to identify systemic problems within ministries, and then program by program, systemic implementation problems.

Promoting learning in organizations has also been important. Even though our experience is limited, moving to self-evaluation is resulting in two things. Firstly, it means ongoing and increasing institutional learning. Secondly, with organized citizen participation there is greater transparency because those ministries that evaluate with citizen participation do so by negotiating any changes to the program. Transparency thus goes hand in hand with evaluation — or rather participation goes hand in hand with self-evaluation. All of this allows for participation of citizens, as well as from within the organizations being evaluated. We are progressively moving in this direction.

**Weaker Points**

The weak points are more numerous. There is a lack of in-depth evaluations, linking expenditure management through the budget cycle (Box 2). We are beginning to incorporate evaluation results into the revision of the Multi-year Plan and the budget. But we want to ensure that the results-based information provided by the different ministries meets minimum quality standards. In designating the programs to be evaluated we want to take advantage of all of the work done by the different ministries and use this information in our evaluation.

We have a problem with the quality of information and indicators, and related to this is a lack of technical capacity in the line ministries. Chile has talked about their difficulties here. So how can we make line ministries capable of understanding this monitoring and evaluation language? This is a core problem, and is made more important by our strong decentralization reforms, which we started to implement during 2005.

**The New Multi-Year Plan, 2004-2007**

So this is the monitoring and evaluation system that we have created and it is part of the new 2004-2007 Plan (Box 3). As we started only in 2000-2003 we have limited experience but we have evaluated this four-year experience and we have changed a few important aspects to improve the evaluation system.

We need to monitor and assess the plan from a strategic viewpoint. What do we mean by strategic? Brazil’s Plan contains several over-arching goals and more than 30 specific ‘challenges’ or objectives. We separated quantifiable from non-quantifiable challenges. We don’t know how the non-quantifiable challenges are going to develop, but we can look at these from a qualitative viewpoint also. Human rights is an example of an issue that is harder to quantify, especially with the statistics that we have in Brazil.

So the selective evaluation of programs is important. And we need a participatory approach to evaluation. So we are linking together two systems, one that is broad and comprehensive, combined with a deeper system.

As mentioned earlier we have three types of evaluations in our system. First, a comprehensive self-evaluation, secondly a decentralized but networked evaluation system under which the ministries come together with decentralized units; and, third we have selective, in-depth evaluations that are done separately through an independent commission for monitoring and evaluation, involving a participatory approach — citizens, academia, and staff of the sectoral ministry.

---

**BOX 3: Brazil-System of M&E for the 2004-2007 Multi-Year Plan**

<table>
<thead>
<tr>
<th>What Changes on the New Plan?</th>
<th>Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Strategic viewpoint approach (30 objectives).</td>
<td>• We need to ensure that the evaluations influence the annual plan and the budget.</td>
</tr>
<tr>
<td>• Creation of an M&amp;E system.</td>
<td>• We need to ensure that ministries have the internal capacity to support program managers in M&amp;E processes.</td>
</tr>
<tr>
<td>• Evaluation of selected programs.</td>
<td>• We need to spread a culture of evaluation and its use in the formulation and implementation of government policies and programs.</td>
</tr>
<tr>
<td>• Shared approach on evaluation.</td>
<td></td>
</tr>
</tbody>
</table>

---

23
We need to ensure that the evaluations influence the annual plan and the budget. However we have expenditure rigidity and inertia and sometimes we cannot make changes because legal provisions state that certain budget items cannot be amended or changed.

We need to ensure that the ministries have the internal capacity to support program managers in the M&E processes. Our experience thus differs from that of Chile. We need to spread a culture of evaluation and its use in the formulation and implementation of government policies and programs. I think this is a challenge for all our governments.

**Institutional Arrangements for Evaluation**

Firstly, we have the line ministries. Within each ministry there is a committee for the coordination of programs, and of course the program managers. Then there is a unit for monitoring and evaluation. For the 34 ministries there are some 26 such M&E units (UMAs) which are important interlocutors in the evaluation system. They provide advice to the committee for the coordination of programs within their ministry and they consolidate M&E data and information, especially evaluation data, and ensure it is incorporated in the SIGPLAN database. They also propose methods for monitoring and evaluating the programs and provide support for the preparation of the management plans for these programs. We are now asking every line ministry to provide us with a program management plan, something we did not do when we began work in 1999, when we failed to realize its importance from the operating viewpoint. Finally, the M&E unit promotes internal training in methodologies of monitoring and evaluation.

The unit is not only focused on the ministry or the sector but also has to look at related multi-sectoral issues. Units are part of a broader network which must help us think about problems and issues from a broader government perspective, not just from a sectoral perspective.

Then we have the Ministry of Planning. Within it we have the M&E commission (CMA) which is responsible for establishing a general framework and guidelines for the M&E methodology and for recommending methodologies for the annual evaluation of the plan. The CMA will select, in consultation with the ministries, those programs that are going to be evaluated. The Commission will also evaluate large projects —$4 million and over. We have learned a lot from the Chilean experience and I think we are going to be able to establish a system for ex ante (rather than ex post) evaluation of large projects, because this is better for assessing performance and is more cost-effective. We are going to start this year. However, we don’t have great expectations and we will focus only on new projects.

A small technical committee provides technical support to the CMA and this technical committee is also going to be involved with other units in the executive branch in its work.

Within the Ministry of Planning we have my secretariat —Planning, Investment and Coordination. The Federal Budget Secretariat deals with budget development and execution. And we have the Institute for Applied Economic Surveys and the Secretariat of Management.

Then we have the Ministry of Finance which includes the National Treasury which manages cash flows, etc. There is some tension between the budget roles of the Ministry of Planning and the Ministry of Finance, but this tension works well and there is good collaboration. The Ministry of Finance also includes the Economic Policy Secretariat.

There are four other organizations which are involved in the work of the M&E Commission —including the Office of the President of the Republic, which has an obvious interest in evaluations, the Federal Secretariat of Control, and the Ministry of the Environment. It is important to have the latter involved as the Court of Accounts (the national audit institution) of the federal government has stated that we need to have environmental impact assessments for all our projects, and these need to be incorporated into any ex ante evaluation. Finally there is the Institute of Statistics.

**Evaluation of Selected Programs**

There are several general characteristics of the way in which the Ministry of Planning evaluates the programs selected. Firstly we need to evaluate the processes and the impact of the programs. These external evaluations have participation from academics, representatives of users and the program team, and external experts and institutions that we employ for this work. Program teams may resist evaluation. They think that individual knowledge overrides general knowledge and that empirical knowledge is better than scientific knowledge when it comes to improving results and management. Overcoming this resistance and changing this mindset is a big challenge.
We know that there are user expectations from programs and we are committed to a gradual implementation of participatory evaluation. The program manager needs to know what the user thinks in order to adjust the management of the program. We need to increase the level of social control to increase the effectiveness of our programs. But what is real and what is perceived? We need to know that things are done well and thus help citizens. Also, it is very important that we agree on results with society—not society in an abstract manner but rather specific groups in society. There are many techniques that can be used in this consultation process. We are not really concerned about how scientifically rigorous our communication is but we want to know how people perceive the results and the outcomes of the program. We have to understand that these agreements with society are also important in framing future changes to the program.

Lastly, we have the evaluation plan for each program. We also have, of course, management plans for programs. When we select the programs to be evaluated we look at the key aspects to be evaluated, the cost and timing, the information required and avenues for social participation. On costs, to evaluate a big program we are going to need about $400,000 and for a smaller program perhaps $150,000. Of course, we would be able to spend more money if we had a longer period of time in which to evaluate the program.

The evaluation plan also discusses the training and certification of technical staff in the M&E units (UMAs) in each ministry. We are trying to create a program for the certification of M&E specialists or a cadre of specialists who are going to be invited through a public bidding process.

The SIGPLAN information system has to be consistent with sectoral ministry systems. It is one thing to look at a process, but a different thing to monitor and evaluate physical aspects.

Lastly, one of our challenges is that senior management become interested in monitoring and evaluation systems. When senior management understands that these are important, then things are going to be better.

**Discussant: Romulo Paes de Sousa**  
*Secretariat of Evaluation and Information Management, Ministry of Social Development*

I would like to look at Brazil's M&E system from the social development viewpoint. Social development is a concept that has to do with building development from such a social viewpoint. The idea is to have the citizens understand that things are going to become better and, political leadership is also going to be involved.

We would like to know who our clients are—who the users of these programs are. There are almost 50 million users of our programs. These persons have different needs when it comes to social assistance policies. They are a highly vulnerable group. Some need everything and all programs and policies have to be designed to be delivered to this population. Another segment of the population has different needs: for example, they need incentives to become members of the labor market again. They don't need everything, just a portion of our programs.

So we have different populations with potentially different needs. This is important because in the last group we have a population that is close to the poverty threshold and may fall into it and we need to do preventive work to avoid them needing these programs.

We have talked about program design and we saw many programs for the generation of work and income. But these programs don't really take account of the fact that the community doesn't have the ability to absorb these programs and that we have to look at different populations. For example, indigenous peoples, the poorest people, have no experience with money and these programs are not going to be able to solve this quickly.

So we have two choices. We have a moral choice, because we have to pay our social debt and our population needs to be included. We also have a pragmatic choice, because we are dealing with groups that may be included rapidly in the market, for example, people who are currently unemployed, but who can be rapidly reinstated in the labor market.

We are talking about two different vectors here, one that has to do with social protection and the other one that goes outside the model, which has to do with social development. In Brazil we have 21 programs in the Ministry of Social Development for families, young people, elderly persons, etc, covering nearly 35 million persons and involving about $4.6 billion annually. So we are talking about many people, many programs and a lot of money. We have to decide what we are going to evaluate. Roughly eighty percent of our resources are allocated to 20 percent of the programs, and that is a general pattern. We cannot evaluate everything. So we are building a database to support the decisions on what to evaluate.
Let us look at the example of activities related to work and income generation. Whereas other programs are sometimes very isolated, here we can see a more systemic linkage. The activities on income and job generation need a more systematic and integrated approach to evaluation because there are both government and public actions, there are semi-public actions and then there are non-governmental actions. Access to information on these actions is important so people know what we in the Ministry of Social Development are doing. It is also important for us to know what others are doing, to avoid overlapping and waste of resources.

There are also many follow-up processes within the government and between governments, and studies of the different strategies and mechanisms for social control. There are studies of good practices. There is impact evaluation. And there are opinion surveys on the different actions that are taken and qualitative studies of the results.

To conclude, the most important step that we are taking right now is to conduct more in-depth or larger evaluations that would allow us to understand the quality of those programs and to link them to planning and budgeting. This will assist in building a more integrated system so that these studies can shed light not only on the policies and programs of the social planning ministry, but also enhance the policies of other ministries, especially in the social sectors.

**Discussant: Marcos Holanda**
*Director of the Economic Research Institute in the Planning Department, State Government of Ceara, Brazil*

In Ceara, we are new to monitoring and evaluation, but we have taken some initiatives based on the experience of our federal government and of Chile and Mexico.

First of all it is important to understand that monitoring and evaluation is a tool within something much wider —results-based management. M&E is not an end in itself; it is basically an instrument to support a management process based on results. We have indicators, objectives of the government, and from that we use monitoring and evaluation as a tool.

Secondly, in the Government of Ceara we have adopted an institutional structure to support monitoring and evaluation that is similar to the one at the federal level and similar to the Chilean experience. But in Ceara this process is managed not by a committee, but by a commission of management made up of four secretariats dealing with planning, finance and administration. This integrates the main cross-cutting secretariats and it strengthens the implementation of the results-based process and of monitoring and evaluation.

It is a results-based system, not an M&E system. In this way it is different from the federal government. In addition to the four ministries there is a technical group which supports this work, which is based on the technical committees, made up of specialists within those four ministries. The Ministry of Planning has technical groups from its own different secretariats. The greatest challenge here is integration within the government in order to reach those results.

In Ceara, when starting systems of monitoring and evaluation, we have to answer five basic questions when trying to sell this approach to the other secretariats and other public entities. We have to be very clear about why we are doing results-based management, how we are doing our evaluation, why we are doing it, and for whom, and where we are going to implement it. If we cannot clearly present the answers to the secretariats we cannot expect a very great interest on their part.

Another issue that has been mentioned here is the perception that monitoring and evaluation are not the same thing. They are integrated processes, but they are different.

From the operational viewpoint, in Ceara, we divide the work. There is no single ministry or secretariat that is responsible for operationalizing the tasks and the work. Results-based management is more delegated to the individual entities, and the monitoring and evaluation to the Ministry of Finance.

My comments reflect the way we are working in Ceara compared with the federal government, and the fruits of experiences and the successes that we already have in trying to adapt it to our realities.

**Discussant: Yasuhiko Matsuda**
*Senior Public Sector Specialist, The World Bank*

As a discussant of the Brazilian case, I want to present a few ideas and a few reactions to the presentations and the comments that I have heard so far, plus what I have observed over the past few years of working in Brazil.

My first observation is that in comparison to some of the other countries, federalism in Brazil provides us with a very interesting situation. The Brazilian Federation is a great laboratory for institutional experimentation, and today we also have participation from the state level in addition to
sectoral ministries. Observing this monitoring and evaluation agenda in Brazil, we can see different approaches at different levels of government and among different sub-national governments.

One institutional particularity, at least as it is impacting the way in which the monitoring and evaluation agenda is evolving in Brazil at the federal level, is the multiplicity of central agencies that are involved in setting the directions for the entire government. This is different from Chile, where a strong central budget office within the powerful Ministry of Finance, has single-handedly led the development of the monitoring and evaluation system as part of a broader results-based management system.

In Brazil, we have at least three central agencies at the federal level, each with a stake and slightly different perspectives on this and related topics. This was presented as a strength of Brazil’s institutional arrangement, and I agree that it can be. But it can also increase the transaction costs involved in developing government-wide policy. My impression is that the transaction costs have been rather high, and developing the M&E system for the entire government has not been easy. So that is one reflection of the particular way in which the federal government and the executive branch is organized in Brazil.

It is interesting that in the case of Ceara, there is a similar type of multi-sectoral commission leading the development of the system. One can speculate whether that is something merely inherent in the Brazilian institutional design or whether it is just coincidence that the federal government and the state government of Ceara happen to rely on the multi-sectoral approach, and that other states and municipalities are different.

I say these things because this is not a technocratic exercise. There may be a tendency to over-engineer the design of these systems. We would like to see a good monitoring and evaluation system that includes different levels of sophistication including impact evaluation — something more like the Chilean approach — and all the gradations of evaluation sophistication in between. We would like to see a system that informs the budget and policy-making processes, and one that improves management within ministries. For each of these things, one could easily come up with a long list of institutional features which, if put together would create a monster, and I don’t think it is possible to develop such a monster overnight.

One of the most impressive features of the Chilean case, which I know a little bit better than the other countries represented at this conference, is the gradual development of the system over a little more than a decade. The Brazilian experience also has been evolutionary at the federal level. It started not with the attempt to develop an M&E system, but to strengthen the government-wide planning apparatus, and its linkage to the budget formulation and program management. Monitoring and evaluation is a component of this agenda. This has been evolving over almost the last decade, but the evolutionary path seems to have been rather different. To me the Chilean evolution seems to have been much better controlled and designed. The Brazilian evolution has been more stop and go, depending on the emphasis by the different governments, and the political support that the technical team has received, notwithstanding the continuity of the technical team.

So it is important to understand these institutional dynamics and avoid over-engineering the design of these systems. I think this is quite difficult to avoid because we tend to rush into technical rationality and forget about the institutional dynamics that drive these change processes.

Finally, is there any uniformity among these ministries in terms of how they approach monitoring and evaluation? In fact, in some ministries there was quite a great diversity within the ministry in terms of M&E initiatives, and the central part of the ministries didn’t always know what the various operational parts of the ministries were doing in the name of monitoring and evaluation.

And some ministries, because of the nature of the goods and services that they provide, were subject to heavier political interference through the budget process and so on, in comparison to some others. We had the strong impression that the former were very unlikely to be able to develop robust monitoring and evaluation mechanisms. Their budget was messed around by politics, and they had programs on paper, which may be different in reality, although the technical people were trying their best to execute what they could. There was a great divergence between what was intended on paper, budget and plan, and what actually happened on the ground. I suspect that this sort of situation applies to a great number of sector ministries in most countries in our region. Yet I think we have a tendency to over-engineer these systems and also to ignore the political reality in coming up with the institutional design. This is one of the challenges, not only for Brazil, but also for most of the other countries in the region. This is integral to the discussion of monitoring and evaluation systems as a whole.
Colombia’s National System for Evaluation of Management and Results

Country Presenter: Manuel Fernando Castro
Director of Public Policy, Department of National Planning (DNP)

Introduction

I will first provide the context of the evaluation and results management system in Colombia, which has received very direct support from international financial institutions. I will stress the institutionalization of this system. The initial development of the system took place between 1991 and 2002, with a deeper reform process since 2002. I will discuss how the system has strengthened and then look at the existing elements, the institutional framework in Colombia, the progress to date, and the main challenges.

Context

Colombia’s model of evaluation or results management goes beyond a results-based budget. While this is an important element of the model, the system seeks to support the modernization of the Colombian state and to support institutional change, so that our government’s human and financial resources become performance oriented.

The system began as a part of a very strong modernization process in Colombia in 1991, covering a series of economic and political reforms. But there was also a very important constitutional change; the new Constitution mandates that the state should have proper M&E tools, as well as a system which brings together all the different institutional efforts desired by a reform-oriented government. Then there came a national development plan, which is the document that supports the four-year government plan.

Main Problems

But there were problems during this stage, including a lack of coordination among the separate processes of planning, budgeting and evaluation. The institutional arrangements for the public budget in Colombia involve two central offices. The Ministry of Finance programs the resources and the National Planning Department manages public investment —including transfer payments — and

In Colombia there was a great deal of awareness about how public funds were being used because of enormous corruption scandals before 1991. There was a general understanding that huge amounts were being spent on high-profile projects and infrastructure projects, which, at the very least, had transparency problems.

The system was conceived along three lines, with these three elements being integrated:

- monitoring or follow-up of results;
- strategic evaluations; and,
- accountability.

The system was set up by using results-based indicators in public institutions, which while seeming basic today, was a new approach then. The introduction of indicators was one of the great initial achievements, together with different instruments that were used to monitor results —for example contracting with the public sector at the national level, and a number of other instruments that were introduced into the culture of public management, including performance-based results. There was a great deal of focus on follow-up, and the issues of evaluation and accountability became well conceptualized.
is responsible for their evaluation. Establishing the link between all the different actors at various stages has not been an easy task in Colombia.

Another problem was that the system did not define the scope of government to be covered — whether it was going to monitor only the central level or the other levels of government. While there were efforts to monitor the other levels of government from the center, this created problems in a decentralized country such as Colombia.

Thus we had not clearly defined the different linkages, tasks and responsibilities of the system, which up until 2002 created complex institutional problems. We have learned from this and we have tried to adjust accordingly. Significant problems persist with the systems in Colombia in spite of the fact that we have moved forward with a very good M&E system.

If the system does not have good information it cannot be used as a tool for quality evaluation. So information and its related problems have been quite an issue in Colombia. In 2002, Colombia had 940 indicators, which made it really hard to monitor results. We needed to simplify the indicators and we also wanted to see what the impact of that simplification would be on government management.

We found that there was not a clear separation between the two concepts — monitoring and evaluation. Evaluation is more exhaustive and more complicated and it goes beyond monitoring. Monitoring does not establish any causal relationship between outcomes or results and the factors that brought about those results, whereas evaluation seeks to do that. So we had to redesign the system and clarify the concepts.

When we commenced this implementation, we took all these things into account. Additionally, there were no incentives to stimulate demand for evaluation from the different institutions, and there was no dissemination of results, which undermined accountability.

**Fiscal Issues**

The fiscal situation in Colombia also had to be taken into account. When the evaluation and results management system was initiated there were no major fiscal problems, although there were some management problems and a need to modernize the state. However, social expenditure grew by four percentage points of GDP between 1994 and 2004, and after 1998-99 the government’s deficit was a very persistent and delicate problem. Thus all M&E instruments had to be used to try to improve budget allocations; we were finding it very hard to efficiently allocate resources. So the fiscal situation was a very important issue, and the state has made efforts to bring about reforms in light of the fiscal problem. The quality of expenditure from intergovernmental fiscal transfers, which are significant in Colombia, is also an issue.

However, more expenditure had not brought about better results, specifically in social indicators. Between 1995 and 2000 education expenditure grew 30 percent but the official matriculation levels grew only 18 percent. Hospital expenditure grew 113 percent between 1996 and 2001 but the number of persons receiving health care remained constant. So how could we better spend money and what impact are we having? The government started to think about using M&E tools for better performance management. We looked at the original concepts of the system and the three main components.

In the monitoring and results-based budgeting component we simplified indicators, reducing the number from 900 to 300. Then we started monitoring the outcomes or the results, measured against presidential goals, for which the ministers and the technical teams in the line ministries were responsible.

We looked at political, financial and institutional restrictions. We wanted to tell each sector how much money it was going to have. When we developed goals only from the technical viewpoint, the politicians were not committed to them. Thus we had to meet both political needs and technical needs and there was a dialogue between the technical staff and the politicians. For education they agreed that in the next four years we were going to open up 1.5 million student positions in schools.

Once we strengthened the initial design of the system, we looked at what we call strategic evaluations (evaluaciones focalizadas). In 1998, we started a new process of conducting impact evaluations, driven by the government. We tried to work with the social sector whose programs for reduction and mitigation of poverty were the first to be evaluated. The results were good, so we started to analyze the policy implications, and we recognized the need to strengthen and extend impact evaluations to other sectors.

On accountability, an important issue is that the public should be aware of the results of government policies.
So both information and accountability were included in our program. We wanted the information to become public; monitoring information should be made available to the citizens in general, and the person reporting the information has to be accountable for it. We also wanted to establish quality controls for such information; while reporting to Congress is a very important accountability instrument, the reports were very voluminous—more than 1,000 pages. No Congressman would read them so we had to simplify and put results information in user-friendly language.

We attempted to establish a very clear relationship between the priorities proposed by the government during election campaigns, and the content of the government’s national plan and the goals that were set. We wanted to make a link between these promises and the resources allocated for these priorities. So we had to link the priorities of the government with the budget. We had made efforts to do this in Colombia in the past, but we had not been able to articulate this. Now we believe that even though we are still at an early stage, we will get there in the near future. The idea is to be able to look at the government budget and see how much money is given to hospitals, to education, to paving roads etc. This requires new ways of classifying government budget expenditures—to reclassify the investment budget and align it with our goals.

All this technical information—M&E reports, impact evaluation reports, management reports, etc.—has to flow in simple language using simple wording.

A system of this nature should work on the basis of very transparent assumptions and also incentives. These incentives can be budgetary, of course. Many countries have budgetary incentives; those in Chile are very interesting. Incentives can involve rewarding managers who do a good job, and sanctioning managers who don’t perform well.

**Institutional Structure for M&E**

In Colombia, the highest social and political decision-making body is the CONPES (Consejo Nacional de Política Económica y Social), the Council of Economic and Social Ministers. The CONPES decides what is going to be evaluated and what is not going to be evaluated. A four-year agenda is set for evaluations. Below the CONPES we have a group of officials with technical and political responsibilities. Then we have the President’s Office and the President’s Council. There is an Inter-ministerial Committee for evaluations and results-based management to make the link between results and decision-making. Ministers are invited to listen to the committee’s views. There is the National Department of Planning, and the Ministry of Finance, and we try to have a system in which evaluations influence planning and budgetary decisions. We have the line ministries and the different entities that are ranked as ministries. Also, we have what we call the decentralized sector, for example, ECOPETROL (Colombian Oil Company). This is the structure of the system at the national level. All these institutions come within the system of results indicators or results-oriented management.

In terms of accountability we have political control, which is the accountability of ministries to the president, and social control, which is the external accountability of government—to citizens, Congress and oversight organizations.

The progress made and remaining challenges that we face in each of the three components of the system are now examined briefly.

**Monitoring and Managing for Results**

Here we are not only talking about all the indicators, but also focusing on the most important results in terms of the government’s goals for major programs. Four-year goals are set at the beginning of each Administration, and we are able to monitor the achievement of these goals (Box 4).

An effort was made to develop a programmatic classification for the budgets for 2004 and 2005. Next year, for the first time, we are going to present to Congress a results-based annex to the budget—although this will cover only the investment budget, because in Colombia we have not yet been able to develop a consolidated budget, covering both the investment budget and the operating budget. Our objective is to encourage discussion about the budget, not only on global expenditure, but on results. We have a web-based tool (SIGOB) that I will explain later. It provides citizens with real-time access to government management information and they can see how well schools are doing, how much money is being allocated to road paving, etc. They can find out the person responsible for each activity and they can see how the government budget is being allocated.
There are also three elements that are important. We are monitoring not only goals and results from the four-year government plan but we are also trying to evaluate other government strategies — for example, poverty-related strategies, millennium development goals, and so forth. The idea is to look forward to year 2019, which will be the 200th anniversary of Colombia. This will enable us to see where we are going and what kind of impacts we are having on the competitiveness agenda of the country.

We need to have M&E sub-systems at both the sectoral and sub-national government levels, and to ensure that these sub-systems will be able to link with the central system. But not everyone in the line ministries has appropriate monitoring and results follow-up tools. We also need to replicate these national systems at the provincial level. The objective would be a system that has the same quality standards and criteria as the federal system, as well as all the information needed to manage the sector. Not all the information generated by the system is going to be of interest to the President or to ordinary citizens, so we need to identify the information that would be relevant to the ministries or the sector. These sectoral sub-systems have to fit together and be articulated with the federal system, so that we are able to measure sectoral results. The same thing applies at the municipal level. We are establishing measurement and monitoring systems and results-based budgeting, and we are replicating the system that we apply at the national level. The idea would be to consolidate the whole system to provide municipalities and provinces access to it. The federal government would give the technical support needed to consolidate all these sub-systems.

The final challenge here is to have results-oriented budgeting at the national level. At present we can’t give more money to organizations or individuals who do a good job because we have budgetary restrictions. However, we have to try to institute a fund such as the one which Chile’s finance ministry oversees, and we will raise this issue with Congress.

---

**BOX 4: Colombia-Achievements and Goals 2005 – 2010: Monitoring and Budgeting for Results**

<table>
<thead>
<tr>
<th>Achievements</th>
<th>Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Focus on results.</td>
<td>• Sub-systems for M&amp;E and results-based budgeting in strategic sectors.</td>
</tr>
<tr>
<td>• Goals for the government are set in National Development plan (PND).</td>
<td>• M&amp;E and results-based budgeting at municipal level.</td>
</tr>
<tr>
<td>• Clear distinction between M&amp;E.</td>
<td>• Results-based budgeting for investment and recurrent spending at the national level.</td>
</tr>
<tr>
<td>• Roles, procedures and responsibilities.</td>
<td>• Budgeting incentives.</td>
</tr>
<tr>
<td>• Results-based budgeting (2005).</td>
<td></td>
</tr>
<tr>
<td>• Results based discussions in the Congress.</td>
<td></td>
</tr>
<tr>
<td>• Real time access to M&amp;E information through internet (SIGOB).</td>
<td></td>
</tr>
<tr>
<td><strong>Monitoring Government Strategies:</strong></td>
<td></td>
</tr>
<tr>
<td>• Mission to Design a Poverty Reduction and Inequality Strategy</td>
<td></td>
</tr>
<tr>
<td>• Millennium Development Goals</td>
<td></td>
</tr>
<tr>
<td>• Colombia’s Century Vision 2019</td>
<td></td>
</tr>
<tr>
<td>• Competitiveness Agenda</td>
<td></td>
</tr>
</tbody>
</table>
**The Evaluation Agenda**

We are developing a four-year and an annual evaluation agenda. We need to ensure technical resources are available to support this agenda. In prioritizing these evaluations, criteria we will need to consider include, for example, the characteristics of the population concerned and the number of people affected by these programs, the relevance of these programs for a specific sector, how innovative or creative the programs are, and therefore the possibility of replicating them. This will help us decide what should be evaluated, because we cannot really evaluate everything (Box 5).

Impact evaluations are not the only form of evaluation; there are other, less costly, but equally important kinds of evaluations such as management and process evaluations. So our idea would be to have a flexible evaluation agenda, covering a diverse range of evaluations.

We have succeeded in increasing the amount of budget that is subject to evaluation — from 4.13 percent of the investment budget in 2002 to 13.4 percent in 2004; our idea is to go to 23.9 percent in 2006, and to maintain it at that level until 2019. To achieve this we need to develop our evaluation capabilities. For example, impact evaluations cannot be done only with the resources that we have in our country. We need to build partnerships with international companies and local firms and with academia. If we want to carry out more evaluations we need to reduce their cost. Evaluations which cost US$3 – 5 million are too expensive, even if they are in themselves cost-effective. These costs are too high and we have to extend the evaluation agenda.

We have other challenges in Colombia. We need an external evaluation of our M&E system, including an assessment of its impact on policy and decision-making. We are also evaluating the program of public sector reform and modernization.

We also need to incorporate all evaluations into the government's budget decisions. But the institution responsible for the budget in Colombia's Ministry of Finance has said nothing about linking evaluations with the budget. It needs to establish guidelines and quality standards for this.

The evaluation agenda for the 2000-2006 period includes the social support program and social assistance programs,

---

**Box 5:**

**Colombia-Achievements and Goals 2005 – 2010: Strategic Evaluations**

<table>
<thead>
<tr>
<th>Achievements</th>
<th>Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Four year and annual evaluation agenda.</td>
<td>• Extend the evaluation agenda.</td>
</tr>
<tr>
<td>• Inter-sectoral Committee on Evaluation and Results-based Management.</td>
<td>• Develop evaluation capacities.</td>
</tr>
<tr>
<td>• Evaluation Criteria:</td>
<td>• Reduce costs.</td>
</tr>
<tr>
<td>• Amount of resources</td>
<td>• Increase the annual investment budget evaluated to 23.9% in 2006.</td>
</tr>
<tr>
<td>• Characteristics of the population</td>
<td>• External evaluation of the Colombian M&amp;E system.</td>
</tr>
<tr>
<td>• Relevance for the sector</td>
<td>• EOP expenditure evaluation.</td>
</tr>
<tr>
<td>• Innovation</td>
<td>• Minimum quality standards.</td>
</tr>
<tr>
<td>• Replication potential.</td>
<td></td>
</tr>
<tr>
<td>• Kinds of evaluations:</td>
<td></td>
</tr>
<tr>
<td>• Process — Management</td>
<td></td>
</tr>
<tr>
<td>• Impact.</td>
<td></td>
</tr>
<tr>
<td>• Increase in annual investment budget evaluated, from 4.13% in 2002 to 13.43% in 2004.</td>
<td></td>
</tr>
</tbody>
</table>


Towards the Institutionalization of Monitoring and Evaluation Systems in Latin America and the Caribbean

for example, child care and family welfare. These programs have been operating for about 40 years, but we don’t know their impact. We also have programs for small and medium size enterprises. In Colombia, a large proportion of the budget goes to the provinces, and a large proportion also goes to health and basic services, but we don’t know the impact of these expenditures. So we need to incorporate these items into the evaluation agenda.

We have to look at how decentralization is working and whether it is having a good impact on health and education goals, etc. We have a number of social and other programs that are part of this agenda; some of these programs have non-fiscal revenue coming from decentralized entities of the state, and they have never been evaluated. For example, SENA, the program for training workers, has been in existence for about 40 years and receives the equivalent of US$300 million annually, but it has never been evaluated.

Accountability

The main challenges in accountability are the need for policy guidelines on public information and what we call the public access to information law. Without such a law there will not be any legislative or institutional regulation governing the information that needs to be provided to citizens — the content, quality and reliability of that information. We have introduced some general policy guidelines and have made more information on government performance publicly available but these have not yet been formalized by a law (Box 6).

To promote accountability with civil society we are trying to implement a model with the private sector and with NGOs which have established a partnership known as Free Colombia. The idea is that governments should generate good quality, reliable results information for social control. And civil society organizations should monitor and evaluate the results, guided by experts whom they could mobilize. This would not be a substitute for regular accountability processes, including through Congress or within the Government. However, it would give more credibility to government results information and would reduce the possibility of governments using that information for political propaganda.

We need to train public officials in the concepts and tools of monitoring and evaluation through the regular courses we have started with the state and private universities in the main cities. The idea is that these experiences will be replicated at other levels and in other locations, using the same tools that are used at the national level. As discussed

<table>
<thead>
<tr>
<th>Achievements</th>
<th>Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Policy guidelines.</td>
<td>• Policy guidelines on public information and law on public access to information.</td>
</tr>
<tr>
<td>• Design and implementation of periodical information dissemination.</td>
<td>• Develop framework with civil society.</td>
</tr>
<tr>
<td>• Redesign reports to Congress.</td>
<td>• Dissemination of manuals and methodologies.</td>
</tr>
<tr>
<td>• Quarterly follow up and management control for the President.</td>
<td>• Strengthen information quality.</td>
</tr>
<tr>
<td>• Annual budget report on television.</td>
<td>• Develop alliances with regional universities.</td>
</tr>
<tr>
<td>• Quality control of information.</td>
<td>• International seminars, workshops and courses on M&amp;E.</td>
</tr>
<tr>
<td>• Promote accountability with civil society (Colombia Líder).</td>
<td>• Consolidate technological tools for public information.</td>
</tr>
<tr>
<td>• Manuals and methodologies.</td>
<td></td>
</tr>
<tr>
<td>• Alliances with universities (courses).</td>
<td></td>
</tr>
</tbody>
</table>

Box 6:

Colombia-Achievements and Goals 2005 – 2010: Accountability
previously, we need to standardize evaluation methodologies and apply those standards at the different government levels as well as in different sectors.

We also need to consolidate all the tools that we have for public access to information, using a web-based information system that contains all these elements. The important thing is that the common citizen can go online and see it in real-time; and this is the same monitoring tool the government itself uses.

The important thing is that there should be good accountability to the citizenry, because whoever is in charge of providing that information has to follow certain guidelines to ensure that people have access to it. So whatever goes wrong with government performance, citizens will know about it.

**Discussant: Carlos Pinzón**

Senior Advisor to the Executive Director of Brazil, Colombia and the Dominican Republic, World Bank

My comments are from the perspective of having been at the Ministry of Finance in Colombia for only three years. Although the institutionalization of the M&E process in Colombia has been underway for more than ten years it seems clear that concrete progress has been made during the last three years. It is a significant achievement to subject up to 14 percent of budget expenditures to evaluation.

It should be acknowledged that the progress has something to with the nature of the present Administration and its effort to strengthen monitoring and evaluation in general. Both friends and critics of our President say that he is focused on details. However, he stresses comparisons between present results and results from the past and this new approach has led to progress in M&E.

Currently there are three main processes that are being followed under M&E. Firstly, SINERGIA is the best known. It is notable that most of the expenditures being evaluated are in Colombia's investment budget. Secondly, we have another system where performance-based projects are being evaluated and benchmarks are being set to see how different policies are being implemented. Thirdly, we have a monitoring process which is known as the Community Councils. The President usually goes to these council meetings throughout the country. Many different commitments are made by the central government and these commitments are followed up and results are analyzed and evaluated in terms of the different projects, and somehow their impact is also evaluated.

I would like to refer to the challenges we face and to compare them to what we had in the past. The first challenge is that although these M&E processes are highly institutionalized, having been in place for more than ten years and having produced results, nevertheless we have seen an increase of seven percentage points of the size of the budget in relation to GDP. So it is important to maintain the process of institutionalizing M&E, although it is worthwhile to question here whether this level of expenditure should become institutionalized. Here I should add that Colombia is a very legalistic country; everything has to be done through laws or legislative means, and there are laws for everything, even though things don’t work.

With all three of these M&E processes, the degree to which the public is consulted on policies is unfortunately very low. In other words, we have the tools available but the public really does not have ownership of them. There is still the hope that the evaluations will become institutionalized and lead to changes in the budget allocation process; I believe that we would then have a greater public ownership of those M&E processes.

Another challenge we face is the cost of impact evaluations. Until now these evaluations have been conducted largely by international consultants, and this has resulted in good-quality evaluations but which are very expensive for a country like Colombia. So how can we achieve a quicker transfer of skills so that evaluation can be done locally and, at the same time, maintain credibility of the evaluations?

Another point that has come up is the creation of sub-systems for M&E. As was said, we have a large central government package or set of proposals and in addition, other M&E processes proposed by local governments. Yet there is limited capacity to do credible monitoring and evaluation at that level.

The important thing with these M&E processes is their cost-effectiveness. What will we get out of all this effort? As long as these evaluation processes cannot be linked to budgetary decision making — that is, unless we have some form of performance or results-based budgeting — we are going to end up with real problems. As much as 95 per-
cent of Colombia’s budget continues to be very rigid and the only feasible means of changing this is to amend the organic budget law, but this is not taking place. These evaluations should serve to engage society and the government on the need to refocus or reallocate public expenditures toward those activities that are more efficient, and not to continue with a budget where expenditure patterns are rigid or inflexible.

**Discussant: Gladys Lopez-Acevedo**  
*Senior Economist, The World Bank*

As already mentioned, the monitoring and results-based component of SINERGIA is the oldest and DNP has worked very hard on the latest one, which is the social control component. SINERGIA has gone through several changes. A government regulation has been drafted with the aim of further advancing SINERGIA in an integrated and comprehensive way. Examples of the goals put forward in the new legislation and presented in this conference include aligning the presidential goals with the results-based component of SINERGIA. In this regard SINERGIA has made tremendous progress because the goals are now managed by DNP, rather than by the Office of the Presidency as previously. This change will ensure consistency with SINERGIA’s system of performance indicators.

The other change highlighted was the creation of an inter-ministerial evaluation committee responsible for defining the evaluation agenda for every administration and amongst other activities for developing the technical guidelines for evaluation. The intention of this committee is to build a coalition of M&E champions which could ensure the transition of the SINERGIA system to the next administration. Continued work in this area will be critical for the institutionalization of the system.

On the social control component, DNP, together with the other ministries, has worked very hard on establishing strategic alliances with civil society to make information accessible. An example of this is the recent release of the system online, so that any citizen can access it and track the performance of expenditures. It is very important to make this information accessible to the public.

As mentioned, there are challenges as the system evolves because it is important to ensure consistency among the components as it expands to the line ministries and to sub-national governments. Also, to achieve a truly results-based budget, DNP will have to work very closely both with Congress and the Ministry of Finance.

Finally, building institutional capacity for monitoring and evaluation will require increasing the skills of the private sector and of academia to undertake impact evaluations.
M&E of Public Expenditures in Peru

Country Presenter: Fernando Zavala
Deputy Minister, Ministry of Economy and Finance

The Need for a Monitoring and Evaluation System

Beginning in the 1970s we attempted to set up a system of indicators and monitoring in Peru. It didn’t have the impact we wished and we are now redefining it.

Budget revenues are increasing, and thus there is potential for increasing government expenditures. We need a good monitoring and evaluation system so that this will lead to more efficient expenditures. In Peru we began the decentralization process a couple of years ago so we need a system that allows us to monitor decentralization. We also need more information for civil society which actively debates public policy, but is not always well informed. All of this means that there is an urgent need to develop a good M&E system.

In Peru, 52 percent of the population is poor, and, more than 21 percent are extremely poor; only 41 percent of the extremely poor have access to potable water.

In short, resources need to be used efficiently. Increased spending does not necessarily mean an efficient use of resources. Expenditure on social programs has been increasing, but this reflects salary increases; it does not mean that the infrastructure of schools or the availability of educational material are being improved, or that the teachers who are earning more are providing a better service. We have not seen a corresponding increase in student test results. And maternal mortality in Peru is among the highest in Latin America, together with Bolivia, Guatemala and Haiti.

So this is why we need M&E. This does not mean that there has been no progress in M&E, but we need to move ahead even further.

Thus far in practical terms, we have a law for fiscal transparency to promote fiscal sustainability and growth, and the law is being complied with, in general terms. We have a macroeconomic multi-year framework, under which we are programming public resources over the following three or four years. This information has been published since 1999.

We are also undertaking ex-ante evaluation of investment projects such as infrastructure and other types of capacity-building. But we need to broaden the concept of public investment in the national budget.

We have a national strategic plan, and sectoral and institutional strategic plans; that is, three different strategic planning systems. Our mistake was not developing a relationship between these plans and the national budget. So there was no incentive to meet the objectives. Citizens did not understand the indicators being used to evaluate the sectors and the institutions. And an effective sectoral viewpoint was not developed because there were sectoral indicators with no responsible party associated with them.

Given these problems we have formulated a new vision. We also have a management-by-results agreement between the National Budget Office and state entities and similar agreements with public enterprises. In recent years there have been two important mechanisms involving civil society and public management. The first is a national agreement with the participation of political parties and civil society to review strategies for the future. Secondly, we have a poverty
unit which monitors the poverty impact of public expenditures. So there are some tools on which we wish to base this new monitoring system.

The fiscal responsibility law and the multi-year macroeconomic framework have fiscal sustainability indicators. The national system for public investment (SNIP) deals with investment projects, but not with current expenditures. Then we have the integrated system for financial administration (SIAF) which is managed by the Ministry of Economy and Finance and produces budget execution reports for each sector and governmental unit.

Who should carry out the monitoring and evaluation? Perhaps it could be done by the Prime Minister's office or perhaps a more independent institution. After several months of discussions we decided that since we didn't see other sectors doing it, and because it operates across all three planning systems discussed above, that the Ministry of Economy and Finance should be the organization to establish the M&E system.

There are pros and cons to this decision. We made our decision on the basis of the technical capabilities and systems that we already had within the Ministry. We had other mechanisms but they did not meet the needs we had identified. But I think in the future, our monitoring system needs to be managed by another public sector institution.

In the new monitoring and evaluation system, SSEGP (Seguimiento y Evaluacion del Gasto Público), we need to look at results and to relate them to budget allocations. The weak points of our previous monitoring systems were that they had no relationship with the budget. Thus the executing units had no incentive to achieve results. And we did not have enough information and tools to reformulate programs — and we still don't. But we have a number of independent evaluations that have identified serious problems in the social area.

So in summary, we needed a national system for monitoring and evaluation, to support multi-year programming, and for it to be linked with the budget process. And we want the M&E systems to be useful for civil society's oversight of government performance.

This is the analytical framework. The World Bank and Chile, Brazil, Mexico and Colombia have helped us. In Peru we sometimes start things a bit later, but we are able to learn from the good experiences of other countries.

We don't use the term 'government reform', because this is often perceived as related to budget cuts. Rather we talk about results-oriented or results-based management. We talk about organizational structures, strategic planning, incentives, M&E, and control and auditing. The ultimate objective, of course, is to improve the well-being of the population.

One of our problems in establishing indicators in the past was that all the indicators were impact-based. This meant that the results were very broad, with no real ownership. None of the sectors felt that they were directly involved with, or responsible for, impacts.

I mentioned previously the example of education. We have problems with our educational system and our students do not reach the achievement levels that we want, specifically when they are tested using international benchmarks. Perhaps the student does not have access to the materials at the start of the school year. So a very clear and operational indicator in the framework of results-based management is to identify the time when the student needs to get the materials.

The Monitoring and Evaluation System
The objectives of the system for monitoring and evaluating public expenditures (SSEGP) are to improve the quality, availability, demand for and use of information, so as to improve public sector management and decision-making processes. We want the information to be used to design policies, but we also want civil society to participate in the use of that information. There are several civil society associations that participate in and work with the monitoring system.

We also have information from the sectors. Although we worked with the sectors from the very beginning, we made a mistake in having the indicators coming only from the Ministry of Economy and Finance. Now they come from all sectors and this helps create ownership. Information from the national statistical office is also used.

In Peru, specifically with health, and to a lesser extent in social and food distribution there were programs that provided information only within their own sector with no independent or external evaluation. We are now using the information that we have to create a national system, which will integrate all the sectoral information and provide a more comprehensive system.
We want to look at the strategic impact and what we call 'social accountability'. We have two different roles of actors in our SSEGP system. First, as producers or participants, they contribute to the selection, generation and dissemination of information, and, second as users they demand and use the information created by the system in their work.

We have a range of actors in our SSEGP system. Policy-makers define priorities, provide information, and use system reports. The municipalities and other executing units also give information and receive it, as does civil society. The oversight or control institutions — the National Audit office, the Ombudsman office and Congress, which have been participating from the outset, use the reports to discharge their supervision and control functions. But control organizations only receive information. In Peru we have a very good control system through the National Audit Office which looks at budget execution and financial procedures, but its work does not extend to results-based management. We are trying to encourage control institutions to participate in the implementation of the system. As time goes by and we adjust the system they may also become information providers.

There are different types of evaluation for public programs. When developing M&E we could start with impact evaluation or program reviews, or several approaches at the same time. We started with program reviews in Peru.

**Evaluation Component**

At the beginning of the 1970s different evaluation methodologies were implemented. These started with particular projects related to loans from the World Bank and the IADB. We started a process to assess the impact of these projects. We also included other projects that were not only social in nature, but were investment-related — some of these projects may only have a neutral impact on the economy if they don’t have a good monitoring process. We had ex-ante evaluation only for public investment projects but they were quite isolated from the budget process and the information was not widely disseminated — perhaps only within the executive branch of government or within a specific sector (Box 7).

**Monitoring Component**

Performance indicators need to be linked to strategic outputs and to produce information at the sectoral and institutional levels. How many performance indicators should we have? The lesson we have learned is the fewer the better. With 26 sectors and 2,400 institutions, including 1,800 municipalities, how are we going to implement the system? We have national, regional and local levels of government but we need to focus on the national level, which is the largest. In a few years we are going to be able to further decentralize and assign more functions to the municipalities, each of which will have its own M&E system.

Use of the information, at the outset, is important. We need to have better citizen participation, and oversight of the actions taken by the government — a culture of oversight.

<table>
<thead>
<tr>
<th>BOX 7: Peru-Evaluation Component</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Background</strong></td>
</tr>
<tr>
<td>Perú’s evaluation culture</td>
</tr>
<tr>
<td>started in the 70s (Ex Post: qualitative and results oriented evaluations).</td>
</tr>
<tr>
<td>The first serious evaluation</td>
</tr>
<tr>
<td>methodologies started in the 90s (Foncodes (1994) and PROjoven (1997)).</td>
</tr>
<tr>
<td>Impact evaluation was</td>
</tr>
<tr>
<td>consequence of the requirements of International Financial Institutions.</td>
</tr>
<tr>
<td>Programs such as PROjoven and Fonempleo managed to institutionalize evaluation practices: <em>Impact Evaluation</em> (Yamada 2005).</td>
</tr>
<tr>
<td>To improve the effectiveness of the projects the Sistema Nacional de Inversión Pública-SNIP was implemented in 2000. <em>Ex Ante Evaluations.</em></td>
</tr>
</tbody>
</table>
We have evaluated public investment expenditures, which comprise eight percent of public expenditure. But we only evaluate two percent of the recurrent budget; we need to evaluate 98 percent. We are phasing this in and I think that we are going to cover about ten percent of recurrent expenditures, specifically core programs such as health, education and food. We also have to define what programs are able to be evaluated and are going to be subject to evaluation (Box 8).

**Participation and Oversight Component**

We need a social oversight and participation program. At the Ministry of Economy and Finance we are looking at issues which will go beyond the current administration. So civil society has to lock in these systems to ensure that they continue even after a particular administration has left office. The M&E system will be sustainable if civil society and the business sector help to lock it in.

So we want the whole process to support the budget system but we also want the process to work at the social oversight level. The strategy for participation is to use existing systems, both at the civil society and the sectoral level. We started with civil society participating in the preparation of the budget at the regional and at the local levels. Within certain parameters they decide what the budget is going to be — they decide on the projects. The monitoring system now has to support this participation, particularly at the municipal level.

**Conclusion**

Right now we need a system that can achieve results in the short term and give us some insights for the long term. When we tried to develop an M&E system six years ago, the model did not work. Now when we are trying to implement the new system, this is costly so we have to think carefully as we move to establish it.

We have to create a culture of M&E, meaning a climate in which program evaluation is accepted as worthwhile.

**Discussant: Carlos Ricse**

*Ministry of Public Health, Peru*

My first observation is that when we focus on M&E it is sometimes viewed as an end in itself. But M&E is only a tool to help make decisions. It is something that contributes to government management but is no more than that. Evaluation is a tool to support planning and budgeting, while monitoring has to do with institutional management and execution — to achieve efficient use of allocated resources.

I want to say a few things about evaluation as well as talking about monitoring and management and planning.

In Peru the budget is very rigid, with 80 percent to 90 percent of the expenditure being allocated to items that are fixed or difficult to change. Perhaps we have less than 15-20 percent of resources that are allocated to goods and services or what we call discretionary items. This rigid budget is illustrated by the case of health. We have goods and services and payment of personnel and social benefits. In 2005, however, the overall situation became even more rigid, through the raising of salaries for doctors, as well as for policemen and teachers.

How are we going to have good planning if our budget is so rigid? What opportunities do we have to use evaluation in these circumstances? Rather we have to look at the tools we use to allocate resources and we need to have better planning, while not distorting political priorities, at various levels — national, sectoral and local, so that we can achieve results-oriented performance.

I would like to focus now on monitoring and on management. Once resources have been allocated, they should be targeted towards their desired final outcomes. Activities or actions that are part of an operational plan must be well-designed in order to achieve results or impact. The indicators should be able to measure the logical sequence between
resources or inputs, outcomes and impacts. Monitoring and evaluation must complement each other in order to determine how we achieved those results. A good evaluation should combine different monitoring and evaluation aspects which should provide information for both external and internal users of this system to make good decisions based on that information.

For example, the indicator of maternal mortality rate in Peru, which is one of the highest in Latin America, is strongly influenced by the percentage of institutional or hospital deliveries. In hospital deliveries we need skilled professional staff, good provision of medications, and appropriate use of equipment by these professionals. But to achieve this we need certain inputs and a budget that supports this activity all year long. There are two related subsystems for administrative data, geared towards measuring these professional inputs and processes, and the results they produce. This information gives us an idea of how we are doing at these different levels. Such information is available in Peru through national registries at the Ministry of Public Health.

We have developed specific indicators for each one of the levels. When we measure maternal mortality rates we gather that information through public health surveys, where we have a baseline, let us say, of 85 deaths per 1,000 live births. And we have information that shows us how we have been progressing at the hospital level, for example the success rate of the use of oxytocin which will prevent miscarriages, and the use of other medications used for hypertension or preeclampsia. These are other measures of performance that we use.

Also, we monitor each of the specific tasks under those activities, for example to identify how the staff at public health centers can use oxytocin or other medications and whether they are using them properly or not. We have a range of additional information on the physical and financial performance of those institutions. And we have information on the training agenda for staff.

To conclude, I re-state my three main points. M&E is not an end in itself. It is a system that supports the decision-making process. Secondly, evaluation is only an input into the political decision-making process. Finally, monitoring is an additional element of good management to ensure that the allocated resources are properly used.

**Discussant: Livia Benavides**  
*Senior Social Sector Specialist, The World Bank*

Peru is just starting its M&E system and at this time it is facing more challenges than real success. There are some aspects which are very important to highlight.

On institutional arrangements the Ministry of Economy and Finance has already established a network, with a focus on the social sectors. The Ministry also has a very important and useful integrated financial management information system (SIAF). This system is well recognized throughout the region and is facilitating the implementation of the broader M&E system — including at the local and municipality levels, which is important, given that Peru is going through a decentralization process at the present time.

There are some areas that are very advanced with their information systems and they have become net exporters of information to the Ministry of Economy. On the other hand, there are other sectors which have difficulties in managing their information. They are probably causing gaps in information which will not make it possible in the short term to monitor public expenditures. There needs to be training and skills-building in those areas. So it is a challenge to establish the system within the Ministry of Economy and to receive the information from the other sectors.

One encouraging feature is that the number of indicators is under control. From the beginning the Ministry of Economy has been very concerned that the number of indicators in the national M&E system should be very limited and to leave the sectors to do their own, more detailed performance monitoring for their own purposes.

I believe that the indicators being used in health and education are the most important ones for these sectors. However we are concerned about those in social protection, especially those in food assistance programs and social infrastructure programs, because those programs do not currently have a very clear logical framework connecting inputs to processes to outputs to outcomes. The monitoring is really in the area of inputs, and such indicators are not much help in decision-making.

Finally, since this M&E system has started recently, it is not yet having a major impact on decision-making — on budget-related decisions. This will be one of the greatest challenges in Peru, where so much of the budget relates to recurrent expenses and especially when it comes to non-results.
related salaries. There are shortages of funds for both educators and health workers. The indicators show that human skills have major effects on the results or impact indicators.

As for program evaluations, Peru faces an even greater challenge to create an evaluation culture. In general, program evaluation has been taking place more due to outside pressure, with only a few exceptions. We need to change this. There are some programs, about to begin, where it will be very important to support monitoring and evaluating from the very beginning by having a clear design for the evaluation. This applies to social protection programs in particular.
WE HAVE HAD FOUR VERY GOOD COUNTRY PRESENTATIONS TODAY. I HAVE SOME KNOWLEDGE OF BRAZIL AND CHILE THROUGH HAVING WORKED THERE WITH THE BANK AND OECD. THE EXPERIENCE OF COLOMBIA AND PERU ARE NEW TO ME, AND I THINK THOSE TWO ARE PARTICULARLY USEFUL ILLUSTRATIONS OF AN ATTEMPT AT A COMPREHENSIVE APPROACH TO M&E. THESE PRESENTATIONS RAISE FIVE ISSUES WHICH WE CAN CARRY OVER INTO TOMORROW'S DISCUSSION.

FIRSTLY, THIS WORKSHOP IS ABOUT INSTITUTIONALIZING M&E. TO WHAT EXTENT HAS THE EVALUATION PROFESSION OR COMMUNITY (IF IT IS A COMMUNITY OR PROFESSION — OR CAN ANYBODY WITH A GOOD BRAIN DO IT?) FOCUSED SUFFICIENTLY ON INSTITUTIONAL ISSUES AS OPPOSED TO TECHNICAL OR METHODOLOGICAL QUESTIONS? INSTITUTIONALIZING IS MORE THAN JUST ESTABLISHING SYSTEMS. WE HAVE HAD SOME VERY IMPRESSIVE PRESENTATIONS OF QUITE MATURE SYSTEMS, PARTICULARLY THE CHILEAN SYSTEM OF EVALUATION. OTHER PRESENTATIONS ALSO INDICATED ORGANIZATIONAL STRUCTURES AND COMMITTEES, COMMISSIONS ETC., THAT HAVE BEEN SET UP TO INSTITUTIONALIZE THE M&E PROCESS. BUT THE INSTITUTIONAL ISSUE IS USING THESE STRUCTURES — WHETHER THERE IS ANY FINAL IMPACT FROM THESE EVALUATIONS. THERE IS PERHAPS A RATHER NAÏVE VIEW THAT M&E, OR COMPONENTS LIKE PERFORMANCE BUDGETING, ARE LIKE 'BOXES' WHICH YOU JUST PLUG IN AND THEY THEN START TO WORK AUTOMATICALLY AND DELIVER RESULTS. WE KNOW THAT THAT IS NOT THE CASE, THAT THERE ARE BIG ISSUES OF INCENTIVES, BOTH POSITIVE AND NEGATIVE.

HOW DO WE PROVIDE INCENTIVES FOR THE USE OF THE RESULTS INFORMATION? OECD WORK, THAT I AM GOING TO TALK ABOUT TOMORROW, USES THE WORDS 'STICKS, CARROTS AND SERMONS' AS THE THREE POSSIBLE INCENTIVES; STICK BEING THE NEGATIVE INCENTIVE, CARROT BEING THE POSITIVE INCENTIVE TO ENCOURAGE YOU, AND SERMONS BEING PREACHING THAT EVALUATION IS GOOD AND, THEREFORE, SHOULD BE DONE SIMPLY BECAUSE IT IS GOOD.

THE SECOND POINT IS THE CHALLENGE OF M&E IN A DECENTRALIZING ENVIRONMENT. HOW DOES M&E WORK WHEN SO MUCH OF THE SERVICE DELIVERY IS DONE AT THE SUB-NATIONAL GOVERNMENT LEVEL, WHERE ALL THE INFORMATION IS HELD, EVEN IF WE ARE TALKING ABOUT NATIONAL PROGRAMS AND NATIONAL PRIORITIES? HOW DO WE MAKE EVALUATION WORK IN A DECENTRALIZED SYSTEM? THIS IS OBVIOUSLY A MAJOR ISSUE IN BRAZIL, AND WE HAVE HAD COMMENTS ABOUT ATTEMPTS TO REPLICATE NATIONAL M&E SYSTEMS AT THE LOCAL LEVEL IN COLOMBIA AND PERU.

THIRDLY, M&E COMPRISSES A RANGE OF DIFFERENT TOOLS, ARRANGEMENTS AND SYSTEMS. WE HAVEN'T TALKED ABOUT ALL OF THEM SO FAR; PERHAPS WE HAVEN'T EVEN TALKED ABOUT MOST OF THEM. WE HAVE TALKED ABOUT FORMAL EVALUATION SYSTEMS. WE HAVE HAD SOME DISCUSSION ON THE LINKAGE WITH THE BUDGET. WE HAVEN'T HAD A VERY EXPlicit DISCUSSION OF ONE OF THE TOOLS, PERFORMANCE BUDGETING. RESULTS-BASED MANAGEMENT IS ANOTHER ONE OF THE ARRANGEMENTS OR TOOLS OR PACKAGES, WITHIN WHICH AN M&E APPROACH MAY BE EMBEDDED. ONE AREA THAT HASN'T BEEN MUCH DISCUSSED IS THE AREA OF PERFORMANCE AUDITING. IN A NUMBER OF COUNTRIES, THE NATIONAL AUDIT INSTITUTION HAS A MAJOR ROLE TO PLAY IN THE OVERALL EVALUATION FRAMEWORK.

THE FOURTH POINT IS THE ISSUE OF PERFORMANCE INDICATORS, STARTING WITH CHILE'S DESCRIPTION OF ITS PERFORMANCE INDICATORS. CLEARLY, THERE CAN BE PROBLEMS WITH PERFORMANCE INDICATORS. WE CAN MEASURE THE WRONG THING. WHO LEGITIMIZES THE PERFORMANCE OBJECTIVES, AND WHO SAYS THIS IS THE WAY WE
should be measuring performance? Do the indicators relate to the organizational objectives? I took one comment on the Chile presentation, as a possible view that perhaps some of the Chilean evaluations are rather too narrow. Of course all evaluation is suboptimal. We are not reviewing every component of a program from a zero base. The question is where should we draw the boundaries and who legitimizes that decision.

The final point is how much evaluation is enough? How much can we afford to do and how do we use our scarce evaluation resources in the most cost-effective way? I was quite surprised at the modest cost of the Chilean evaluation system. Chile must pay its consultants a very low fee. It must be a great honor to participate in Chilean evaluations.

But looking at the Operations Evaluation Department and other parts of the World Bank where evaluation work is done, some of the costs there seem to me to be rather modest. But there is the obvious question of how much is enough, what to evaluate and when, and how to get the most cost-effective use of scarce evaluation resources. Keith Mackay has raised the important issue of whether enough is being spent on evaluation.

But it is not just a money question. There are technical issues of quality. We are not saying that anybody can do evaluation; we are saying that there are certain required skills. Maybe there is a profession called evaluation. Maybe we need licensed evaluators with codes of ethics, like the accounting profession or something like that — although hopefully they will be more ethical than that profession.

Conference Facilitator: Sonia Ospina
Associate Professor of Public Management and Policy, New York University

When talking about the goal of institutionalizing these systems, there are technical dimensions and there are political dimensions. Even though both of these are interrelated, for analytical purposes we can look separately at these two elements.

Under the technical, there seemed to be some recurring themes. One has to do with the design of the system itself and the objectives in the different systems. Each has objectives or goals, but some systems emphasize some more than others. Is it macro public policy or micro spending? Because all of this goes to the details of implementation and follow-up.

Second, under the technical dimension, there are questions of institutional design, and here we see several points. At what level do we evaluate? The question of decentralization is an important one especially in the processes going on in Latin America, with very important implications in terms of going from an evaluation of national programs to the evaluation of the other two levels of government, the state or provincial and the local. This is very important in terms of the institutional design, as well as governmental relations of a vertical type.

Another issue of design is the relationship between planning, budgeting and evaluation, and related to this, the relationship among the different organizations in charge of these processes, and their integration, as well as jurisdictional issues regarding the design of evaluation. In the Peru presentation, we saw a fourth system that needs to be brought in, which is personnel evaluation of the budget and also personnel evaluation in public administration, as a part of the institutional design that needs to be considered. And very much related to the other two is the question of expenditure re, the concept of expenditure as an abstract concept that is being used as if it were the same in all of the countries, when in fact in each one of the systems it seems to center around the question of what can be evaluated, and this creates very interesting variations that may be studied later. There is also the rigidity of budgets, which is a problem of all Latin American countries and not exclusively of Peru or Colombia.

Related to institutional design and what is subject to evaluation, is the issue of what cannot be evaluated according to the methodologies we have available. We should not abandon the option of using evaluation for these activities; we may be able to create new methodologies in order to evaluate what cannot yet be evaluated. So perhaps we need a little bit of imagination and creativity.

The discussion of the third technical element, capacity, showed some interesting concerns regarding the endogenous capacity of countries to confront the challenges of their systems and to achieve evaluation at reasonable cost. Issues raised included transfer of technology in order to be able to do evaluations within the country; also external evaluations among the different sectors and organizations within the country. The capacity is not balanced or equal. There is greater capacity in some ministries and sectors than others. In terms of capacity, I also include certain problems raised concerning performance indicators — the concern about
the quality of the information. This is a problem of capacity, as well as system design. Reducing the number of indicators seems to be the great truth, but there is also the question of quality.

Lastly, in terms of capacity, there is an implicit debate regarding how much self-evaluation and how much external evaluation. Self-evaluation may be done because there is insufficient capacity for external evaluations. This is a sufficient reason, but if it is also important to create capacity and organizational learning, then self-evaluation is important.

In the political dimension, there were only two large elements. One was the enormous concern regarding citizen and civil society participation, the importance of bringing social control and surveillance into the processes. There are two sides to this coin.

When there is pressure, there are the issues of how to offer the kind of information that is needed, and also how to generate citizen demand, because reference is made to an uneducated citizenry. So an education process is necessary. Then there is the question of access, how to disseminate the information in a user-friendly fashion; when we talk about civil society participation, this is more political than technical.

And the second political element is the relationship with Congress. It looks as if there is no demand from Congress for these systems and that is a question of concern in implementing these systems. It seems that Congress is the invisible player in Latin America and it looks as if, in some other countries, it has a much stronger and clearer function.
M&E of Social Programs in Mexico

Country Presenter: Gonzalo Hernandez
Director General of Monitoring and Evaluation of Social Programs, Social Development Secretariat (SEDESOL), Government of Mexico

Introduction

I would like to outline the efforts that we have made in Mexico, and, specifically, in the Social Development Secretariat (SEDESOL) to develop an M&E system. This is an important public policy challenge. When I arrived at the Secretariat a few years ago, I wanted to use matched samples and propensity scores, and other issues from my academic days. But it is important to have knowledge of public policy and not just of technical issues.

Firstly, we have to consider what the ideal M&E system might be. What is its objective? What are we doing? What do we have to do in the future to achieve such system? I will also discuss the creation of the system in SEDESOL, the way it works and the challenges Mexico is facing.

In SEDESOL we have a very interesting structure and we need to look at where this fits in the general structure of M&E in Mexico. When we think about social policy in Mexico we only think about social assistance programs, but that is not all of social policy. We would have to identify and examine social problems. Only after that will we be able to look at programs and operating systems.

Monitoring does two very important things. First, it monitors processes. It allows us to know whether a program and its objectives are working, and if not if this can be corrected. Also, it allows us to know if a given objective, such as reducing malnutrition or illiteracy, is being achieved as a result of the program. That is what we mean when we talk about monitoring processes and integrated evaluation of programs. But the story does not stop there. Evaluation and monitoring provide important information which allows us to redesign programs, budgets and operations, and to improve social policy.

M&E then are a catalyst for the provision of information, and this allows us to know where we stand vis-à-vis programs, where we have to go and also how we are going to use that information to improve social policy.

Demand for Monitoring and Evaluation

If nobody uses the evaluation information, the system is incomplete. We have to create a system through public policy, a system that automatically provides this information. A businessperson has a good monitoring system and I don’t think it would make much sense to talk about M&E policy in the corporate world. A businessperson knows, or wants to know, the impact of actions on profits. So what is somewhat automatic or immediate in the corporate world is not automatic and immediate in the social assistance world. In the social world one has to build this.

The first question is who wants knowledge of the benefits of a social program. In the corporate world clearly shareholders or company owners want to have an M&E system. But in the social assistance world, it is not clear who wants M&E studies.

In Mexico, for a number of decades, we had many social assistance programs. But when we asked about the specific
impact on the population of those programs, nobody answered. There was no specific request from people to know what that impact had been. Then, in Mexico, we got democracy. We wanted transparency. Then people began to request information.

An M&E system requires someone who wants to know what is happening. So when we have an M&E system, we need to identify this person and arrange for the information to be provided in a systematic manner.

The first simple question is what exactly is the benefit of a social program? The corporate world knows that the most important thing is profit — that is what they monitor. But when we talk about social programs we don’t know exactly the benefits that we need to look at or which is the most important benefit. So we need to systematize the process to have better information flows. We need to reflect on how we are going to use the information that is generated to make decisions related to the budget. We need to institutionalize the system to answer this question.

The Development of Evaluation in Mexico
How did the first evaluations come about and how were they carried out? During the 1990s, the first isolated evaluations of social programs, such as rural assistance and food and nutrition programs were conducted, and there were many efforts made to incorporate evaluations from the very beginning, but not very successfully. One of Mexico's programs is Progresa (Oportunidades), which was instituted in 1997. Its impact evaluation was instituted in a very specific and clear manner. That program had a very clear objective, and the evaluation had almost immediate results, taking into account two main actors. First, Progresa had IADB resources and the IADB required an external evaluation of the program. So here is a first actor requiring an evaluation — an international donor in this case. Secondly, the political actors at that time welcomed well-conducted evaluations of this type of social program.

Starting in 2000, the Congress required by law, an annual evaluation of all the programs conducted by the Federal Executive Branch. How did this occur and where did this demand come from? In 1999, having a more pluralistic Congress than in years before and facing a very difficult Presidential election in 2000, the then opposition in Congress demanded certain things to avoid a situation where social programs would be used to obtain or to buy votes, as perhaps had been done in the past. So the Congress started to require an external evaluation of all social programs, to achieve greater transparency and accountability. So it is important to bear in mind that Congress demands an external evaluation process not because of a budgetary concern, but more for accountability reasons. This distinction is important.

Then in 2001, the Under Secretariat for Planning and Social Evaluation was created at SEDESOL. Its main objective is to implement evaluation of social programs. This mandate of SEDESOL, given by Congress, led to the creation of the Under Secretariat and General Office, which is in charge of continuing with this mandate. In 2002, the Minister of Finance and the Comptroller General (the latter is now an independent public office), made an agreement to improve evaluations and thus evaluations presented to Congress must now meet certain minimum standards.

Finally, in 2004, Congress approved a Social Development Law which, among other things, institutionalizes the evaluation process and creates a National Evaluation Council, which will come into effect by the end of 2005. So once again it was the Congress which initiated a law creating a National Evaluation Council. This will give greater autonomy to the evaluations conducted in the future.

We should note two important influences on the launching of this system. First, there was the initiation of democracy and transparency in Mexico, without which I believe an M&E system would have never been very successful nor in great demand. Secondly, there were better techniques and tools of evaluation, which have made better evaluations of social programs possible.

Strengths and Weaknesses
Our first strength was that we changed the rules of the game, and this initiated a cultural change in the area of evaluation. By 2000, all of a sudden, overnight, a new evaluation of social programs was required. This caused a lot of surprise among many actors but it led eventually to the creation of a new evaluation culture or at least accelerated the move to that culture. The actors, including program managers, are forced to think in terms of evaluation, not always favorably, but not unwillingly. Nowadays, at least at SEDESOL, program managers know that an evaluation is going on and that they have to pay attention to it, whether for good or
bad. Actors are made accountable for the program. This institutional change whereby Congress required by law an annual evaluation of social programs thus accelerated social and cultural changes.

Now for the weaknesses or difficulties. First, the actors were not technically trained for this. The importance of evaluation was not explained at the ministry level. We were lacking technical training of evaluators and of decision-makers and policy-makers. Also we did not start with a good performance indicator system. So we had evaluation year after year where we were lacking very important inputs — performance indicators and other important information.

As mentioned above, Congress required external evaluations of federal social programs, not initially to make budgetary decisions, but for accountability reasons. Nevertheless it was necessary to link these evaluations, with a clearer, open budgetary cycle. Guillermo Bernal, the Budget Director, will talk more about this. The Ministry of Finance still needs to do a lot to incorporate the evaluation process and results into budgetary decisions.

Components of the Evaluation System

There are four different components of Mexico’s evaluation system and I will explain the rationale behind each of them.

Firstly, there is the external evaluation required by Congress of all social programs. This leaves aside, for reasons I will explain later, the second component — impact evaluations. However we have impact evaluation of only some of the programs, as I will also explain later. The third is a major component of the evaluation system that we are building, the creation of a results-based monitoring system whose main objective is to strengthen internal evaluation of programs while also serving as an input to and source of information for external evaluations — and for improving performance of social programs. Fourthly, we have an official measurement of poverty in Mexico, which started in 2001 and is intended to measure the extent of progress made.

External Evaluation

The external evaluations done for Congress are carried out by an external evaluator (Box 9). They are reported to Congress, to the Ministry of Finance, to the line ministry, and to the Comptroller-General annually, either in September or October, so that evaluations are taken into account in preparing the budget for the following year. The Ministry of Finance sets some minimum standards which the evaluation must meet, such as coverage, impact, cost-benefit analysis, and perceptions of users. There is a whole list of elements or factors that must be complied with even though in practice not all of them can be complied with because of lack of resources or shortage of time, but the most important are coverage and targeting specific objectives.

Here I would like to emphasize a point. Three years ago, when we evaluated social programs, our attitude was that of a judge or critic. This attitude did not lead us anywhere. In our annual evaluations it has been better for us to take the helm of the ship, together with the evaluators and the managers of social programs, by involving them from the very beginning of the process in the drafting of the terms of reference and by meeting together to discuss them (Box 10). Only in this way, we believe, can we have evaluations which are objective but at the same time useful, so that the information that will come out can also be used by program managers.

What do we do with the external evaluations? Two years ago, when we completed evaluations we submitted them officially to program managers, who would read them or file them or put them away. We don’t know exactly what they did with the results. Then we would submit the evaluation to Congress, the Ministry of Finance, and everybody involved. For the last few years we have made a small change that has brought great benefits: we now officially submit a
very summarized version of the evaluation to each program manager, showing the weaknesses and challenges that the external evaluator has found and asking them officially to tell us how they think the problems identified by the evaluation could be solved.

Thanks to this we have changed the way we used the information at SEDESOL. Nowadays, social program managers know they must have an action plan to solve problems detected by the external evaluation. This action plan is something that the Minister has on his or her desk, as does the ministry’s internal comptroller’s office. This approach has made significant changes in evaluation. Firstly, we know that evaluations are taken into account in decision-making, at least more than before. Secondly, the evaluator improves his work and spends more time on each recommendation because it is clear that it is going to be used in decisions about the program. Thus, we have better quality evaluations because the evaluators know that their evaluations are being used. Programs have had important changes made to them based on this new approach.

For the 2005 budget negotiations on each program, the Secretary had in front of her the official responses given by the program managers as evidence that the programs were leading to some benefits and that they were complying with the recommendations of the evaluators. That was a positive change at SEDESOL.

Thus the actors are required to do evaluations. Important findings are made, some more important than others, in order to improve programs; but also, important weaknesses are found.

However more training of evaluators is needed. The external evaluation is done for all the ministries, and every year we are evaluating about 150 or 200 different programs. Mexico does not yet have the technical capacity or human resources to conduct so many good quality evaluations. This also complicates the hiring or bidding process for evaluators.

In addition, the programs and the evaluators don’t always have available the basic information to do a good analysis. This is not helped if, as in many cases, the evaluation is seen as something coming from the outside and being imposed, and many times it is not done because of budgetary reasons.

**Impact Evaluations**

These are a bit more scientific, resembling those of Progresa and Oportunidades. A lot of emphasis has been put on SEDESOL having impact evaluations, but also supplemented by qualitative measures, which are as important as quantitative measures. We also get resources from the World Bank and the IADB to conduct impact evaluations and from the Mexican fund for science and technology, in order to increase the funding for impact evaluations, at least for the technical side of them. We have about eight or eleven programs, such as the milk program, nutrition program, habitat program, housing program, and youth and opportunities programs, in which we are focusing on improving evaluation. All these programs involve a large amount of resources. We do not want to have an impact evaluation for all programs, only the most important. The Ministry has taken evaluation very seriously, even though program managers and directors are interested in the program continuing on regardless, through different administrations.

*Progresa* survived as a six-year program and it even survived the transition to a different political party, because it was having great impact. What many program operators now understand is that if they don’t have a good impact evaluation that identifies good results, then the program is not going to survive.
Results-Based Monitoring

It is good that the Congress requires us to evaluate programs, and this has led to an important culture change, but at the same time we lack basic inputs. The external evaluator requires a minimum amount of information that is not always available. Many times it is almost as if we hired an auditor and we tell him ‘now that we have hired you, why don’t you do the bookkeeping, because we haven’t done that’. The ideal situation, as I presented earlier, would be for each program to have a certain level of internal evaluation so that the external evaluation only verifies certain data or corroborates certain information, and gives a general assessment of the program. To do this — and this is something we are working on, we need to strengthen the results-oriented monitoring system. This is still a major challenge.

In looking at a social program we can use the structure from inputs to processes to outputs to achieving goals. We already have many indicators in the first part. We have good budgetary and financial indicators, which are also required by the Ministry of Finance, Congress and the Presidency. Thus we are required to have indicators to see how the initial phase is going. However we have too many indicators which are not always useful, either for the Ministry of Finance or the Presidency or Congress, at this first stage. But for the later stages, we do not have enough results-oriented indicators. So with the help of the World Bank and others we are trying to develop a results-based monitoring system which has significant indicators throughout all the processes of the social program, not only at the beginning. This should lead to better performance and produce better information for future evaluations.

We want to convince program managers that it is important to have a good evaluation system. Even if Congress did not require us to do it by law, it would be useful for us to do it anyway to make good decisions. Such an approach would make it possible to have better and more efficient external evaluation and also later on, to have fewer indicators, which would be more useful for decision-making.

What steps are needed to strengthen this monitoring system? We had a very useful readiness assessment done by the World Bank about a year and a half ago. The World Bank came to SEDESOL and conducted this assessment in a one-week visit. We conducted M&E workshops which were a great success. The three day workshop we had in February 2005 was also very successful, with about 150 attendees. Many of them asked: ‘Why didn’t we have those workshops or seminars four years ago?’

Also, we have approached the Minister for Finance, the President and others to explain the system, and to seek their feedback on and support for this system, because they are also users of information. In due course we want to have fewer but more useful indicators. This is an important issue not only for SEDESOL but for our partners. We are in the process of hiring a consultant to strengthen this system using four pilot projects. We want to promote a dialogue among programs in order to exchange experiences in improving monitoring systems.

Poverty Measurement

Starting in 2001 a group of researchers was required to submit to Congress and to the government an official measurement of poverty in Mexico. About ten researchers got together and after long discussions and debate three official measurements of poverty were formulated. We know they are not fully satisfactory because this is very difficult to achieve, but at least we have three ways through which the government can measure poverty reduction. However, there is the challenge of linking those measurements with the impact of social programs on poverty. So this component is not necessarily linked to the other indicators, and this is a remaining challenge. Next week this team of experts will present the official measurement of poverty for 2004 and we will be able to see the trend since 2002 and 2003.

Conclusions

We have made a lot of progress in knowing that there are users of evaluation information —the Congress, citizens, external donors (Box 11). They are all good users and demanders of this information and their pressure will accelerate the provision of information.

We have also made great progress in identifying the benefits of each program. We have made great efforts in impact measurement but we still need to do a lot more to generate the required information. We still need to do a lot to strengthen and sometimes even to create results-based monitoring systems.

We have also made progress at SEDESOL in what to do with evaluation results, based on the system I already explained.
We have a lot more work to do if we are to better link program evaluations with budgetary decisions.

Overall, I believe the challenges are very significant, but at the same time, I believe that what we have done in Mexico at SEDESOL during the last three years has been very significant.

**Discussant: Guillermo Bernal Miranda**  
*Head of the Budget Office, Government of Mexico*

As a commentator, I would like to step back from my role as a public official in Mexico. I have been very fortunate to have had the opportunity to talk with Gonzalo Hernandez. We had never been able to personally speak about these matters in the past, and this gives me an idea of the problems we have in Mexico today.

**Oportunidades** is a unique program in the way it was established and accredited and also because of the very important monitoring and evaluation results it has provided. It is set up within the Treasury at the Ministry of Finance, contrary to the wishes of the other participating secretariats — social development, education, and health ministries. **Oportunidades** is based on the budget that had been allocated to a wide range of programs under the Secretariat for Social Development, SEDESOL. It is precisely this evaluation that allowed us to overcome the resistance of the different secretariats, as well as of Congress itself, which was critical for the program.

In Mexico, the legal requirement to have goals and indicators within the budget began in the 1980s, when planning legislation was instituted. This law established that there would be a development plan and sectoral plans and that these would then be reflected in the programs and budgets submitted to Congress. The budget had to include indicators for each program. So for many years there was the bureaucratic custom of defining more than 3,000 indicators in the budget, in very broad terms and with goals that were usually generated by the budget managers. Basically it was a dialogue about numbers but not about goals, and it was also reviewed in this light by Congress, with kilos and kilos of paper produced but not used. Rather, budget discussion was on the different items or inputs for different offices or organizational units. So Congress controlled the allocations, say, for blue pencils and green pencils, and it allowed the Executive certain flexibility to decide whether to purchase blue or green pencils. Congress was not interested either in results or indicators of the programs.

As Gonzalo Hernandez well said, there has been a change in Congress and there is now the obligation to evaluate programs annually, in order to see whether they should be continued or extended under the budget. However, this culture is not yet deeply rooted in Congress and most of the evaluations are of poor quality for the reasons given by Gonzalo — limited human capital and rigidities in contracting evaluation consultants. For example, the system facilitates the contracting of public universities, but not private universities or NGOs to undertake evaluations which might be more desirable, especially if the cost is lower. Thus the evaluations are of low quality and most do not meet the requirements for decision-making regarding budget allocations for programs.

During the last three years, we have moved significantly ahead in results-based budgeting. Sometimes, without Congress ‘buy in’ we have sought the discussion and approval of the budget at an aggregate level; we have tried to put aside the presentation of the budget by office or organizational units. Rather, we have tried to organize it in such a way that the programs are linked up with agencies, which we refer to as institutional entities. These institutional entities are the ones that receive an indicator. Up to 2002 there were no results indicators for the program **Oportunidades**.

---

**BOX 11:**  
**Mexico-Goals**

- To achieve an integrated M&E system.
- To strengthen the M&E system and internal evaluation.
- Have a stronger link with budget evaluations.
- Improve the skills of local evaluators…
- …and also of the potential users of the evaluations (Congress, etc...).
- Improve qualitative evaluations.
- The transition to a new National Evaluation Council will be neither easy nor automatic.
- There is practically no evaluation at the state or municipal level.
This allows for the different offices under different ministries, to participate in a single evaluation effort, to be reflected in a single result or a simple indicator, generally generated by the institutions that participate in the implementation of the program. In this way we have reduced the number of indicators from 450 to 350, and we think an optimum would be 250 indicators for the budget.

There have been some fortunate events. President Fox set up a system of presidential goals under which government offices make a commitment with the Presidency. Thus, important goals for government offices become the President’s goals and then these are translated into the budget.

There are numerous problems yet to be resolved — for example, allowing managers to have a greater participation in the definition of indicators, and to improve Congressional review of the budget along the lines of the program and its indicators, and not just on the basis of specific allocations by office. Certainly the legislation provides that there should be better quality evaluation and more funding for evaluation. I generally do not agree with additional expenditure, because I think money not spent is well spent. However, I believe we are going to find a solution to the political problem in Mexico, concerning the conflict between the Executive and the Legislature in approving the budget.

**Discussant: Mark Hagerstrom**  
*World Bank Country Sector Leader for Human Development*

For me, it is extremely difficult to say enough about SEDESOL in the area of evaluations. I believe that the role of SEDESOL in terms of institutionalization and recent progress in result-based management is very important. We have heard a great deal about SEDESOL’s **Progresa** and **Oportunidades**, but we need to see whether or not it is an island of excellence. We need to evaluate the impact of SEDESOL on monitoring and evaluation in general in Mexico, and it may be possible to provide lessons for other countries on the promotion of M&E.

We know that SEDESOL represents only around ten percent of the social federal expenditure, perhaps a little less. But it is important to understand that SEDESOL has had enormous impact on M&E in Mexico. The Fox administration has demonstrated a kind of empowerment in terms of **Oportunidades**. The administration always talks about the importance of accountability and the impact of expenditure. Showing the rest of the world what SEDESOL is doing provides a standard for other programs.

Then we have the impact of SEDESOL on the cabinet. Today the cabinet is always talking about impact and there seems to be a kind of competition among the different ministers to demonstrate that their programs do have an impact.

The government also has a program for measuring poverty. That has a significant role in terms of promoting M&E and also for the first time we have, in relation to the Bank, an opportunity to review federal public expenditure. I believe it is the first time that the government has undertaken this type of analysis.

Certainly, SEDESOL brings information into the light of day. Information is not simply something that gathers dust on a desk at the offices of the Ministry of Finance.

Now, we can talk about other sectors. We have the example of the **Seguro Social** for which an evaluation has been provided, and I am sure there are other examples from the Secretariat of Labor.

In education, we began with an evaluation of a compensatory educational program for poor rural children and while there was a great deal of initial resistance, the evaluation is now part of the program. Another example, which included World Bank support, was a program for providing loans to students and there is now also a program to set up quality schools. In education, there is also an evaluation of training of judges, being carried out by the Rand Corporation.

We are now undertaking a very important effort in impact evaluation with the Ministry of Education. We should note the importance the Minister of Education attaches to **PISA** (Programme for International Student Assessment) — the students exam or test results. PISA is a voluntary scheme and a request needs to be made to the OECD (Organisation for Economic Co-operation and Development) to participate in it. But the results of these tests are being made public and coverage is being broadened to include states. This will have an enormous impact, because those states that assume responsibility for basic education are the ones who are taking on this evaluation. There hasn't been a great deal of publicity about outcomes in education.

As occurred in SEDESOL there was a restructuring in the Ministry of Education under which they set up an office for planning, which reports directly to the Minister for Education and is in charge of evaluating the inputs and
the outputs of the programs, as well as monitoring of the processes. This is a very recent restructuring that deserves support.

We already know that in the delivery of social services, the states play an important role in Mexico. I would like to know what Gonzalo thinks of the strategy of promoting M&E at the state level. I think that is the next important challenge in Mexico.

Finally, there has been some feedback on the impact of SEDESOL on M&E. The World Bank’s chief economist, in my area of human development and social services, is someone who has earned his fame in the area of evaluation and Oportunidades. He is now distributing methodologies and promoting M&E in many other countries. So we at the Bank should thank Oportunidades and SEDESOL for this.
PART THREE:

Synthesis Session
Institutionalizing Monitoring and Evaluation – Issues and Experience in OECD Countries and in Latin America

Conference Rapporteur: David Shand
World Bank Consultant

Section 1: General Issues in Evaluation

‘EVALUATION PERSUADES RATHER THAN CONVINCES, argues rather than demonstrates, is credible rather than certain, is variably accepted rather than compelling’.1

Introduction
Evaluation has been described by the Organisation for Economic Co-operation and Development (OECD) as a systematic, analytical assessment of performance.2 But this is best regarded as monitoring — assessing whether performance is at the level it should be. Evaluation goes beyond monitoring by also searching for explanations of performance, so that performance can be improved. This distinction between the roles of monitoring and evaluation seems to have been, in the end, well brought out in the discussions at this conference.

The concept of evaluation is defined in multiple and even conflicting ways and there are the usual problems of terminology, let alone those of translation from English to Spanish. The terms reviews, scrutinies, audits, assessments, policy analysis, etc, are often used interchangeably with the term evaluation. Any decision can be said to involve an evaluation of some sort. But this may debase the concept of evaluation.

On the other hand while we may avoid debasing the term evaluation in this way, there may also be a need to demystify the concept, so that it not perceived as a necessarily highly technical exercise, beyond the understanding of mere managers. This raises the interesting question of whether evaluation is a ‘discipline’ which requires its own specialized training and skills — or whether ‘anyone’ with a good analytical brain can do it. What is the role of subject specialists as opposed to professional evaluators? Increasingly the idea of mixed teams, comprising both, seems to be accepted as good practice. The country presentations in this conference have shown cases of this approach.

Evaluation may have a history of unfulfilled promises. To some extent this may reflect unrealistic expectations of what it can do. Evaluation is subject to both methodological and management or behavioral limitations. In terms of methodology there are many subjectivities and judgments involved. Some have suggested it is more an art than a science. Much depends on the objectives ascribed to what is being evaluated. Even then, how do we determine cause-effect relationships in complex programs? The difficulties and cost of impact evaluations have been a recurring theme in this conference. And even if an evaluation is methodologically sound, or perhaps as sound as it can be, how do we ensure it is linked with decision-making processes?

It can be argued that many of the problematical issues concerning evaluation are institutional or behavioral rather than methodological — so it is entirely appropriate that this conference focuses on this issue. Evaluation is not just some ‘black box’ that can be plugged in and automatically deliver results. Its institutional framework must be carefully designed. For example, what are the incentives to carry out evaluation and then to use its results?
Related to this, evaluation information is always to some extent imperfect. Quite apart from any methodological issue there are costs of obtaining information. How far do we go in obtaining relevant information? Some have described this as avoiding ‘paralysis by analysis’; the issue remains of where we cut off the information-gathering. This raises the important issue, frequently raised in this conference, about how much resources we should be allocating to evaluation. Evaluation may expensive. We not only need to use scarce evaluation resources efficiently but we need to decide how much is enough. Some OECD countries have suggested a benchmark of a certain percentage of program costs, say 3 or 4 percent. But there will be a reluctance to allocate such funds unless there is a demand for evaluation and it provides clear benefits to decision makers.

A number of other issues or choices about evaluation present themselves — and are discussed in more detail later. However, they are not necessarily mutually exclusive so it is not a matter a choosing one or the other:

- What is the desirable timing of evaluation —should the focus be on ex ante evaluation before a program or project commences, or ex post? Or both? Clearly, they are not mutually exclusive and each is widely used, as indicated by the country presentations, but perhaps with greater emphasis on ex post, or evaluation of existing programs in many cases, perhaps reflecting that data exist for ex post evaluation which may not exist for ex ante evaluation.

- What should be the scope of the evaluation? In other words what is accepted as given, or to what extent should the objectives which govern the evaluation of the program be questioned. And who legitimizes these objectives, thus defining the scope? All evaluation is sub-optimal in the sense that we are not reviewing every aspect of every program. But who decides what is to be accepted as 'given'?

- What do we perceive as the purpose of the evaluation? Is it to make a judgment on performance, to determine whether some agreement or contractual relationship has been met, something stressed on OECD discussions of 'new public management' which is allegedly based on a contractual approach? Or is it to determine how the performance of the program can be improved? Several of the country presenta-
Finally, there is the challenge of increased decentralization in most governments. How do we manage evaluation where national governments have passed major service delivery responsibilities to sub-national governments? Who is accountable and where is the information? What information is needed by the centre to monitor (for what we may call ‘strategic control’ purposes), and what information is needed by program managers for their operational decision-making? A number of the country presentations have stressed the need to develop M&E systems at the sub-national government level, and to link these with national M&E systems. But it is worth noting that in many OECD countries, including the United States, sub-national governments are ahead of the national government in developing their own M&E systems. This is in contrast with the general position in Latin America.

We now consider in more detail three key issues raised above.

Performance Indicators and Evaluation
Many governments have invested considerable time and energy in the development of performance indicators, usually based on programs. Often this has been approached with more enthusiasm than skill. Indicators are best regarded as part of the monitoring function — indicating whether performance is on track. They do not of themselves explain why performance is as it is.

There are many issues or problems with performance indicators, apart from the more general question of legitimizing the objectives on which they are based — in many cases performance indicators have been developed before objectives have been clarified or legitimized. In some cases there are too many indicators; while there is need for comprehensiveness there must be a balance with simplicity and understandability. There is a need for intermediate indicators when impact indicators are difficult to develop or have limited ownership. And given that ‘what is measured is what counts’ we must be sure that we are measuring the right thing. The list of indicators which may lead to perverse management behavior to achieve them is long. Some measures may be vague or subject to manipulation and involve playing the ‘numbers game’ — hospital waiting lists and crime rates are two such examples.

Performance indicators have been used primarily to compare performance internally within an organization with expected or budgeted performance. This has been the focus of the country presentations, where the difficulties of comparing performance between programs have been mentioned. However, international performance indicators are now increasingly associated with the rise of performance benchmarking — within and between organizations and governments, a development embraced by many OECD countries. This raises further problems of possible misuse (or over-use?) of such indicators. But there is a need for targets or baselines, whether external or internal.

The need for systems to collect performance information has also been noted at this conference. Most government information systems are focused on financial information, related to planning and budgeting systems, illustrated by Peru’s SIAFI system, rather than on information concerning outputs, outcomes and impacts, which tends to be collected in an ad hoc manner with consequent concerns about its reliability. But the Latin American region seems relatively advanced here compared to OECD countries, as illustrated by systems such as Brazil’s SIGPLAN and Colombia’s SIGOB, the latter being accessible by the public. This may reflect the strong emphasis on social control in Latin America, discussed below, and the existence of national planning systems, which as discussed later is not a feature of OECD countries.

What is Performance?
This has a number of dimensions, which can be set out in the following hierarchy. Problems of terminology (and translation between languages) can arise here. The important thing is that we keep our minds clear about concepts:

- Impact — on broad social and economic indicators, such as poverty and literacy rates or income levels. A key issue here is that these indicators are likely to be affected by many other factors apart from the program being evaluated; determining cause-effect relationships may be problematical. There has been considerable emphasis in the country presentations on the difficulty and cost of impact evaluations and the potential problem that if they are too broad no organizational unit is seen to be responsible for them.
- Effectiveness — achieving defined program objectives, which are generally referred to as outcomes.
• Efficiency — in terms of producing the activities or services (as reflected in outputs) at minimum cost, consistent with a defined quality level.
• Service quality — as reflected in measures such as timeliness, accessibility, suitability and continuity of service, such measures being sometimes reflected in service quality commitments (of which the United Kingdom Citizens’ Charter was an early example). These measures have both efficiency and effectiveness components. We can note that Chile has defined service quality as a separate component in its performance indicators.
• Good processes and systems — this may cover the adequacy of information systems, planning and management processes, internal controls, organizational structures and the quality of management and staff. Such systems and processes are evaluated against templates of good processes and systems established by the ‘management consulting’ profession.
• Financial performance — achieving budget targets, covering either revenues or expenditures or both. A key point to be made here is that evaluation is not just about impacts, outcomes and outputs; it must also focus on inputs. And reliable information on inputs is needed for evaluation — a point further stressed in the discussion on the New Zealand and Australia experiences in introducing full accrual accounting, as set out in Section 2.

Using Monitoring and Evaluation
There are a number of areas where monitoring and evaluation information may feed into decision-making. But these different uses are not mutually exclusive. These uses can be summarized as follows:

• For budget decision-making — improving the allocation or resources in terms of government priorities and seeking savings. Performance budgeting is a tool often used here. While this is not the place for a detailed review of such a complex issue as performance budgeting experiences, as a tool is does not appear to have achieved the (perhaps unrealistic) results expected of it.
• For the development of national or sectoral plans. However, a key issue is implementation, which has perhaps been a weaker feature compared with preparation of these plans, through their linkage with other mechanisms such as the budget.
• For performance management of programs, activities, organizations. This involves regular monitoring by program managers and their taking corrective action where necessary to improve performance. The Chile and Brazil presentations have noted the importance of M&E feeding into program management or program improvement plans.
• For establishing accountability (internal and external) and social control. As discussed above, performance information may be used to establish whether contractual undertakings have been met, for example in service delivery contracts. At a broader level, the external reporting of performance indicators and program evaluations can inform citizens on how well public funds are being spent. This social control element seems particularly strong in Latin America, perhaps reflecting its more recent democratization than OECD countries, where perhaps consultation with citizens and program users and public access to information, have been more long accepted parts of the government system.

All of these uses relate to the overall goal of decision-making to improve public sector performance. But it can be seen that linking with decision-making processes is key — with the budget, with performance contracting arrangements or with the development of national and sectoral plans.

As will be seen in the Section 2 summary discussion on OECD country experiences and the Section 3 summary of the experiences of the five LAC countries illustrated in this conference, the approaches to linking with decision-making processes differ considerably, reflecting their different approaches to public sector management and resource allocation. For example, no OECD country cited has a national plan. While the approaches are not mutually exclusive they do give rise to the question of which will have the greatest impact on improving performance or accountability. For example, in what circumstances might performance budgeting be the best lever or mechanism to achieve this improvement?

Unfortunately, there are many examples both from OECD countries and Latin American countries of what are in some cases high-quality evaluation and extensive
performance indicators which have no clear link with decision-making processes.

**Section 2: OECD Perspectives**

**Key Issues in Program Evaluation (OECD, 1997)**

The OECD undertook an extensive review of program evaluation practices in a number of its 29 member countries in 1997, with the objective of developing best-practice guidelines. While some of these points have been covered above, the list below provides a comprehensive outline of key issues as perceived by the OECD for successful program evaluation. Despite the lapse of time since this study there is no reason to believe that these issues are still not applicable:

- **Gaining support from the top** — from elected officials in both the legislative and executive branches. This requires at a minimum that evaluation by managers be encouraged. Desirably it also involves elected officials wanting and being prepared to use evaluation information in budget or planning decisions. However, transparency about results may not always suit their political objectives — but it can be argued that they should not keep the results secret from themselves. The challenge of getting Congress interested in M&E has been a continuing theme at this conference.

- **Generating effective demand for monitoring and evaluation at all levels of decision-making.** While there may be general acceptance that it is in everybody’s interest in the long run, a short term view may predominate. There is a need to consider incentives for evaluation, which may be classified as ‘Sticks, carrots and sermons’ — negative incentives, positive incentives and persuasion that evaluation is inherently good.

- **Setting realistic expectations** and particularly making it clear that evaluation by itself will not automatically increase performance and provide new resources. As mentioned above, evaluation is not just a box to be ‘plugged in’. Unrealistic expectations, inevitably followed by disillusionment, have led to cynicism in some quarters about monitoring and evaluation.

- **Systematizing evaluation activities** so that they become a regular activity, perceived as a normal part of political and management life, rather than some special or ad hoc exercise which may therefore appear threatening.

- **Linking with the budget process as far as possible.** It is important to stress that the link is not mechanistic but that the budget process offers potential to improve public sector performance. Performance information thus informs rather than drives the budget process. However, there is a record of mixed success in performance budgeting both in OECD and developing countries. One issue which may make performance budgeting problematical is inflexibilities in budget allocations, which limit possible reallocation of budget resources — an issue raised in several of the country presentations.

- **Locating the evaluation function for greatest impact.** This raises the question of which institution does it or leads it? Should it be the finance ministry, the planning ministry, the presidency, an independent organization (research institutes, a national evaluation organization, management consulting firms), the national audit institution, or line ministries and program managers themselves?

- **Planning evaluations** so that the objectives, scope and information requirements are established at the beginning of the exercise. Proper planning is a pre-requisite for success, to ensure the evaluators know their task and scarce evaluation resources are not wasted.

- **Timing evaluations appropriately** — to fit in with the policy and decision-making cycle to have impact. Otherwise, evaluation information may sit unused or become dated.

- **Meeting user needs and ensuring relevance.** Evaluations must be able and be intended to be used. This has two dimensions: evaluation reports must be understandable by those who are supposed to act on them; and the subject matter must be relevant. There is less point evaluating a program which a government is committed to retaining regardless of the evaluation, or putting it another way, where it ‘does not want to know’.

- **Involving the stakeholders**, to improve the quality of the evaluation and its likely impact. There are always multiple stakeholders including the government itself, the taxpayers who are funding the program,
and the consumers of the program — who may be a quite different sub-set of the population from the taxpayers. The desirable involvement of civil society has been a strong theme at this conference.

• Ensuring methodological quality. There should normally be arrangements for quality control such as peer or external review, particularly in a major evaluation. Ethical standards in evaluation are also important. Given the inherently subjective nature of evaluations (as discussed above in Section 1) it is desirable to be explicit about the limitations of any individual evaluation. It is important to avoid a situation where evaluation is mis-used to produce pre-determined findings, for example by setting program objectives in a particular way or by direct ‘cheating’ by misusing data or ignoring unfavorable information.

• Communicating the results to stakeholders. This needs to be done in a clear and timely way. The evaluation must be understandable by its intended users. Limitations and qualifications on the evaluation should also be provided. Options rather than single choices in using the information should desirably also be provided.

• Monitoring or follow up — to ensure the results are used. This will depend on the way the evaluation feeds into decision-making processes — budgeting, planning and management improvement, as discussed above in Section 1.

• Having adequate staff and training in evaluation. Evaluation requires some specialized skills, for example an understanding of statistical and survey techniques, and in this sense evaluation is not for ‘amateurs’. However, subject specialists and others having good analytical skills may contribute as part of the evaluation team.

OECD Country Illustrations

The following country examples illustrate a range of issues and approaches to evaluation. They do not aim to provide a total picture in relation to each country, but rather key, notable issues. The fact that they are mainly of the so-called ‘Anglo-Saxon’ tradition of management may reflect that evaluation and performance measurement generally are part of that tradition.

New Zealand

It has been said that new Zealand’s now well-known and radical public sector reforms which commenced in the late 1980s and were set out in the Public Finance Act 1988 have been much studied but less copied. The most notable reform to evaluation focused on output budgeting and performance contracting with chief executives. Budget allocations became appropriations for the purchase of agreed outputs from the ministry or agency concerned. In addition, the chief executive of each ministry or agency had a separate performance agreement with the relevant minister concerning the provision of these outputs. This focus was arguably more on accountability through contracting arrangements than performance management or improvement — although more recent evidence suggests that the contractual approach has been more in form than in substance, and arguably more an ex post evaluation than ex ante.

Outcomes did not feature directly in the budget process or in these contracting arrangements. This reflected the view that outcomes, being influenced by many external factors, could not be part of a contracting arrangement, which should instead focus on outputs controllable by the organization. Outcomes were seen as the responsibility of ministers. Policy development was clearly separated from service delivery. Ministers would arrange for whatever evaluation they considered necessary to determine what outputs might be required to achieve their desired outcomes. Perhaps it is not surprising that evaluation, left to ministers, withered—leading to significant loss in evaluation capacity and information on outcomes. Although mechanisms were developed to link outputs with higher level government objectives through ‘Key Results Indicators’ and ‘Strategic Results Agreements’ these do not seem to have worked satisfactorily.

In the past two to three years the need to revitalize program evaluation seems to have been recognized. While budgeting still occurs on the basis of outputs, chief executives are also required to manage for outcomes — ‘budgeting for outputs but managing for outcomes’ is the stated principle. Chief executives are required to work with ministers on determining linkages between outputs and outcomes, so as to achieve this.

On the other hand, considerable attention has been paid throughout the New Zealand reforms to developing good information on inputs, namely through the development of
accrual accounting (and budgeting) to produce information on full financial costs.

**Australia**

A formal (ex post) evaluation system linked with the budget was established in the late 1980s. Each portfolio (ministry or group of ministries) was required to develop an evaluation plan for programs to be evaluated on a cyclical basis, with evaluation implemented by them but overviewed by the finance ministry. An expenditure review committee of Cabinet was established to review all new policy proposals (which were required to provide objectives and performance measures) and to consider major evaluation reports. Evaluation reports were required to be published.

With a change of government in 1996, this approach was de-emphasized, and major aspects of the New Zealand approach to output-based budgeting (including the adoption of accrual accounting into the public sector to measure costs of inputs) were adopted. Formal evaluation requirements were abandoned and evaluation was further devolved to line ministries/agencies.

Coupled with a tendency of the government to seek a greater amount of policy advice (or evaluation related activity) from outside the civil service, this has weakened the central policy capacity and evaluation skills, previously focused in the finance ministry.

An outcomes and outputs framework was instituted in 1999, involving regular collection and reporting of performance information. There was also an emphasis on benchmarking performance indicators within the public sector and making comparisons with the private sector to determine whether activities might be contracted out to the private sector, and to use performance indicators for developing service delivery agreements between the ministry and the minister. This performance information is now largely internal to the ministry and its minister with little finance ministry involvement.

National Audit Office reviews have been critical of the quality of performance information, including the adequacy of accrual accounting information at the unit level, highlighting insufficient incentives for ministries to prepare it seriously, because central overview, for example through the finance ministry, is lacking.

Australia is now moving to put more emphasis on outcomes, compared with the previous focus on outputs, by developing outcome performance reporting and budget appropriations for outcomes.

The National Audit Office maintains a broad and well-resourced performance auditing role, which enables it to evaluate the economy, efficiency and effectiveness of government spending, although it is not within its mandate to question government policy objectives.

**United Kingdom**

The UK previously reflected an agnostic view of evaluation, typified by the concern expressed by a senior Treasury official at the time of the OECD’s 1997 study that ‘it might lead to demands for more expenditure’. Previously there had been regular but ad hoc expenditure reviews including fundamental expenditure reviews by Treasury, examining all aspects of a ministry’s activities.

With the development of (autonomous) executive agencies from the early 1990s came an emphasis on performance targets, particularly those related to efficiency and service delivery. However, concerns developed that there were too many indicators/targets and that they were not necessarily the right ones, not necessarily reflecting the organization’s real objectives and subject to ‘game playing’, with the emphasis being on achieving the targets irrespective of the costs.

While there is still emphasis on measurable targets for each ministry/agency through Public Service Agreements, mostly focused on inputs and processes, and through service delivery, there are now a reduced number of indicators and targets, and the focus is more on outcomes and outputs.

Moving beyond a narrow focus on performance indicators, there are now centrally–driven, two-yearly spending reviews, using extensive evaluation information. These focus on outcomes and priorities, and also examine efficiency and possible savings. Each ministry/agency prepares a submission on its operations which is reviewed by Treasury, Cabinet Office and a relevant Cabinet Committee, resulting in a Ministerial agreement and indicative spending allocation for the next three years; that is, medium-term funding is provided to achieve medium-term targets.

Thus evaluation/policy analysis issues are embedded in ministries/agencies, with Treasury and Cabinet Office playing a review role. In addition, there are major evaluations in key areas — education, health and cross-cutting issues such as the ‘new deal for young people’.
France

France has only recently moved to develop performance indicators for government programs and to develop more generally a results focus in the public sector. However, one longstanding and noteworthy institutional feature is a government body — the Scientific Council for Evaluation, which oversees major evaluations in the public sector. Evaluation is thus seen as requiring qualified professionals and as being carried out externally and independently.

Canada

In the 1990s, Canada developed many overlapping performance management initiatives — strategic planning, management by objectives, program reviews, etc., with considerable confusion about their inter-relationships. None of these appeared to have significant input into decision-making processes.

However, there is now a close linkage of performance information with the budget, with emphasis on reallocation through a Cabinet Expenditure Review Committee.

There is extensive annual performance reporting by line ministries/agencies. While the information appears to be little used outside the ministries/agencies themselves, they do use it for program management purposes. While there is an annual Government-wide performance report which aggregates much of this information, it does not appear to have any direct link with decision-making processes.

A formal evaluation policy now issued by the Treasury Board requires a 3-5 year cycle of program evaluations, to be carried out by all ministries.

In addition, the Office of the Auditor-General, as in Australia, has a well-established and broad mandate in performance auditing.

United States of America

The US has a long history of program evaluation through the work of the Government Accountability Office (GAO) which reviews many aspects of government programs, both at the request of Congress and at its own initiative. And within the Executive branch there is a long tradition of evaluation in some departments, well-illustrated by the systems development analysis of the Department of Defense, commenced by Secretary MacNamara in the early 1960s and continued even after the formal requirements of the Planning, Programming, Budgeting System (PPBS) which ceased in the 1970s.

More recently, the Government Performance and Results Act 1993 (GPRA) requires agencies to develop five-year strategic plans and performance indicators, over a medium-term period. While this path-breaking legislation is used by the Executive Branch in developing the President’s budget it does not appear to be used by Congress in appropriating the budget.

The Executive Branch through the Office of Management and Budget has also developed the Program Assessment Ratings Tool (PART). This is a formal program assessment tool covering efficiency, effectiveness and service quality issues, and is used by the Executive Branch in the budget process, thus reinforcing the GPRA. Agency scorecards are also used to compare the performance of different agencies, focusing on management issues such as service quality, systems and contracting out.

Sweden

Sweden illustrates a long-established evaluation culture, reflecting the consensus nature of its society and the willingness to review and discuss policies and programs to achieve this consensus. There are many evaluative bodies and extensive consultations on policies and programs through Commissions, which bring together all stakeholders and produce public reports to feed into political decisions. Evaluation is thus an accepted and understood part of political and managerial decision-making.

There is also extensive use of performance information in the budget process — annual results information supplemented by regular in-depth reviews.

But recently there has been discussion as to whether the information is more for display than decision-making. Budget decision-making appears to have focused on efficiency rather than effectiveness issues, with the latter being largely handled outside the budget process.

Section 3: Review of Five Country Presentations

The following summary of the five country presentations at this conference is intended to highlight key or unique features in each country, as a basis for further discussion. Other overall issues are raised in the earlier summary of the
first day’s discussion and are raised in the concluding session overview later in these proceedings.

**Brazil**

A major issue in Brazil is linking performance with both the national planning system (specifically the four-year multi-year plan) and the budget — and also with program management. There were major changes from 1999 onwards with the introduction of the new national planning system.

The government realizes that it has more work to do on its M&E system. The performance indicators are perceived as not yet satisfactory and, in particular, intermediate indicators are needed because of difficulties with impact indicators. The quality of evaluations needs to be raised and line ministries still lack capacity in M&E.

As in other countries, there are several levels of evaluation. The evaluation system is based on the principle of annual, results-based self-evaluation (but overseen by the Ministry of Planning) and there is also evaluation across programs by sectoral ministries and a centralized system of selective, external evaluation of large programs through an independent commission for M&E, overseen by the Ministry of Planning.

A key issue is decentralization — how to adequately monitor and evaluate across a federal system, where many national programs are delivered by sub-national governments — by the 27 states and 3000 municipalities, about which there is currently limited information.

M&E is supported by a central performance information system, SIGPLAN, maintained by the Ministry of Planning.

The system is perhaps over-ambitious and in any case appears not to have had a major impact on resource allocation as yet.

**Chile**

Chile’s M&E system is commonly acknowledged as well-functioning and containing many elements from which other countries can learn. The system is centralized through the central budget office of the Ministry of Finance. There are a variety of consistent tools which are well-integrated into decision-making processes, to achieve an overall performance focus in government management.

Monitoring is carried out through performance indicators covering effectiveness, efficiency, economy and service quality. These indicators also inform the budget process and are reported to Congress. They are also relevant to management improvement initiatives through comprehensive management reviews and a management improvement program.

Since 1997, there has been a formal system of (desk-based) external program evaluation. Evaluations are required to be public, independent, reliable, timely and efficient. Evaluation results are clearly used — to make budget adjustments or major program redesign, or program abolition — or to confirm program effectiveness.

Chile has been moving to more in-depth impact evaluations, using more sophisticated methodology. And the system of comprehensive management reviews examines the overall performance of individual organizations.

Overall there appears to be a demand for high-quality evaluation from Cabinet — and, apparently to a greater extent than in other Latin American countries, from Congress.

**Colombia**

Colombia has a long track record of working on an M&E system, which was initially developed between 1991 and 2002 as part of a strong public sector modernization process. The SINERGIA system has been in operation since 1994 and M&E developments have deepened since 2002, with strong top-level political support. A key driver of the reforms has been fiscal problems and the need to improve the quality of public expenditures, through improving resource allocation in the budget and the development of the national planning system.

The system has three components — results monitoring, strategic evaluations, and reporting, with a strong focus in the latter on accountability and social control through informing the citizens. Indeed the web-based system SIGOB, used for government management, provides citizens with real time access to this information.

Strategic evaluations are in-depth impact evaluations, with the Council of Ministers deciding which programs will be evaluated. However, the cost of these evaluations is high.

There has been a lack of coordination between planning, budgeting and evaluation actors, and more work is required to link evaluation results with the budget.

Given decentralization initiatives in Colombia, there is a desire to build an M&E system which links all levels
Towards the Institutionalization of Monitoring and Evaluation Systems in Latin America and the Caribbean

Mexico

Previously, Mexico had an extensive system of performance indicators, prepared in the Ministry of Finance and reported to Congress. This appears to have generated an extensive amount of paper, but the performance information does not appear to have been used.

The change from one party rule has seen demands for greater transparency and accountability, and for preventing the misuse of public funds for political purposes. In 2000, Congress required that all programs be externally evaluated each year. An independent National Evaluation Council oversees these evaluations, which are reported to Congress, the finance ministry and the line ministry, and become part of the next year’s budget dialogue. However, many of the evaluations are of poor quality, because of lack of evaluation capacity.

Apart from external evaluations, other components of the Mexican M&E system are selected in-depth impact evaluations, a results-based management system, and a new system of poverty measurement. The system of presidential goals established by President Fox may influence the choice of impact evaluations. There is still a need to convince program managers of the value of results-based management.

Evaluation in social programs is covered by the Ministry of Social Development’s SEDESOL system. This covers about 10 percent of the budget and evaluation has covered the umbrella program Oportunidades, based on the budget allocated to a range of social programs. Evaluation of this ‘mega program’ has been considered useful and has provided a demonstration effect for other areas of expenditure.

Peru

After a series of ‘false starts’ Peru has commenced a national M&E system (SSEGP), as part of wider efforts to promote results-based management, and this is focused on the social sectors and on investment expenditures. Beginning in the 1970s, Peru had a system of indicators and monitoring which did not work, because of lack of linkage of plans with the budget. Also, impact indicators that were developed were too broad — nobody owned them and therefore nobody acted on them.

There is a concern that expenditures have increased without corresponding improvement in results, and that M&E is needed to increase the quality of public expenditures. M&E is also seen as important for civil society oversight of the national budget; civil society participation should help ensure the system survives any change of administration.

Performance indicators have been developed at different levels of national government — for sector, organizational unit, activity, etc. Based on past experience, their approach can be described as ‘the fewer indicators the better’.

Peru is undergoing a decentralization process but the system does not yet extend to sub-national government, where a large part of government service delivery takes place.

Panelist: Harry Hatry

Director of the Public Management Program, The Urban Institute, Washington, D.C.

The work that has been done and is being planned in the five countries is very impressive. Let me give you some observations and suggestions. I see that you are planning a regional network for M&E, and I have some suggestions for this.

My first point is a definitional issue — the distinction between the terms evaluation and monitoring. It seems that the five countries are using it differently. I think there is some confusion and a need for clarification. There are several types of evaluations, as has been shown. There are impact evaluations, which usually have comparison groups. Then there are ex-post evaluations. Then there are management reviews. I think that some of the Chilean examples of small, short studies are probably management reviews. There are process evaluations, and case studies. Which type of evaluation are you talking about?

Usually, evaluations are thought of as being in-depth studies. You can only do a few of them and they are often much more expensive, as is the case with impact evaluations. So a program can only be evaluated every several years. Usually evaluations require special skills and are most often done externally, as we have heard.

Monitoring — and for the term, monitoring, one could also use the term performance measurement — is much more closely related to the terms managing for results or managing by results. Usually here we are talking about the regular tracking of performance indicators. This regularity is very important to be useful for managers and for managing.
It has to be done frequently (certainly at least quarterly, if not monthly) depending upon the indicators, and with timely reporting. Internal people do most of it, although they should often benefit greatly from external consultants to help them design the process.

The second point is that the devil is in the details. It is easy to talk about these concepts and principles. Where you really get into trouble is in the implementation. Performance indicators are the first example of that. Throughout the discussions we have heard very little about the specific indicators and data collection procedures being used. It is very difficult to determine the quality of these activities and extremely important that they be done properly.

On indicators, I want to point out that by this time, the year 2005, almost every program imaginable has been subject to some sort of evaluation or measurement and the websites are full of this. So my first suggestion for the regional network is that you think of putting together a series of common, typical, good-quality indicators of outcomes, particularly ones that are already available so that everyone doesn’t have to start from scratch.

The distinction between outputs and outcomes has been a big problem in every country — every foreign country, my own country, every city, county, state. Making this distinction is a problem; it is not always clear which is which, and it is nice to have the definitions of inputs, outputs, impacts clarified. Again, I think the regional network can help in this. It doesn’t mean everybody has to follow the same approach, but at least you would have some general guidance.

My third point is to highlight something that has been raised periodically in this conference, but not emphasized enough: training. How do you do training, who gets trained, and how much training is provided? This is extremely important for evaluation. You all recognize this and it is obvious. And I don’t mean training just for budget staff, but also emphasizing training of sector staff. I think that has been neglected by the donors, for the most part. They haven’t done enough work with the sectoral ministries. I think your countries are much more exposed to this issue, and you realize that you have to do it. It is also extremely important to get members of Congress and their staff involved. I do not believe that Congress is not interested in results and outcomes. And if I look at the United States and Ukraine and the work we in the Urban Institute are doing there, the reason they are not doing it is that they don’t get good information which they can understand. I don’t blame them for not using such information, but they are an important customer and they can also be given help, encouragement and training.

How should the training be done? You can use external trainers including international consultants but they should be primarily used for training the trainers. Internally it should be a responsibility of the government. My suggestion is that for the training of the trainers at each ministry, have perhaps at least two people who receive that training, and then are responsible for going back into their own ministries, departments and programs to train people within them. I have seen that work in one of our states in the United States and I think it is, in the long run, a fairly effective approach.

For your regional network, another suggestion is that not everyone should have to develop training materials and curriculum. Rather, you might collectively come up with a set of written training materials. There is a lot of training material available around the world now on performance measurement and evaluation, not always available in Spanish, but it can be pulled together. So putting together both common training and written materials and curriculum would be very useful. Obviously, it will have to be adapted to each country but there is no sense everybody doing it themselves, particularly those other countries which have not yet made significant moves to develop M&E.

My fourth suggestion is that as part of an ongoing performance measurement and monitoring process there should be explicitly written into the process explanations where the targets are not met or even where they are exceeded by a certain percentage. Perhaps you might require that the program, the agency or the ministry explain variances from the performance target of say, plus or minus ten percent. For example in two of our states, Texas and Louisiana, the legislature has required this. A variance report has to be provided by the program or the agency if the actual value shown in the quarterly reports differs by more than five percentage points from the target. The explanations can be presented in many different ways. They don’t necessarily have to be terribly sophisticated; they can be informal, qualitative explanations. You can use focus groups with citizens, and you can also use these explanations as a basis for subsequently developing the agenda for in-depth evaluations.

The fifth suggestion is a technical one. While many countries and their agencies which are doing evaluation
realize this, it is surprising how often it is not practiced. The data on outcomes or impacts should be disaggregated in different ways. Obviously people are interested in data disaggregated by geographical area — the district, the province, the state, the city, the location. They should be broken out by demographic characteristics — age group, gender, income group, race and ethnicity. Certainly human services staff are aware of this issue. We talked yesterday about maternal mortality and a disaggregation by the causes or perceived reasons for those deaths. We’ve heard about the reasons for traffic accidents. If you’re talking about pollution work in the environmental department, pollution sources need to be known. We should disaggregate outcomes information by those different groups. The information then becomes much more useful to everyone.

My sixth comment or suggestion concerns incentives. Some of you point out the lack of incentives for M&E, and it is a key issue. On the one hand, the research that we have done at the Urban Institute does not make me very sympathetic towards performance-related pay for individuals. While it is a very simple idea it seems to me it is more trouble than it is worth and almost counterproductive. If you have to do something like that, use it for groups, but I’m not a proponent of pay for performance.

I think there are other types of incentives that are more practical and more useful. One is simple recognition awards. We don’t do enough of that. Agencies and programs that have achieved their targets or exceeded their targets annually should be explicitly recognized; this is very powerful.

One of the lessons from the United States and I do not know how applicable it is to other countries, is to relax some of your restrictions on use of money. For example, the size of contracting that you can do or the hiring policy. You give agencies the flexibility to make trade-offs if they succeed in achieving outcomes.

Perhaps the one I like best, however, is a very simple idea. It is that a key motivator in government is that the upper level managers, the upper level leaders, explicitly express interest in outcome and results type information.

One of the emerging ways to provide incentives or motivate staff is what we call ‘how are we doing’ sessions. Shortly after the (let’s say quarterly) performance report comes out, the manager at whatever level meets with his or her staff and discusses where they have done well, where they have not done well, why they aren’t doing well, what can they do to improve — are there some actions they should take. Then in later meetings they go back and see if it worked or not and whether further change is needed. In the City of Baltimore, Mayor O’Malley has been meeting regularly with each of his department heads. He goes over all the statistics — inputs, outputs, outcomes and whatever other information is available, and he discusses with them in a constructive way what they are doing, to try to lead to improvements. This is very impressive.

My seventh point is that we need more emphasis on sub-national governments. There is no question that management for results, performance-based budgeting, etc., are equally important at those levels of government. There is no reason for local levels of government not to introduce performance measurement. Actually it is somewhat easier at that level. However I don’t think that donors have done enough to help lower levels of government, nor do I think countries themselves have.

The term ‘performance partnerships’ — between levels of government and between programs — has occasionally been used for this type of thing. Some of you have noted the problem of interactions of goals. You have classic examples of improving the economy versus hurting the environment, or vice versa, improving the environment and hurting business. With such interacting issues there should be performance partnerships as a very useful mechanism to get people together.

My last point is really a tactical suggestion. I suggest dropping the term ‘evaluation’ — at least for your line agencies. Keep it if you want for central use but the term ‘evaluation’ is threatening to most people. I think it would be more useful to talk to your management not so much about that word, evaluation, but rather about ‘improved program performance’ or ‘improved outcome management’.

Panelist: Philip Joyce
Professor of Public Policy and Public Administration,
George Washington University

My comments reflect my experiences in the U.S., which are not only at the federal level, but also recently in a project looking at the 50 state governments. I would echo other comments about the importance of looking at sub-national governments. In addition I think that at the present time there are many more interesting things going on at the state and local levels in M&E than at the federal level.
It is probably obvious to everyone here, but it is hard to institutionalize M&E. So it is important to consider some implementation issues. In particular I want to focus on making the connection between successful performance measurement systems and the allocation of resources because I think that is perhaps the most difficult thing to do.

My first point is that it is important not to confuse having a lot of performance measures with making progress. Sometimes people think the more measures they have the better they must be doing. There are as many systems that have suffered from having too many measures as from not having enough. Particularly in countries that may have too many measures, there tends to be proliferation of output measures and not enough measures of outcomes. Their outputs are not necessarily linked to the outcomes, even when they have outcome measures. That is, their indicators don’t tie in with the goals of the programs.

I think it is also very important to spend some time on the logical connection between the output and outcome measures. In my work last year in Mexico, I noted that both the Health Ministry and the Social Development Ministry, SEDESOL, had done a good job of thinking about the relationship between the measures. That was not true for all of the other ministries.

Having worked in ministries of finance, I have some sympathy with Guillermo Bernal’s point of view that money not spent is money well spent. If you are in a ministry of finance and looking at performance measurement systems, then if you see indicators which are completely under the control of the line ministry or even worse, where success is defined as spending more money, you ought to be very concerned.

Secondly, I want to say a couple of things about the linkage between results and funding. One is the fallacy, in practice if not in theory in a lot of places, that planning and budgeting are separate processes. This causes two problems — that sometimes planning is done outside resource constraints and that resource decisions are made without thinking about planned goals. We have to think about the plans and the resources in tandem.

Another and perhaps equally important fallacy is to think that the linkage between results and funding is somehow mechanistic. I dislike the term ‘performance-based budgeting’, even though it is widely used, because it encourages the notion that if you get the performance measures right you will know how much funding somebody should receive. This is almost certainly not true. An even worse formulation of this is where people say ‘If ministries meet or exceed their targets we will give them more money, and if they don’t, we will take money away from them’. My question would be, how do you know that? If they didn’t meet their targets, perhaps this would signal that they need more money.

So I very much prefer the term ‘performance-informed budgeting’, because that is the reality. You have the performance information available when the budgeting decisions are made, without implying some mechanistic relationship or link between the two.

My third point is the importance of thinking about performance-informed budgeting comprehensively. Too often there is a focus on the legislature or congress and on the finance ministry, on the assumption that these are the people who budget and who allocate resources. In fact, the budget process has several different stages — budget preparation, budget approval, budget execution, audit and evaluation. A lot of what is done in the line ministries could be called budgeting, because as the Mexico presentation pointed out, budget execution is also about allocating resources. So we should think comprehensively about resource allocation and budgeting, as involving not just the initial allocation of resources, but also the management and ultimately the audit and evaluation of the use of those resources.

Fourthly, I want to reinforce the importance of incentives; that actors at any stage of the process will only reform if they have incentives to do so. This means that if the current way of doing things meets their needs they will resist reform. So if elected officials think the most important thing is being re-elected and they are being re-elected at a very high rate, a challenge to the status quo will not necessarily be supported. So you have to think about what people want that they don’t have now. Elected officials may be concerned about inadequate responsiveness to constituents. The finance ministry may think that it is being asked to make decisions on results and on funding without enough information. The line ministries may think there are too many central controls and may want some relaxation of those controls.

Finally, I want to mention some important implementation issues, and the first involves leadership. Virtually all successful reforms of this type have required sustained and consistent leadership not only in the line ministries, but also in the centre of government.
carry out reforms and nobody in the centre wants to, there are limits to how far they can go. The second implementation issue is that the system has to be designed so that participants get the information they need. It did not surprise me at all that there was initially no questioning from the Mexican Congress because if you inundate such an organization with performance measures they are not going to look at them all. You have to think about providing a limited number of very high level measures. That doesn’t mean they are the only measures available in the government — other measures might be used at the ministry level although they should be logically connected to the higher-level measures. Or to put this another way everybody needs to have the performance measures they need.

I would add that implementation can occur either from the top down or the bottom up, and there are advantages and disadvantages to both. Top down gives more control to the ministry of finance, but a bottom up approach is more likely to get line ministry buy-in. But regardless of whether you try to do it top down or bottom up, I would encourage any country that is moving down this path not to try to do everything at once.

The most important piece of advice I can provide in terms of implementation would be that you start with the 'low hanging fruit'; that is with the organizations that are already well on their way, because they can be used as examples for other ministries. The bottom line is that these reforms take a long time and you should avoid promising too much too soon. Otherwise, invariably three or four years down the road, someone will brand your reform as a failure — when it is not a failure, and simply hasn’t had enough time to succeed.

Panelist: Inder Ruprah
Principal Economist, the Inter-American Development Bank

Essentially we are discussing whether evidence-based policy and evidence-based government is replacing other factors that influence policy-making and policy recommendations — the nebulous world of pressure groups and the pragmatics of day-to-day decision-making. So the first interpretation of this conference is that M&E systems are gradually replacing GOBSTTTT, which is a technical term for ‘good old boys sitting together talking turkey’.

It is very clear that there is no one unique institutional system to handle all the different tradeoffs — internal, external, etc. and each country is struggling with its own value systems and has its own history to resolve those tradeoffs. The search for the perfect model is a deceptive one.

As others have said in this conference the main issue is not techniques. The issue is improving accessibility of the evidence, increasing ownership of the evidence and getting appropriate buy-in from actors, leading to subsequent actions.

Communication of evaluation is one of the biggest tasks. Last week I was in Paris at a DAC/OECD meeting about setting standards of evaluation. I made a presentation in which I tried to demonstrate that before/after photographs or naive outcome evaluations are terribly misleading. So if the issue is development effectiveness of a program, impact evaluation is the most appropriate technique, be it embedded in process evaluation or institutional analysis. My presentation was a complete failure. The only thing that people could remember was the photographs. There were these wonderful photographs of before (a three-legged dog, and a little child with no shoes, and a broken down building) and after (the boy with shoes, the four-legged dog, and a four-garage house).

So what are the ‘three elephants in the room’ that we need to put on the table. One elephant is that evaluation is intrinsically a political issue. The decisions on what to spend, how to spend it, and on whom to spend it, are all difficult to make. Political and personal actors are involved. So the evolution of an M&E system and how to communicate its findings, whether internally or externally, are all subject to political constraints. Thus institutionalization of the monitoring and evaluation framework is essentially a question of how to set and enforce a structure of incentives and disincentives.

Another elephant is each country’s system of national statistics. I am surprised by the disconnect between evaluation systems and countries’ statistical institutes. We know from our evaluation work that it is relatively inexpensive to insert a set of additional questions into a survey which is already planned; this information would then be used in rigorous impact evaluations, for example. Yet because of the disconnect this potential is usually not exploited. As I said, the problem is not techniques; the problem is data.
The third elephant, or elephants, are the multilaterals. We are not donors. We are essentially lenders. So this is a real political problem. A country borrowed, an institution lent, the debt service is in the future, an evaluator comes and says, ‘Look, you guys didn’t achieve anything with this program.’ So that aspect is important. A second aspect is whether all programs have to be evaluated in each country, given the extent to which multilaterals share ideas. Let’s take Oportunidades. Is it necessary to go to each country, each year, to evaluate conditional cash transfer programs? I would say no, but the politicians would say yes. They want to know if the conditional cash transfers are working in their municipality.

I will end with the task of what we are doing with our ex post evaluations. We cannot participate in the conception of a project. We can’t even be doctors. We are not allowed to look at projects that are alive. Instead, we are a type of mortician. Someone in my office said that is an incorrect metaphor: we are archaeologists, because we actually have to look at projects that died four years ago.

There is an enormous ‘gravy train’ coming, mainly from the multilaterals organizations, and I think there is an initiative in the United States to force the multilaterals to have a certain percentage of their programs evaluated. That is an enormous amount of money that is going to be pouring into evaluation in Latin America. Here I think there is the issue of training and technical competence. We are finding that in most countries, in the universities and NGOs, there is an enormous amount of highly qualified technical capacity.

In the IADB’s Office of Evaluation and Oversight we write terms of reference for our evaluations and we outsource our evaluations. We have created a system of registration of evaluators called EVALNET. We try to avoid relying on a very few people in a few universities in the United States and England, etc, to do evaluations. As I said, we find very good quality proposals and high evaluation competence within the countries themselves.

It is very important that this large new funding for evaluation in the region be used to enhance local capacity — at the ministry level and at the sub-national level.

Notes

Country Reactions:
Reflections of the Five Country Presenters and Views of Other Participants

Sonia Ospina (Conference facilitator)

We now want to focus on the future. We have heard a day and a half of important and relevant information and some great stories, and some real challenges have been raised. It is time now to start thinking about what we do about these issues. I now ask the five country presenters to give us their reflections on the challenges that they most want to address as they move forward.

Marcela Guzman (Chile)

First of all, there has been a lot of discussion on the complex issue of performance indicators. It is very important to keep in mind that there are different levels for the use of this information and hence a need for indicators at each level. I do not believe there is an optimum or ideal number of indicators. The issue has more to do with the circumstances of each country and with the different hierarchical levels at which different information is used.

My second point has to do with institutionalizing the M&E systems in each country. Obviously it depends on how a system is designed and implemented and this is influenced by the institutional capacity of the country. We have to consider not only how those systems can be developed but also how they can evolve and become stronger. We need to determine the basic requirements for a country to be able to implement these systems and processes. Countries themselves have to consider the resources they need for this task.

My third point goes back to the issue of indicators. Sometimes there are measurements which do not have any direct relation to outputs or other results, and our experience suggests that determining this relationship is not easy. It is not that we don’t have to be concerned about the impact of a program; rather that we are not always able under normal circumstances to have an impact evaluation, even if it is desirable. A program manager should have information on what goods and services are being provided and on the impact of those goods and services.

Another point is the problem of inertia in government programs. In our experience the development of M&E systems frequently contributes to removing that inertia. Similarly, human resources are one of the most important assets that a government has but the allocation of human resources can be very inflexible, with the result that they are used in stagnant programs. So evaluation should focus on this issue too.

Ariel Pares (Brazil)

The issue of the number of performance indicators has been discussed a lot. This is a very central issue. I think we need to focus more on the quality of the indicators and the data, rather than the number of indicators.

A second point is that two types of evaluation have been presented here — not as opposites but as two types. One is self-evaluation by managers or operational staff, and the other is external evaluation. Both could be important and it is important to have linkages between them. I was a bit surprised about the amount of discussion here on self-evaluation. In Brazil many people said that self-evaluation would not work. But I think it is promising, while recognizing that it can be improved. Both self-evaluation and external evaluation are needed, because they have different objectives.
There is a need for comprehensive evaluations, which may be based on a requirement of, say, the President. This will create an increased consciousness and understanding of the need for evaluation. But I don’t see external evaluations as necessarily the solution.

A third point is that certain evaluation initiatives have their origins in fiscal problems, which would be the concern of the Minister of Finance. The Minister of Planning is in charge of our M&E system. It was a concern with the quality and costs of programs, and their impact, which led to the development of evaluation in our case.

Another point, which was highlighted in the Mexico presentation, is who is requiring the evaluation — presumably someone who needs better information for better decision-making. However in our case, and I see that this is also the case in other countries, the Congress is very much absent from this type of decision.

Finally, we have not explored much the issue of citizen participation. But everybody thinks this is important. The issue is how to do it, and this is a complex issue which depends on local conditions. I believe that citizen participation and evaluation should go hand in hand.

Manuel Fernando Castro (Colombia)

I have five points which are on the agenda for monitoring and evaluation in Colombia. They also appear to be important for other countries.

First of all, there is great concern about the level of government expenditure and a view that M&E systems should be an additional or supplementary tool for fiscal discipline. As we heard earlier the development of M&E systems in Europe and the OECD countries generally took place in the 1970s in response to two factors: firstly, the need for fiscal restraint reflecting the difficult fiscal situation in those countries; and secondly, a general dissatisfaction with public services in those countries.

In Latin America in 2003, only two countries had a fiscal surplus, which means the other countries have fiscal problems. On the other hand there is general dissatisfaction with public services in those countries.

In Latin America in 2003, only two countries had a fiscal surplus, which means the other countries have fiscal problems. On the other hand there is general dissatisfaction with public services in those countries. If results-based budgeting is presented as a tool for better budget decisions, there is a greater possibility of achieving a good linkage with the budget.

As has already been pointed out during this conference, M&E is not a technical issue alone. While the technical side is important, I believe the main task ahead of us is to strengthen the political muscle of the M&E systems and their support from higher political levels — the President and the Congress. M&E must provide the information that is needed and wanted at that level.

The Colombian experience well illustrates this point. From its beginning, SINERGIA has been a well-designed technical system which works well inside the planning and budget departments. But in the past this technical information was not fed into political decisions. We realized we had to find alternative arrangements to link the technical capability of these institutions with the political muscle.

A third point is that M&E systems cannot be isolated from the need to improve information systems and public information in general. M&E is a tool to generate quality information. In France nowadays, 90 percent of public information comes from administrative data such as institutional records or registries within government. Only ten percent comes from other sources. In Latin America, the opposite is true. Of course, this would imply the need to strengthen information systems in LAC countries.

The fourth point is transparency and accountability. M&E systems cannot only be for governments; they are a public asset and must strengthen and sustain accountability mechanisms. Civil society and the private sector should be involved as users of the information.

Finally, there is the need to train more people as evaluators, and with better evaluation skills — not only in academia but in other institutions. This will help us to reduce the costs of evaluation.

Carlos Ricse (Peru)

I would like to highlight a number of key issues. One is that we should understand that the importance of public expenditure issues is not only related to the size of the budget deficit, but also to the quality of public services. So you can have a zero deficit, but at the same time still have an expenditure problem.
Another is to consider carefully who is the end user of M&E information? Is it the Congress, the President, or civil society? It seems as if there is a ‘one size fits all’ approach. But a Congressman will never read a 1,000 page report. So we should keep in mind who the user is. The political sustainability of an M&E system will depend on the population as a whole being aware whether public money has been well spent or badly spent.

I would like to add four specific points. We are concerned about the development of skills to implement performance or results-oriented management. I am not only talking about the information process — the technical side, but also the outcome of an M&E system — utilization of the M&E information. Here we have a large gap to fill.

Secondly, a constant concern for the Ministry of the Economy in Peru is that the system is complex and that in implementation and development we have to proceed carefully, taking small steps, and not rushing things.

A third point has to do with the development of skills and the sustainability of those skills. We have to find institutions that make it possible to sustain this type of development. The universities of Peru are very important here.

Finally, there is the quality of information in M&E systems. We need to eliminate duplication in information requirements by the national government. The sectoral ministries continue to complain that they are always being asked for the same information by different actors.

_**Gonzalo Hernandez (Mexico)**_

It is good to know that we in Mexico are not alone in this M&E process, although this is not the first time that we have heard about other countries’ experiences. It is always good to know that we can support each other even if countries are facing different problems. However we talk about providing mutual assistance but we never do anything about it. For example, as a concrete proposition we would like to know more about the relationship between evaluation and the budget in Chile.

We evaluate for many purposes — to improve programs, to provide better support for political decisions, to improve accountability and, of course, to feed the political process. But evaluation is only one of the tools available to do this.

An issue raised earlier was including performance indicators in the budget. To do this we need good data collection and if we disaggregate this information by groups it will become more complicated and expensive. For example, we are asked by our states and cities about the impact of the habitat program in their state or city. However, the national budget is not broken down by states or municipalities so it would be costly to obtain that information. Developing a better link between evaluation and the office of statistics may assist in reducing this cost.

We also have to expand the amount of external evaluation which is conducted, perhaps by using the services of international consultants. We have worked closely with experienced experts and with consultants, but are now relying more on recent graduates — I believe this new group of qualified people will be more cost-effective.

Finally, let us remember the point made by Manuel Fernando Castro: the political dimension of an M&E system is probably the most important. This dimension is different in each of the five countries highlighted at this conference, but there are still important lessons for us all to learn from each other.

**Guarocuya Felix Hernandez (Director of the National Planning and Budget Office, Dominican Republic)**

Even though our population is much smaller, our problems are very similar to other countries at this conference.

An important issue that has come out of this conference is that M&E is not an end in itself. It is only a means to an end. It is only a tool to, for example, improve the quality of public expenditures.

I would like to refer to the territorial element — other levels of government, which has been touched on in the case of Brazil. This has been missing from the debate at this conference; how can we improve M&E at other levels of government?

**Maria Castro (Assistant Secretary of Planning, Guatemala)**

In Guatemala, we have been studying how to move towards an evaluation system, recognizing that in the remaining two and a half years of our current administration it will not be possible to implement such a system. But we want to plant the seed. The experience of Latin American countries suggests this is an enormous challenge, not only technically but also because of the lack of human capital and because
many of our governments are not interested in transparency or improving results. So they would be even less interested in implementing evaluation systems. I think we now have an opportunity in Latin America to move forward in countries where democracy is not just a question of voting, but also a question of governance. But this requires continuing work from the time a President takes office.

My thinking has changed as to what such a system requires and we have made some conceptual errors. Thus I am very pleased to have this opportunity to learn from other countries. We have not, for instance, been able to approach Mexico to learn of their experiences in SEDESOL. We have learned more from Chile perhaps because many of us who work in this area have studied there. We hope that a regional M&E network will be established so that we can learn from each other’s experiences.

From the Audience (Canada)
I would like to make a brief comment about our experience in Canada. I would suggest approaching the question of performance monitoring and evaluation using a systematic risk management approach — where desired results and acceptable risk are considered together.

The second comment relates to the devil being in the detail. Many of you stress the importance of capacity building: on developing the people to implement this system. But you have to be quite clear at the beginning of your process about what you really want to achieve, and where your focus will lie. This is where the detail comes in. For example, should the focus be on measuring impact, or outcomes, or outputs? There are also definitional issues to be considered; these are also addressed in the evaluation glossary of the OECD.

From the Audience (Honduras)
From the five country presentations, I see that many of the specific problems raised are also those we have in Honduras, especially in the area of training and institutionalization. It is not clear who should be responsible for monitoring and evaluation, as we do not have a plan setting out the issues we wish to evaluate.

Much more important than technical capacity is the political will to launch M&E processes. In Honduras there has been this political will. Civil society has been demanding concrete results from programs. Honduras is classified as a highly-indebted poor country, and we are implementing a poverty reduction strategy. So we need an M&E system that evaluates the effectiveness of the strategy.

In the implementation process which began last year we have found numerous problems. One is the weakness of administrative records and the need to improve the information coming from these. Much of the information that we use now for monitoring comes from very costly surveys.

We hope that the M&E network, which is to be created will enable us to continue to exchange experiences.

From the Audience (Uruguay)
In Uruguay, since 1966, the Constitution has required that the budget be a program budget. We made progress, but we haven’t always been consistent and in recent years, fiscal crises have meant that fiscal restrictions predominated. Now we have a new budget and a new government so this is a good moment to go back to our previous work. We need to have capacity to evaluate as well as to formulate programs. We need to clearly establish who evaluates and whether the evaluation is internal or external. And there is the question of how sophisticated our performance indicators need to be.

From the Audience (Mexico)
One of the comments made by my compatriot has to do with what would be the impact on M&E systems of a fiscal crisis. Uruguay has recently undergone such a crisis. One issue I would like to pose is that if we design an M&E system, how might it be threatened by dramatic changes in the economic circumstances of a country? This question must remain unanswered for now, but it is important to any future discussion of M&E.
Closing Remarks

Letitia Obeng  
*Director for Strategy and Operations in the Latin America and Caribbean Region, World Bank*

I will begin with a very quick description of a situation many years ago, when I was working with some village communities on their water supply and sanitation system. We were trying to get them to monitor their own performance in managing their water supply systems. We talked to them about choosing useful indicators that they should use in monitoring their performance in an ongoing manner, the importance of assessing the impact of the water supply systems, and of drawing lessons about how to improve the management of their systems.

What we are discussing in this conference is much more complex, but the same lessons arise from the simple project I mentioned above: the importance of picking the right kind of indicators; of real ownership of M&E systems that transcend governments and continue in a sustained manner even though governments may come and go; and that there needs to be accountability for the implementation of these M&E systems.

You have been discussing the institutionalization of M&E systems in a decentralized context, and looking at linkages with the budget, capacity building, and improving service delivery, and you have also discussed the costs and timing of evaluation — when to do it, and how much to do. You are all searching for the right solutions for your situation — not a one-size-fits-all solution, and drawing lessons from each other. It is clear that this is a subject that is truly important to us in this region and that there is a substantive increase in the demand for reliable and well-functioning M&E systems.

Let me assure you that we in the World Bank are very much interested in supporting and working with you, including on the regional M&E network as are other partners represented here, including the Inter-American Development Bank. This regional network will be an excellent opportunity to move ahead on some of the many important issues that have arisen in this conference.
PART FOUR:

Launching an M&E Network for the Region
We want to share with you a few elements that the World Bank and the IADB have been discussing in terms of a regional M&E network.

The first objective is to establish a country-owned space to facilitate interaction between countries and to provide opportunities for mutual learning in monitoring and evaluation, both technical and political issues — covering the whole chain from the production of data, development of indicators, the use of different M&E tools and instruments and finally linkage with the budget. So it covers the whole chain of the use of M&E systems within sector ministries — their own planning and management, and also the use of this information in national budgetary and policy decisions. As we have seen, progress differs between countries.

The second objective is to contribute to the sustainability, strengthening and institutionalization of country M&E systems. By institutionalization we mean systems that will perpetuate as state policies instead of just administration policies so that they continue, even when the political administration of a country changes. The main goal of supporting the implementation of these systems is to contribute to the improvement of public sector performance, to improve the quality of public expenditure and ultimately, to reduce poverty and increase welfare.

In terms of more specific objectives, we need to improve the flow of information on what is happening with M&E systems in the different countries. We need to facilitate learning about best practices because we know that there are many different experiences, but also to learn from failures and how to overcome political and technical challenges in the different countries, and to increase country capacity. Countries need to know where and how to find help. We have seen in this conference many interactions between different country teams, providing information for the first time on who is doing what and where help can be found.

One of the key principles of a regional network is openness and flexibility, which attracts decision-makers and experts inside and outside the government. The people who are currently in the government could be outside the government a few years from now and their expertise could then be drawn on. Those who are currently experts working for a consulting firm or think tank, as well as those conducting academic research may be implementing M&E systems within the government in the future. Thus, it is critical to foster — through the creation of an open forum — the interaction among those who in one way or another and from different positions are able to form a learning community.

We envisage a network which includes senior officials from sectoral, finance and planning ministries; members of Congress and their advisors; academics, consultants, and experts working on these topics; and, finally, multilateral organizations. So it would be very flexible in terms of membership.

It is important that we hear your opinions about the activities that could be undertaken. One is follow-up conferences, similar to this one. The organization of study tours is happening already — for example, Mexico’s Oportunidades program has been visited by teams from several LAC
countries and from outside the region. The World Bank and IADB have been promoting this also and we should find ways of facilitating these tours.

A dedicated website is another possibility. With the availability of Google this might not be critical, but it could be a good instrument to facilitate access to resource materials. There is a need for more materials in Spanish.

Formal in-country training has also been mentioned as a need. Another possibility is training through interactive distance learning techniques, such as the Global Development Learning Network of the World Bank and similar facilities at the IADB, in which people in three or four countries can be connected simultaneously. This technology can provide virtual conferences or seminars at a much lower cost than this conference.

It is critical that this network should facilitate regional activities and that synergies or learning from each other are exploited.

Finally, this network should complement and leverage the work that World Bank and IADB teams are doing within the countries. The two Banks can facilitate country-level diagnoses and needs assessment, be information brokers, facilitate conferences or study tours or lending activities, and take the lead in specific activities, such as developing a website.

However, country ownership is crucial — the initiative has to be demand-driven, not supply-driven and it has to be part of a collective agenda. Moreover, it must be aligned with the needs; demands and challenges countries are facing to improve the quality of their public policies.

**Roberto Garcia Lopez**

*Chief Advisor in the Inter-American Development Bank’s Development Effectiveness and Strategic Planning Department*

After I left university in the 1970s we talked about administering development or managing development. We didn’t talk about results-based management. We also talked about planning, and I worked for about ten years in the planning institute. But the plans were never implemented.

So I had the idea that we needed to evaluate projects and programs. Then another ten years went by in which I designed programs and projects and evaluated them. But I was told that was not the problem; the problem is how we manage budgets. So I went to the budget area and became the budget director in Argentina.

When I was in planning, we thought that the people who worked in the budget area were just accountants and the only thing they knew how to do was to add and subtract. When I was in the budget area, they said that the people in planning are really crazy and they aren’t able to operationalize their plans.

Concerning our joint idea of launching a regional M&E network, at the IADB we have experience over the last five years in promoting dialogue. We have a program of regional dialogues, covering seven topic areas — including education, environment, central banks, and now transparency and public management. During its first years, this focused on the civil service and social services. Around March of this year they decided that these topics were almost exhausted and that they should start working on effectiveness of expenditures and results-based management. This is a formal network in which the vice-ministers of planning or budgeting participate. The budget director of Chile, was selected as chairman. The participating countries have defined an agenda for 2005-2006, which will centre on four major topics, one of which is the link between planning and budgeting. Another is performance indicators in the budgetary process, following the lines of OECD countries’ experience. Other topics raised have been the inflexibility or rigidity of budgets and off-budget activities. We have a forum where they can discuss their common problems, exchange experiences, and do it in an open manner.

This is a closed network, because only the vice-ministers participate. Organizations or academia participate only by invitation if a presentation or paper is needed. After four or five years of this network we have come to the conclusion that a more open network with a broader participation is desirable.

So we agree that this new regional network initiative can be very useful, but there must be ownership by the countries.

I have four brief comments to conclude. In terms of membership we should also think about including civil society and its organizations. In terms of a website and training materials we should consider whether this informal dialogue should be maintained only among those of us already in the network or whether we should have a dissemination activity
in order to bring others in so as to enlarge the dialogue. The network should discuss concrete ways of conducting results-based management and making it accessible to the citizens, an issue that we have not really addressed in the discussions and presentations at this conference. Finally, we should note that countries are seeking a change in the financing provided by multilateral agencies, placing greater emphasis on budget rather than project financing, so we need to look at ways of strengthening budgetary institutions.

**Floor Discussion on a Regional M&E Network (Editors’ synthesis)**

Ariel Pares, representing the Brazilian planning ministry, agreed that it is very important to create a network and to sustain it. It is more difficult to sustain it, and this requires contributions by all members of a network. He stressed the need to integrate the budgeting and planning functions of government because each has a different perspective, as illustrated by the Brazilian experience.

Marcela Guzman, representing Chile’s finance ministry, posed several key questions for the conference organizers and also for those proposing the creation of the network. First of all, is the need to maintain the network. A key issue here is the membership of the network: What would be the coverage of the network? Who is going to participate? What ministries or representatives of the public and private sectors and perhaps the Congress?

A second point is the network’s modus operandi; this has to do with maintaining the network and ensuring its sustainability and effectiveness. To be judged as effective the network would need to contribute to the work of the countries. And there would need to be clear responsibilities on the part of the IADB and World Bank, as well as clear roles and functions of each one of the countries in order to work together with the other actors.

She argued that the network’s agenda would go hand in hand with its modus operandi. There is a need to prioritize topics from the many issues raised at this conference, and to set up an agenda to address them.

Another participant agreed that these points and questions are very relevant. He also advocated an informal approach for the network. A more formal approach would be possible after a few years, if the network is successful. Membership of the network should be kept open, to see who really wants to participate and make a substantive contribution.

He argued that a network cannot handle more than two topic areas per year, and that these should be based on the list of topics presented at the conference, and also based on the interest of countries. This agenda could be developed by a small committee where not only the two Banks are represented, but perhaps four or five other people, not so much representatives per country, but people from different countries who will be committed to help the network in those initial topics.

The network could adopt an approach similar to that of the OECD, which conducts surveys of the progress made by member countries in particular areas, such as the institutionalization of M&E. Country responses on their progress and priority needs could then be submitted to a technical committee of the network and in turn this committee could do a country analysis. This would allow individual topics to be discussed in depth, which could lead to an agenda being developed, addressing the different needs, and identifying the different actors contributing to this agenda.

Hernando Gonzalez, representing Mexico’s SEDESOL, noted that his country has had some technical exchanges with other countries in the region, on issues such as Oportunidades. People have come to Mexico to see how this program works and to find out about the impact evaluations carried out on it, and the World Bank’s useful seminar on impact evaluation which took place in Mexico City last year, and in which his Colombian counterparts also participated and found valuable.

This illustrates that there are clear demands, where people will benefit from interchange in specific areas. However there are many areas and topics to be covered, and since individual countries are moving at different speeds, some would need more of technical support or assistance from Chile or Colombia. And perhaps other countries do not need to emphasize the technical aspects. So the network would need to be flexible because not all the countries are going to be interested in the same subject at the same time. Also, when a country reports on what it is doing, that will create more interest and demand to share experience. The network could thus support, news bulletins from countries. Each country could report on evaluation findings and other lessons, and this would allow other countries to judge if these results are relevant to them.
Another participant argued that a useful tool for this sort of network is the internet, using a website where information and requests can be posted. On the one hand, members of the network could post whatever experience their countries want to share with others, and also things that would relate directly to the two Banks. And interaction would be facilitated by making available the contact details of the network. This could lead to other important activities, such as visits or conferences or perhaps exchanges of papers among the countries.

Another participant agreed, stating that the benefit of a network is that it would provide economies of scale, and experiences which those who come late to these processes for institutionalizing M&E can learn from those who came before. A network could also share basic information on what they are doing and how they are doing it, such as the design of performance indicators; we should not reinvent the wheel.

He also flagged the issue of how to organize this network and who will be responsible for what. He argued that the Banks, at least in the very beginning, will have to help countries a lot, because many have very limited resources and great needs.

Another model was suggested by one participant: creation of a formal committee with two or three technical specialists from each country that would have a formal agenda. It could discuss technical matters and might meet two or three times a year, with the responsibility to be trained and to train trainers in the countries. Such a committee would work within the broader network that would be more informal in nature.

Manuel Fernando Castro, representing Colombia’s department of national planning, argued that there is a great deal of interest in M&E and evidence-based policy-making and that these justify the existence of a network. This provides a window of opportunity and a network of this nature is a way of keeping that interest alive. It also provides a vehicle for the countries and their public sectors to be more aware of the importance of the instruments of M&E.

One of the tasks of the network would be to disseminate the results of impact evaluations being undertaken in different countries. A second issue is that cutting edge-material is all in English. In Latin America bilingualism is still rare. So where there are relevant texts for strengthening M&E capacity, these need to be made available in Spanish, and here the network could also play an important role. It would also be interesting to offer courses in Spanish in M&E in the countries so that there would be broader access to this type of information.

He concluded by suggesting that the M&E network liaise closely with other networks in the region, such as the one dealing with evaluation of social policies — there are potentially strong synergies in those topics which would help to maintain a high level of awareness among public officials.

This suggestion was echoed by a participant from academia. She noted that the inaugural conference of the Latin American Evaluation Association (ReLAC — Red de Seguimiento, Evaluación y Sistematización de América Latina y el Caribe) was held last year, with the support of the World Bank, IADB and UN agencies. Such conferences and networks provide potential for synergies and harmonization. Similar support was provided for a regional conference for civil society held in Peru; civil society has a potentially important role to play in the processes of evaluation and transparency.

But there have been other regional conferences for M&E in Latin America, which have not been very successful. They had similar ideas: annual meeting, website, and newsletter. We need to learn why some of these efforts have not been so successful, to ensure our proposed M&E network does not make the same mistakes. What are the gaps and asymmetries of information that our network would fill? The network could help us all share lessons about the effectiveness of particular types of program — this would help us build a society for M&E, a club of evaluators.

Another participant flagged the experience of the National Association of Public Management (NASPA) in the U.S., which has promoted regional networks for civil society and public management. In Latin America it has supported the Inter-American Network for Public Administration Education (INPAE), with one objective being to achieve the sustainability of this network. It is now located at the University of Buenos Aires in Argentina. INPAE’s academic programs provide training for public administrators or managers and they provide space for analysis of shared experience at the workplace. These programs actively work with local actors, local governments, and NGOs. NASPA has concluded that it doesn’t matter where the network is based because it can use distance learning technology — the Global Distance Learning Network (GDLN).
Another participant endorsed the need to learn from past efforts at networking. A successful network would generate a cooperative system in which we all win. He emphasized the value of his participating in this conference, because his country ‘is at zero now’. He endorsed the proposal of a speaker from Mexico — that the network should be made sufficiently flexible to cater to different levels of interests of the countries and different levels of development in M&E. Thus would enable members of the network to see, for example, how OECD countries and the more advanced countries in Latin America have progressed. Another issue which could also be addressed is the approaches for financing M&E systems. Is it a percentage of income, of donor loans? How is this financed?

Jaime Saavedra concluded by thanking participants for their valuable comments. He noted that the conference has come at an interesting time and that governments recognize that accountability, which M&E supports, is a real priority. The demand test for creation of an M&E network has clearly been met; this justifies the network and will assist its sustainability.

He noted that the issue of Spanish is critical. Also awareness that these topics need to continue not only in a group such as this, but that outreach is necessary to help persuade others of the priority for M&E, in support of sound governance. The two Banks can help with the startup of the network, through concrete steps such as creation of a small committee with the two Banks and a group of countries — countries that have made substantive progress with M&E systems and other countries such as Guatemala which are in more of a learning position. But all countries can learn, because this is going to be a continuing area of work for all countries.

So there can be participation by countries or personal participation from many of the countries. A committee could define specific actions for the next six or eight months, including perhaps the creation of a website.

It will be critical to identify specific activities to manage the flow of information and then to establish priorities. These could be discussed with the broader group during the first year such as at the next conference. There will also be proceedings from this current conference. We will also circulate information about all of the participants here.

Roberto Garcia Lopez fully endorsed these comments. He added one further issue: we are talking about evaluation, and we should practice what we preach. So we need to evaluate the initiatives and networks that have failed in the past and avoid repeating the same mistakes.
PART FIVE:
Conference Conclusions
Some Key Conclusions

Conference Rapporteur: David Shand
World Bank Consultant

We have had five excellent country presentations, all of them reflecting many common and some different points.

Our first presentation was from Chile, which has a well developed and formalized system of monitoring and evaluation and performance management generally. It has a number of mutually dependent components. No substantial changes appear to be envisaged at this stage. In this sense, it is a fairly mature system. Chile regards it as well-functioning, and international observers of performance management and monitoring and evaluation regard it as a fairly well functioning system. This was also the conclusion of a recent World Bank evaluation of Chile’s M&E system. It is a system centralized in the Budget Office of the Ministry of Finance, although this does not preclude any ministry or agency carrying out its own evaluations for its own management purposes.

Brazil, again, reflects a reasonably sophisticated and formalized system, which has been in operation for four to five years, but has evolved over a longer period. It emphasizes self-evaluation, overseen by the Ministry of Planning. It aims to link national planning and budgeting. But it seems recognized that some changes are needed, that it is not fully mature and that it has to work better in relation to the new 2004-2007 national plan. In particular, there is a need for more in-depth evaluation and the quality of performance information needs improving — the latter point having been a recurring theme in this conference.

Mexico was a particularly interesting presentation for me as I had some involvement with its Comptroller and Budget Office some years ago when I worked for the OECD. If I recall correctly, at that time about nine volumes of performance indicators were presented to the Congress each year and all of this information was audited. But this performance information seems to have been little used. It seems that Mexico recognizes that it needs to learn from earlier false starts. Another factor in recent developments has been the move away from one-party rule and a desire for greater transparency. However, Congress has mandated a difficult requirement that every program be externally evaluated every year. It will be interesting to see how it manages the expectations that gave rise to this requirement.

Colombia has been working in M&E for a long time, starting with major economic and political reforms in the early 1990s. It has had institutional difficulties in getting coordination from the various actors in planning, budgeting and evaluation. It has focused so far more on evaluation of investment expenditures, rather than the recurrent budget. It has a strong emphasis on evaluation as a means of social control — keeping the citizens informed, even providing public access to the government website containing management information.

Peru is a more recent starter, and can learn from the experiences of the other four countries, although it has also learned from its own false starts beginning in the 1970s. Linking planning with budgeting has been problematical and evaluation, as in Colombia, has focused largely on investment expenditures. Like Colombia, Peru has emphasized the important role of M & E as a tool of social control.
and to improve the level of public debate on public policies and programs.

There have been many common themes in the five country presentations.

It seems clear that most countries feel that they lack good-quality performance information for resource allocation or for performance improvement. In relation to the budget system, there is not enough information systematically available on what the government is buying with its expenditures, whether outputs or outcomes or impacts. So a common theme has been the need to better link evaluation information with the budget, and thus improve the quality of public expenditures. At least if the budget gives some indication of how expenditures relate to national priorities, that would be a start, but this is not the case with a number of countries in the region which have articulated national plans. But we need to beware of ‘paralysis by analysis’, and accept that we will never have perfect information.

As was pointed out, the development of monitoring and evaluation has, in some cases (such as Brazil and later Colombia, in particular), been driven by fiscal constraints — the need to do more with less, and in particular, a need to address poverty issues within very tight budget constraints. This need to do more with less is also one of the themes for public management reforms in a number of OECD countries.

Another common issue has been rigidities in the expenditure side of the budget. What is the point of doing evaluation to improve resource allocation when the budget system provides limited flexibility for reallocating resources? Chile has a budget bidding fund, which although not available every year, is designed to provide some additional funds for new programs or changes to existing ones. But this lack of expenditure flexibility is not confined to Latin American countries, and, indeed, exists in quite a number of OECD countries. But perhaps in the LAC Region, there is even less flexibility to transfer or to reallocate civil service staffing than exists in OECD countries.

Many of the M&E initiatives we have talked about in this conference have been driven by central ministries — or, as we say, ‘top down’ rather than ‘bottom up’. This raises the issue of whether there is enough emphasis on ministries and agencies doing their own evaluations, albeit at a lower level than impact evaluation — perhaps just output evaluation might be a useful step in the case of a number of ministries. In Brazil, evaluation is primarily the responsibility of those who are implementing the programs, but that was not the message which came across in most of the other country presentations. However, collaboration with — or participation of — program managers in centrally-driven evaluations is possible, as illustrated by the Chilean system.

Another common theme has been indicator overload — volumes of unused performance indicators. Many people have talked about the need for fewer, but better quality performance indicators, but I think we need to be careful here. Chile has some 1,500 indicators, and Peru said their experiences suggests the fewer indicators the better. Colombia reduced the number of indicators from 940 in 2002 to some 300 today. Mexico used to report some 3,000 indicators in the budget documents. It has reduced the number to around 350, but thinks the optimum number is about 250. We can play a numbers game, but we need to recognize that there are indicators at different levels. The indicators relevant to the higher level of resource allocation through the budget are different from the indicators relevant to the performance management of an individual ministry or agency. While we could argue inconclusively about the appropriate number of indicators to present to Congress in the budget documents, we agree on the need for performance measures that are simple enough to be understood by elected officials, but sophisticated enough to be relevant.

This relates to another common theme of this conference — an alleged lack of interest by legislatures in performance information. Of course, most of us are bureaucrats or former bureaucrats and some of us may a have an inherently negative view of legislative activity. But as was discussed in this conference, we at least owe it to the legislature to present performance information in useable and intelligible form. This surely reflects an element of good M&E practice — that evaluation results and performance indicators need to be conveyed to all potential users in a relevant and useable form. It will be interesting to see the response to the new results-based annex to the budget documents, which Colombia is going to present to its legislature.

Another general theme on which we have agreed and which reflects OECD experiences, is the need to measure the right thing and to avoid indicators which have dysfunctional consequences — because as is often said what is measured is what counts, and thus influences management behavior.

I have four final points. Firstly, in a number of Latin American countries monitoring and evaluation is based on
the national planning system and the objective of feeding results back into that system. When comparing regional experiences with OECD countries, we should note that no OECD country has a national planning system. These national planning systems probably encourage the focus on impact evaluation that we have noted in this conference. We recognized that impact evaluation is very important, but like many important things, it is actually more difficult. In particular we need to understand the cause-effect relationships within the program if we want to make changes to improve program performance. Whose activity is responsible for these impacts? If everybody is responsible, then nobody is responsible for the program's performance — a point made in the Peruvian commentary.

Secondly, there is the challenge of evaluation at the sub-national government level, particularly in a federal system, where a substantial number of national programs may be delivered at a provincial or state level. Even in a more unitary governmental system, it is an issue with respect to municipalities. How do we collect the performance information, what level of central oversight is needed and thus what type of information is needed at the national level as opposed to the local level? How do we replicate the M&E systems at the sub-national level? — a point particularly stressed in the Colombia and Peru presentations.

Thirdly, we seem to be in general agreement about the general strategy which should govern the way we go ahead with M&E. We should avoid being too ambitious, take things step by step, and not raise unrealistic expectations and oversell what M&E can deliver. Otherwise good initiatives will be judged by unreasonable standards, and deemed to have failed. We need to take time to do these things properly. I think the Ceara presentation from Brazil captured the issue very well. If we want M&E and results based management to take root we need to be very clear about why we are doing it, how we will do it, for whom we are doing it and how we are going to implement it.

Finally, this conference has clearly identified the lack of skilled evaluation resources and the high cost of using international consultants. This raises the need for training within the region. And while evaluation may be for experts in a technical sense, implementing monitoring and evaluation is for all managers. That means everybody has to be trained about the role of monitoring and evaluation, as part of performance management. So it is not just the training of professional evaluators; it is also the training of line ministry staff and central ministry staff, whether that be Finance, Planning, Presidencia or Comptroller, — and of course, not forgetting the Legislative Branch. All this will require time and resources.

So a lot of progress has been made, but much remains to be done.
Closing Session

Vinod Thomas
World Bank Country Director for Brazil

In this conference there have been quite a few far-reaching insights into the monitoring and evaluation of programs. These insights should be kept in mind as we put M&E into practice.

One set of insights relates to policy settings — such as how we evaluate the power of ideas, how we evaluate efforts that are ‘out of the box’ in expanding the range of program options. There are several examples from Brazil as well as other countries, and these can illustrate how countries can do things differently or better as they go along. That is more a question of the evaluation of policies or reforms on which programs are based, complementarily with the evaluation of those programs.

At the Shanghai conference the question of scaling up naturally arose as one looked at the scale of operations in China. Using comparable measures of poverty, Brazil and China have say, 15 percent of their people in poverty. For China or India that translates into a very different absolute number — more than the entire population of Brazil. Translating lessons from one country to the other caught the imagination of a lot of people at that conference. For example, Brazilian attendees were struck by the fact that urbanization, which often is seen as adding to poverty in Brazil and Latin America, is seen in China as an instrument for addressing poverty. So how does one view urbanization in the context of addressing poverty differently in two large countries? Another example is the role of programs which subsidize food prices or employment as a way of addressing poverty, compared to direct and targeted transfers, which Brazil and some other countries are trying. There is a stark contrast in policies between China, India or Egypt on the one hand, and Brazil and some others on the other.

Let me now refer to the World Bank Institute (WBI) programs, which have played a leading role in the northeast of Brazil, including supporting a network of practitioners and beneficiaries, and also in the Amazon, where poverty is also an important issue, along with major issues of environmental and social sustainability. So WBI programs which address the Brazilian situation have quite different settings from those in China. These cross-country experiences help us better understand the facilitation role of institutions like WBI and the knowledge exchange that is facilitated by a conference such as this.

There are other interesting comparisons to be made. In China, once programs are decided on, it is extraordinary how everybody pulls together and delivers good results on the ground. In this sense, participation is quite high. But at the level of participation in the design or choice of programs, Brazil and some other countries arguably have a much higher degree of participation.

This conference is taking place at a very opportune moment. Countries want to do more evaluation, to measure impact as far as possible, and then to revise both programs and policies.

Manuel Rapoport
Manager of the Inter-American Development Bank’s Development Effectiveness and Strategic Planning Department

As an institution, the IADB together with the World Bank is 100 percent committed to achieving effectiveness in deve-
development. So I want to talk about development effectiveness or impact. This concept comes together with that of results-based management. To achieve development effectiveness we have to understand where we are going, otherwise we will never know when we have arrived. The way in which we reach our objectives is very important, and we need to look at available resources — human, financial, etc. The idea is to reach our objectives by efficiently managing those resources.

We at the IADB, and the World Bank and other multilateral and bilateral agencies, have committed ourselves, through the Paris Declaration, to harmonization. We are talking about and working on aligning our work with country systems. But that is easier said than done. Different countries have different problems. Different sectors have different issues. So we have to work country by country and sector by sector to reach our goals.

We understand that the transaction costs of doing business with institutions such as ours are very high. We have to reduce that cost and we are moving in that direction. To be able to align our procedures with country procedures we have to work to assist institutional strengthening in countries, at the central, ministry and sub-national levels. We are going to do this together with the countries, using a number of tools, such as World Bank and IADB technical assistance and lending programs.

We have devised a development plan with three pillars — internal, external and corporate. Internal has to do with the efforts and skills of our own people, so that they understand how to work with the countries. But most of our effort is geared toward the external pillar which has to do with assisting countries to strengthen their institutions. We start at the national level, but will move down to the sub-national level. For this external pillar we have a $30 million fund, to be divided between 26 countries. It is seed money that will allow us to build or develop countries’ results-based measurement systems.

So the IADB and the World Bank and you as country officials are all going to work together. We have a new framework and we are changing the way we do business. We are also changing our project base and the kind of programs we support.

Results-based measurement systems are important: for accountability to our institutions, we need to know how this money is spent. You as country officials are also accountable for the government resources allocated to different programs. In this conference, we have been discussing performance benchmarks — which are important but complex. We need to be able to identify clear objectives and the level of public resources we are allocating to achieve our objectives. What are the performance benchmarks to measure whether objectives were reached? We need your help to define them. There will be different benchmarks according to the country and the sector involved. And we need to ensure that there is sufficient management flexibility so that those objectives can be met. If we see that objectives are not being reached we have to be able and prepared to change our path. So we need to be flexible in our new approaches.

**Ernesto May**

*Director of Poverty Reduction and Economic Management in the World Bank’s Latin America and Caribbean Region*

In this conference we have been able to share the experiences in our countries in recent years, in an area of major interest to the development community — monitoring and evaluation.

One key message has been that M&E is basically one tool or instrument. It is only a means to an end, not an end in itself. But M&E can have a catalytic role in improving public sector management. In this conference we have heard about key areas such as the quality of government expenditures, accountability and transparency, and institutional capabilities. M&E in these areas can generate demand to improve the quality of public expenditure, in order to have better accountability and transparency of public expenditures to the citizens, and this would also lead to the need to better utilize human and budgetary resources, and to improve public sector institutional capacity to deliver services.

So the work in the area of monitoring and evaluation can be seen as a catalyst for improving public sector management in all of our countries. Clearly we are not alone in this effort; we are all seeking to improve our M&E systems involving a very large group of people within the region. Obviously, the Anglo-Saxon experience is something that we are going to keep in mind along with the experiences of the OECD countries. But we have also a lot to learn from within the Latin American region.

In this context, an important result of this conference, in addition to sharing experiences, is the setting up of the
regional network on M&E. It seems that there is a clear demand for the sort of support that can come from this network — to strengthen the work that each country is doing and to use this network as an input in the work that we are doing — with the objective of improving public sector management, and leading, as a final result, to poverty reduction and a better standard of living for the citizens of our countries. I hope that this will be only the beginning of a way to more effectively work together moving forward.
Towards the Institutionalization of Monitoring and Evaluation Systems in Latin America and the Caribbean: Proceedings of a World Bank / Inter-American Development Bank Conference